Social information processing biases in antisocial behaviour

Tamara Silvia Satmarean

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Science

Department of Psychology

March 2022

Acknowledgements

I would like to thank my supervisors, Professor Richard Rowe and Professor Elizabeth Milne. Thank you for being models of leadership, integrity, and academic excellence. Thank you for supporting me through all my weird and wonderful ideas, for the countless hours of conversations about research. It has been a privilege, I could not have asked for better supervisors and I will always be grateful.

I would like to thank all the participants who took part in this research, your contribution was invaluable.

I would like to thank God, without whom none of this would have been possible. I would also like to thank my family, my mother Sefora, my sister Flavia, and my brothers Paul and Dan, for their unwavering support and constant encouragement that I can achieve anything. You are my favourite people and I would not be where I am without you.

Abstract

Antisocial behaviours (ASB) and aggressive behaviours have been shown to have a considerable financial and social impact (Eme 2009). Social-cognitive theories and models of aggression have increasingly focused on cognitive biases and their immediate effects on behavioural aggression. This thesis investigated the relationship between cognitive biases, aggression and antisocial traits and behaviours using Crick and Dodge's (1994) Social-Information Processing (SIP) model which provided a structured approach to the examination of cognitive biases (i.e., systematically distorted representations, Haselton et al., 2015). The model postulates a central role of memory and emotion; however, the mechanisms whereby this influence is exerted on stages of the SIP model or behaviour are not well understood.

Three empirical studies were conducted. A combination of targeted and convenience sampling was used in the first study which recruited participants with antisocial traits and a history of ASB and undergraduate students. The dot-probe task (DPT) was used in the first study to measure attentional bias to threat during encoding. Attentional bias was not related to aggression and CU traits; however, a relationship was found between attentional bias to angry faces and SIP measurements. In the second and third studies, a novel dual working memory (WM) and visual search task demonstrated evidence of WM guidance of visual attention in former offender (study 2) and community (study 3) samples. WM guidance of visual attention to emotional faces was linked to aggression, trait anger, emotion dysregulation, and later stages in social-information processing. Additionally, emotion and ability to regulate emotion has been linked with early and late stages in social-information processing. These experiments were designed to provide a holistic view of how information is processed in social situations and of associations between deviant cognition patterns and aggression, antisocial traits, and behaviours.

List of Contents

Acknowledgements 2
Abstract 3
List of Contents 4
List of Tables 7
List of Figures 10
Declaration 13
1. Chapter 1: Literature Review 14
1.1. Aggression and links to antisocial traits and behaviours 14
1.1.1. Conceptualisation, prevalence, and cost of aggression 14
1.1.2. Overlap between constructs related to aggression 15
1.2. Overview of social-cognitive models of aggression 17
1.2.1. The General Aggression Model 19
1.2.2. Crick and Dodge's (1994) SIP Model 20
1.2.3. Huesmann's SIP Model 20
1.2.4. Arsenio and Lemerise's SIP Model 21
1.2.5. The SIP model of aggression 22
1.2.6. The role of emotion in SIP 28
1.2.7. The role of memory in SIP 31
1.3. Attentional bias 33
1.3.1. Introduction to attentional biases 33
1.3.2. Theories of attentional bias 34
1.3.3. Measurements of attentional bias 35
1.3.4. Attention guidance by WM 38
1.3.5. Attentional biases: links to aggression, antisocial traits and behaviours 39
1.3.6. Attentional biases in the SIP 40
1.4. Attribution bias 42
1.4.1. Introduction to attribution bias 42
1.4.2. Attribution bias in the SIP model 43
1.4.3. Attribution bias and aggression 43
1.5. Context for empirical work 45
1.6. Thesis aims and outline of studies 47
2. Chapter 2: Study 1- Attentional bias to threat: Links to antisocial traits and behaviours 48
2.1. Abstract 48
2.2. Introduction 48

2.3. Method **51**
2.3.1. Participants 51
2.3.2 Measures 52
2.3.3. Analysis Plan 56
2.3.4. Procedure 58
2.3.5. Data Preparation 59
2.3.6. Computation and reliability of attentional bias data and self-report measures 60
2.4. Results **62**
2.4.1. Sample characteristics: self-report measures 62
2.4.2. Online attentional bias data collection- Exploratory analyses 65
2.4.3. Results relating to pre-registered hypotheses 66
2.5. Discussion **75**
2.5.1. Limitations 83
2.5.2 Conclusions and future directions 84
3. Chapter 3: Study 2- Working memory guidance of visual attention to threat in offenders* *86
3.1. Abstract **86**
3.2. Introduction **86**
3.3. Method **89**
3.3.1. Participants 90
3.3.2. Measures 90
3.3.3 Dual task 92
3.3.4. Analysis plan 93
3.3.5. Stimuli 95
3.3.6 Procedure 95
3.3.7. Data Preparation 96
3.3.8. Computation and reliability of attentional bias data 97
3.4 Results **99**
3.4.1. Results relating to pre-registered hypotheses 99
3.4.2. Exploratory analysis 104
3.5. Discussion **105**
3.5.1 Limitations 111
3.5.2. Conclusion 114
4. Chapter 4: Study 3- Cognitive biases associated with aggression and trait anger: the role of emotional state and emotion dysregulation* *115
4.1. Abstract **115**
4.2. Introduction **115**
4.3. Method **120**
4.3.1. Participants 121
4.3.2. Measures 121
4.3.3 Emotion Induction 123
4.3.4. Dual task 123

4.3.5. Analysis Plan....124
4.3.6. Stimuli....128
4.3.7. Procedure....128
4.3.8. Data Preparation....128
4.3.9. Computation and reliability of attentional bias data....130

4.4. Results....130

4.4.1. Manipulation check....130
4.4.2. Results relating to pre-registered hypotheses....133

4.5. Discussion....143

4.5.1. Limitations....153
4.5.2. Conclusion....153

5. Chapter 5: General Discussion....155

5.1. Review of research aims and objectives....155

5.2. Limitations....156

5.3. Attentional biases to social stimuli: links to aggression, delinquency, and CU traits....160

5.4. Working memory biases of visual attention to social stimuli: links to anger and aggression....162

5.5. WM and emotion mechanisms in the SIP model of aggression in a community sample....169

5.6. WM biases of visual perception in the SIP model: anger enhances processing until it does not....172

5.7. Clinical and practical implications....180

5.8. Directions for future research....184

5.9. Concluding comments....185

Appendix A: Dual Task- Development and validation....187

Appendix B: Hierarchical multiple linear regression analyses with CU as IV and exploratory correlations including CU traits....190

Appendix C: Exploratory Analyses- Are WM visual selection biases predicted by social-emotional information processing variables?....199

Appendix D: Anger manipulation check....204

References....235

List of Tables

Table 1: Means and Standard Deviations for aggression, CU traits, self-reported delinquency, SEIP-Q variables......63

Table 2: Correlation table for bias scores, aggression, CU traits......64

Table 3: Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces ($BS_{POSITIVE}$)......69

Table 4: Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces ($BS_{NEGATIVE}$)......70

Table 5: Correlation matrix for self-report measures of Reactive Aggression, CU traits, and scores on the SEIP-Q......73

Table 6: Bias scores: computation and reliability......98

Table 7: Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$) 101

Table 8: Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$) 102

Table 9: Hierarchical multiple linear regression analysis predicting attentional bias away from neutral faces whilst holding an angry face in WM ($BS2_{NEGATIVE}$) 103

Table 10: Skewness and kurtosis estimates for the variables included in the pre-registered structural equation model......130

Table 11: Means and standard deviations for state anger across emotion induction conditions, before and after emotion induction......132

Table 12: Correlations between variables included in the pre-registered structural equation model......140

Table B1: Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$) 190

Table B2: Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$) 191

Table B3: Hierarchical multiple linear regression analysis predicting attentional bias away from neutral faces whilst holding an angry face in WM ($BS3_{NEGATIVE}$) 192

Table B4: Correlation matrix for Trait Anger, Aggression, Inventory of CU traits scores, and scores on the SEIP-Q..194

Table B5: Means and Standard Deviations for Aggression, CU traits, Trait Anger, Attributional and Emotional response variables...197

Table C1: Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$) with SEIP-Q variables as predictors...200

Table C2: Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$) with SEIP-Q variables as predictors...201

Table C3: Hierarchical multiple linear regression analysis predicting attentional bias away from neutral faces whilst holding an angry face in WM ($BS2_{NEGATIVE}$) with SEIP-Q variables as predictors..202

Table D1: Wilcoxon signed rank tests: Comparison of emotion induction conditions..........215

Table D2: Mean and median scores for induced emotional conditions: composite scores, before and after emotion induction...216

Table D3: Mean and median scores for induced emotional conditions: anger sub-scores, before and after emotion induction..216

Table D4: Comparing trait anger scores across emotion induction conditions...................216

Table D5: Means and standard deviations for RTs in the high state anger (post visual task) group...224

Table D6: Means and standard deviations for RTs in the low state anger (post visual task) group..225

List of Figures

Figure 1: An integrated model of emotion and cognition in Social-Information Processing based on Crick and Dodge's SIP model (1994), the Model of Response Evaluation and Decision (Fontaine & Dodge, 2006), and the Ecological Model of Hostile Attribution Bias (Dodge et al., 2015) 23

Figure 2: Trial-level bias scores for the 100 ms presentation condition: time series for all participants...62

Figure 3: Trial-level bias scores for the 500 ms presentation condition: time series for all participants...62

Figure 4: Interaction between congruency and target presentation duration in the high-delinquency group. Error bars show one standard error above and below mean.....................67

Figure 5: Interaction between congruency and target presentation duration in the low-delinquency group. Error bars show one standard error above and below mean.....................68

Figure 6: Dual task trial example...93

Figure 7: Changes in reaction times as a function of congruency and emotional value. Error bars show one standard error above and below mean..100

Figure 8: Pre-registered structural equation model of the moderating role of emotion dysregulation in the relationship between WM guidance of attention, hostile attribution, anger, and aggression...126

Figure 9: Interaction between emotional valences of WM template and induced emotion. Error bars show one standard error above and below mean...134

Figure 10: Interaction between emotional valences of WM template and target. Error bars show one standard error above and below mean...135

Figure 11: Interaction between emotional valences of WM and target in the induced anger condition. Error bars show one standard error above and below mean...................................137

Figure 12: Interaction between emotional valences of WM and target in the induced neutral condition. Error bars show one standard error above and below mean.................................138

Figure 13: Path model of the moderating role of emotion dysregulation in the relationship between WM guidance of attention, hostile attribution, anger, and aggression, and standardized parameter estimates...141

Figure D1: Emotional state before induction (all participants) 205

Figure D2: Emotional state after induction (all participants) 206

Figure D3: Emotional state before induction: scores in the neutral emotion induction condition..207

Figure D4: Emotional state before induction: scores in the anger induction condition.........208

Figure D5: Emotional state after induction: scores in the neutral emotion induction condition..209

Figure D6: Emotional state after induction: scores in the anger induction condition............210

Figure D7: Anger sub-scores before anger induction: anger induction condition.................211

Figure D8: Anger sub-scores after anger induction: anger induction condition...................212

Figure D9: Anger sub-scores before emotion induction: neutral state induction condition…213

Figure D10: Anger sub-scores after emotion induction: neutral state induction condition...214

Figure D11: Emotional state before-after: Induced anger group..217

Figure D12: Emotional state before-after: Induced neutral emotion group.........................218

Figure D13: Interaction between induced emotion condition and time (before vs after emotion induction). Error bars show one standard error above and below mean...............................219

Figure D14: Interaction between state anger group and emotional valence of the WM template. Error bars show one standard error above and below mean..221

Figure D15: Interaction between emotional valences of WM and target in the low state anger (post visual task) group. Error bars show one standard error above and below mean...........223

Figure D16: Interaction between emotional valences of WM and target in the high state anger (post visual task) group. Error bars show one standard error above and below mean...........224

Figure D17: Interaction between emotional valences of WM template and emotion induction condition. Error bars show one standard error above and below mean...............................229

Figure D18: Interaction between emotional valences of WM template and target. Error bars show one standard error above and below mean..230

Figure D19: Interaction between emotional valences of WM and target in the induced anger condition. Error bars show one standard error above and below mean...............................232

Figure D20: Interaction between emotional valences of WM and target in the induced neutral condition. Error bars show one standard error above and below mean...............................233

Declaration

I, the author, confirm that this thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Publications:

Satmarean, T. S., Milne, E., & Rowe, R. (2022). Working memory guidance of visual attention to threat in offenders. *PloS One*, *17*(1), e0261882. https://doi.org/10.1371/journal.pone.0261882 – This publication was included in the thesis as Study 2 (Chapter 3).

Name: Tamara Satmarean

1. Chapter 1: Literature Review

1.1. Aggression and links to antisocial traits and behaviours

1.1.1. Conceptualisation, prevalence, and cost of aggression

Aggression is typically defined as behaviour directed toward another person with the intention to cause harm (Anderson & Bushman, 2002). It is an innate behaviour, one that is present throughout the lifetime (Ferguson & Dyck, 2012) and stable throughout development (Dodge et al., 1995, Dodge et al., 2003). Advances in understanding of forms of aggression, risk factors, and ecological influences have resulted in a shift in research toward examination of why these relationships exist and mechanisms underlying the development and maintenance of aggression (Lansford, 2018). The latter has been shown to have detrimental effects on individuals and society both in terms of harm and economic costs. A recent report by the European Union Agency for Fundamental Rights (FRA) showed that in 2020-2021 over 22 million individuals in Europe were physically attacked (Fundamental Rights Report, 2021). Costs of violence and aggression were considerably higher than those of theft or property damage (Home Office, 2018). The cost for violent crime in the United Kingdom between 2015-2016 was estimated at over £37.4 billion (Home Office, 2018) and was believed to be much higher in the United States (i.e., 426 billion US dollars, Mancke, Herpertz, & Bertsch, 2018).

Given the high prevalence and cost of aggressive behaviour, the functions of aggression, referring to the reasons why aggression is being enacted, have increasingly become the focus of research. To that end, a distinction has been made between reactive aggression, which is associated with strong negative emotions and enacted in response to frustration and perceived provocation (Wilkowski & Robinson, 2008), and proactive aggression, which is premeditated, aimed at achieving a desired outcome (Cenkseven-Önder et al., 2016). Under a new ecological framework (Lansford, 2018), risk factors for aggression have been identified by taking into account microdynamics (e.g., parent-child interactions, strengthening aggressive associations throughout development) and macrodynamics (e.g., escalation of aggressive patterns through social interaction). For instance, maltreatment during childhood has been shown to increase the likelihood of perpetuation of aggression in adulthood by 42% (Dodge et al., 1990) and viewing relational aggression on television predicted subsequent relational aggression (Coyne, 2016). Conceptualising risk factors under this new framework highlights the importance of considering aggression as resulting from individual tendencies as well as situational factors.

To address the mechanisms involved in the aetiology and maintenance of aggression and externalizing behaviours, research has focused on social-information processing paradigms which describe a series of cognitive steps preceding behavioural outcomes, having a direct effect on the likelihood of producing aggressive responses (Al Moghrabi et al., 2019). Early risk factors such as childhood exposure to violence and rejection are believed to be associated with cognitive processing biases and subsequent aggression. Dodge et al. (2015) tested this assumption in a sample of children from nine countries and found that across the sample, increased hostile attribution of intent in ambiguous situations predicted aggression. Although social-cognitive paradigms of aggression have initially been used to study aggression during development (Li et al., 2013), studies have successfully applied this framework in adult and offender populations (Coccaro et al., 2017). These findings suggest a closer examination of mechanisms through which risk factors increase the likelihood of aggression would be beneficial. Specifically, understanding cognitive mediators of aggression would provide opportunities for targeted interventions.

1.1.2. Overlap between constructs related to aggression

Research has often linked aggressive behaviours to risk factors for antisocial disorders and ASB. The American Psychological Association (APA, 2021), described ASB as a behaviour which deviates from the norm, encompassing behaviours such as property damage, violation of social norms and rules, and theft. Antisocial traits are those which predict an increase in the likelihood of engaging in ASB (Pfabigan et al., 2012). Comparatively, aggression was defined as the behaviour aimed at causing physical or psychological harm, either by intentionally causing injury or destruction (hostile aggression) or to achieve a goal (instrumental aggression). Most recently the APA defined antisocial aggression as comprising acts of instrumental aggression which have socially destructive consequences (e.g., rape, murder, APA, 2021). There is considerable overlap in the behaviours comprised within these definitions. The definitions are also different to those used to group these behaviours in the criminal justice system and policing, where ASB is used to describe behaviours which cause or are likely to cause harassment, alarm or distress to others (Metropolitan Police, 2020).

Similar to aggression, ASB comes at a great cost and has a considerable negative impact on victims and society (Eme 2009). ASB is present throughout development (Moffitt, 1993). In the United Kingdom, treatment for children referred for ASB has been estimated at £6000 per year (Romeo et al., 2006) with average cumulative costs (over a 23-year period) of childhood conduct problems of more than £37,000 (Rissanen et al., 2021). In the United States, an individual with a lifetime history of violence can cost society up to 2 million dollars (Dodge & McCourt, 2010). ASB is highly prevalent (Ferguson & Dyck, 2012) and predicted by multiple risk factors, none of which can consistently predict it (Butler et al., 2015). Risk factors such as relational aggression, emotional and behavioural disorders, and substance use have been linked to increased engagement in ASB; longitudinal studies however have shown that a history of aggression and ASB are the strongest predictors of ASB (Assink et al., 2015). Moreover, a persistent pattern of externalizing problems such as theft, destruction of property, and rule-breaking during development often form the basis of a Conduct Disorder diagnosis (Searight, Rottnek & Abby 2001), a precursor of antisocial personality disorder (ASPD). Whilst cognitive operations preceding behavioural outcome have a proximal effect on ASB, they are affected by risk factors and individual differences in antisocial and aggressive traits (O'Reilly et al. 2015).

Aggression has also been linked to callous-unemotional (CU) traits and deficits in planning and inhibitory control (Thompson & Centifanti, 2018). As a proxy measure of the affective dimension of psychopathy, CU traits appear specifically linked to proactive aggression whereas the impulsive dimension of psychopathy has been linked to reactive aggression (Raine et al., 2016). There is considerable overlap in antisocial and aggressive behaviours in individuals with a history of violence and aggression. Moreover, given the overlap in measurements of traits related to aggression, established measures of aggression, history of ASB, and CU traits were included in this research in order to examine their relationships with (and individual predictions of) cognitive biases in social-information processing.

1.2. Overview of social-cognitive models of aggression

Social-cognitive models of aggression posited that social learning forms an important component in the development of aggression. Learning was associated with aetiology and maintenance of antisocial and aggressive beliefs and behaviours (Ferguson & Dyck, 2012, Dodge et al., 2015). Social cognitive theories of aggression have incorporated elements of priming and learning from social learning theories into cognitive models which predicted aggressive outcomes as outputs preceded by cognitive operations (Ferguson & Dyck, 2012). Given that aggression and violence occur throughout development (Boxer et al., 2005), prevention research has increasingly drawn on such cognitive theories of aggression for intervention design.

Social-cognitive models of aggression allow the study of cognition within social and cultural contexts (Matjasko et al., 2012). The main strength of such models however, is the emphasis on measurement of cognitive processes immediately preceding aggression or ASB. Social-cognitive models differ in focus and scope; these differences and implications for research are outlined below. Nonetheless, they share the common assumptions that aggression likely results from biases in the sequence of steps underlying social-information processing (Crick & Dodge, 1994), that aggressive individuals have a more extensive repertoire of aggression-related knowledge structures such as schemas and scripts, and that emotion is fundamental to the processing of social information (Arsenio & Lemerise, 2004).

According to such social-cognitive theories, aggressive social schemas, scripts, and beliefs are acquired through exposure to and learning of aggressive behavioural patterns. These create blueprints for future social interactions which guide processing of information in social settings (Huesmann & Taylor, 2006). Further, individual risk factors such as temperament, intelligence, social skills, and ecological risk factors such as exposure to conflict and abuse during childhood, or community violence, together create social-cognitive patterns (Lansford, 2018). These are believed to crystallize throughout development and enhance the endorsement of aggressive and antisocial outcomes in response to a wide range of social situations (Boxer et al., 2005; Kofler et al., 2015). The main focus of cognitive-ecological models was to provide an increased understanding of how cognitive biases immediately precede aggressive behavioural choices with the aim to target deviant cognitive processes through interventions (Dodge et al., 2015).

An idiographic approach is helpful in understanding how risk factors and cognitive biases shape processing of social information and how individual differences in measurements of antisocial traits and behaviours contribute to aggressive behavioural outcomes. Such an approach would contribute to the improvement and effectiveness of intervention research (Poldrack et al. 2017). Haselton, Nettle, and Murray (2015) argued that “the direction and content of biases is not arbitrary” (p.971) with biases conceptualised as effective cognitive shortcuts enabled by previous experience. These shortcuts have been shown to be maladaptive when disproportionately guiding processing toward hostility-related information or hostile interpretations of ambiguous social information, leading to the generation of aggressive responses (Dodge et al. 2003). Hence, understanding the cognitive basis for this distinction, how and why social information is processed differently, is fundamental to the study of decision making leading up to aggressive behaviour, and the main focus of social cognitive theoretical paradigms.

1.2.1. The General Aggression Model

The General Aggression Model (GAM) has emerged as an overarching theory (Anderson & Bushman, 2002). The GAM draws on the social cognitive script theory, conceptualising aggression as an outcome of activation and application of previously acquired aggressive scripts and schemas held in memory (Bushman, 2016). Choice of script and evaluation of its suitability in a given social context is mediated by memory and goal driven processes (Huesmann, 1988). The GAM posited that aggression is always harmful, learned, and automatic with measurements of personality dimensions being an expression of acquired scripts and schemas. Numerous studies however have demonstrated the increased risk of aggression associated with biological factors (Beaver et al., 2011) as well as instances where aggression involves substantial forethought. Moreover, the harmfulness assumption disregards the adaptive and proactive uses of aggression and anger manifestations, as well as the potential use of aggressive behaviours in the interest of protecting others (prosocial aggression, APA, 2021). As an over-arching theory of social information-processing and aggression, the GAM has proven inadequate, failing to provide predictions that can be mapped onto real-world aggression (Ferguson, 2010). Moreover, the theory has been shown to be empirically inadequate lacking criteria for falsification and failing to account for emerging findings in the field of aggression such as distinction between aggression subtypes and roles of cognitive biases in generating behavioural outputs (Ferguson & Dyck, 2012).

1.2.2. Crick and Dodge's (1994) SIP Model

Dodge's (1986) initial SIP framework was developed to explain aggression as a function of on-line cognitive processes. Crick and Dodge's (1994) original model comprised five social information-processing steps, namely encoding, interpretation of cues, choice and clarification of goals, response construction (or selection), and response decision. The model focused mainly on the effect of these steps on behaviour in specific instances. This was reformulated to include the role of latent knowledge structures leading to a shift in developmental research from an exclusive focus on measurement of latent variables to on-line processing, immediate effects on behavioural outputs and links to psychopathology (Li et al., 2013). Initially aimed at understanding cognitive processes leading to aggression in children, the SIP posited that during an interaction, patterns of processing leading up to a behavioural outcome can be measured and the individual contribution of deviant patterns to aggression could be predicted (Crick & Dodge, 1994; Dodge et al., 2013). Over the last three decades, research using SIP models of aggression has consistently linked cognitive biases and individual differences in antisocial and aggressive traits to aggressive behaviour (Dodge et al., 2015; Harper et al., 2010).

1.2.3. Huesmann's SIP Model

Huesmann's SIP model similarly explained the development and maintenance of aggression through sequential cognitive tasks (Huesmann, 1988). However, there were key differences between Huesmann's and Crick & Dodge's SIP models. Namely, Crick and Dodge's revised SIP model (1994) focused on the effect of on-line processing on social outcomes specific to the social context. Comparatively, Huesmann's (1988) model postulated that cognitive processes preceding behavioural responses were automatic and their main focus was posited to be on either script creation or retrieval of previously acquired scripts. Due to the increased focus on latent structures, this model was largely disregarded by the academic community (Li et al., 2013). In contrast, Crick and Dodge's (1994) model addressed a gap in cognitive understanding and focused on context-specific processing, accelerating research into cognitive processing patterns associated with psychopathology.

1.2.4. Arsenio and Lemerise's SIP Model

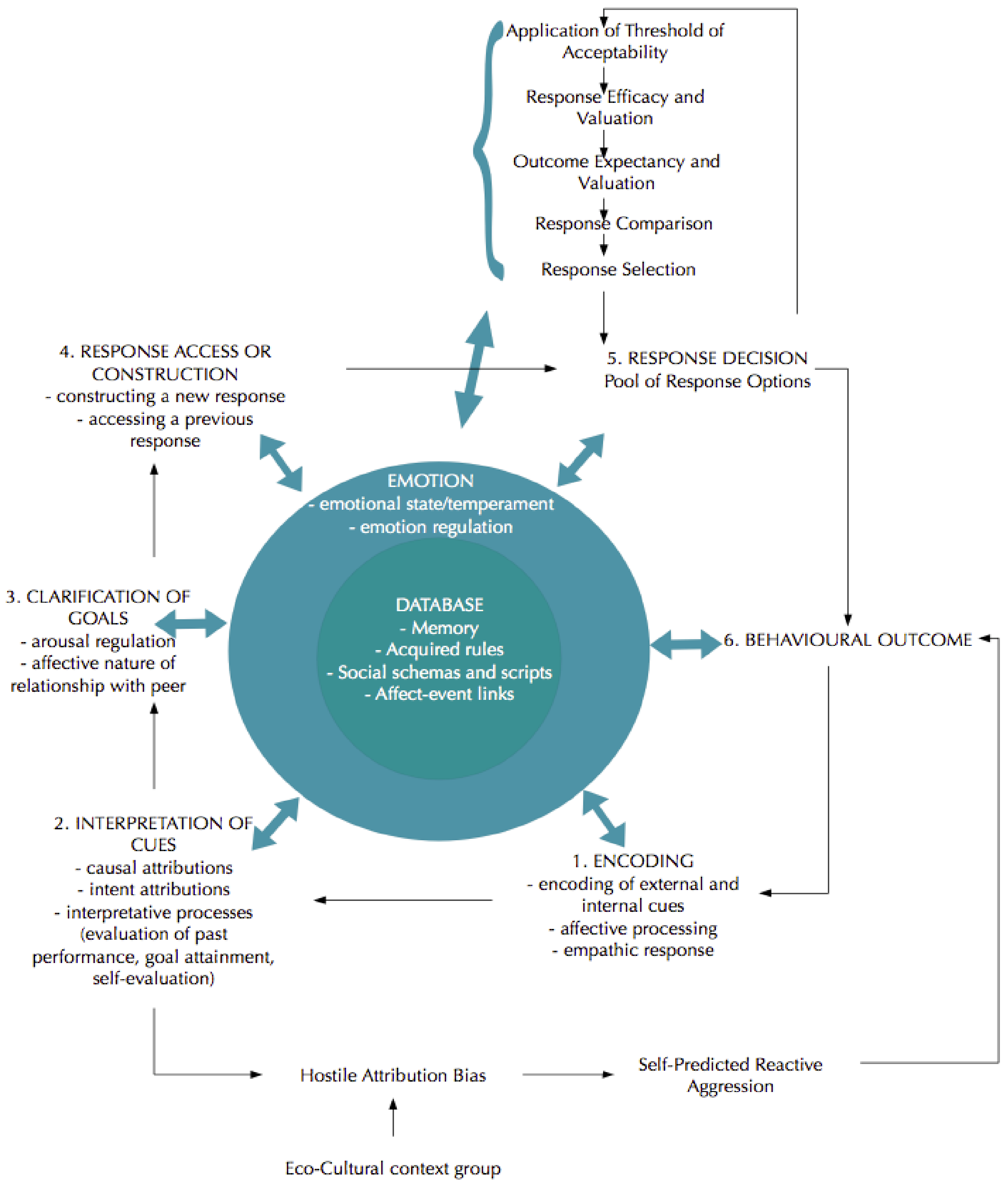
Despite its popularity, Crick and Dodge's SIP model was subject to on-going critique. Building on the SIP model introduced by Crick and Dodge (1994) and Huesmann (1988), Arsenio and Lemerise' (2004) model aimed to explain the aetiology and maintenance of aggression through latent structures, online processing, and crucially, through the role of emotion throughout social-information processing. Arsenio and Lemerise (2004) critiqued Crick and Dodge's (1994) model as ignoring the role of emotion in social-information processing. Indeed, their subsequent research specifically emphasized the dynamic interplay of emotion and cognitive processes in the prediction of aggression during development (Arsenio 2010). Little research has been conducted to integrate the role of emotion in SIP processes; however, findings mostly support that both emotional state and emotion regulation abilities contribute to aggressive outcomes in children (Li et al., 2013). Whilst this model has been useful in its additional consideration of the role of emotion in processing, the vast majority of research investigating cognitive biases in relation to aggression has drawn on Crick and Dodge's (1994) model. The latter has gradually been developed to account for aggression subtypes and individual differences in externalizing problems, and has been widely acknowledged as providing a structured framework for the study of aggression, one with has extensive potential for targeted interventions.

1.2.5. The SIP model of aggression

Crick and Dodge (1994) posited that 5 overlapping stages in social-information processing led to behavioural responses, with deviant patterns of processing increasing the probability of an aggressive outcome (see Figure 1). These steps have been reformulated and refined (Fontaine & Dodge, 2006) as further relationships were identified between cognitive biases and measures of antisocial and aggressive traits and behaviours (Fontaine et al., 2010). These relationships are summarised below for each proposed stage in the SIP model:

Figure 1

An integrated model of emotion and cognition in Social-Information Processing based on Crick and Dodge's SIP model (1994), the Model of Response Evaluation and Decision (Fontaine & Dodge, 2006), and the Ecological Model of Hostile Attribution Bias (Dodge et al., 2015)



1. The first step comprises ***encoding*** of situational cues following attention allocation to relevant information in the environment. Socially competent responders were expected to attend to and encode cues relevant to the given social scenario, and process contextualized information effectively. In contrast, individuals prone to aggression and intense anger were expected to exhibit hypervigilance to threatening stimuli (Dodge et al., 2013). An overwhelming amount of information is available during the course of a social interaction; therefore, the encoding stage needs to be "selective, fast, and automatic" (Horsley et al., 2010, p. 588). The attentional bias literature has consistently linked increased bias toward hostility related information with increased aggression (Smeijers et al., 2020). However, attentional biases have been linked to the spectrum of externalizing psychopathology and antisocial traits such as psychopathy (Kimonis et al., 2006, Domes et al., 2013) and CU traits (Kimonis et al., 2012, Ciucci et al., 2018). The relationships between attentional bias, aggression, and antisocial traits and behaviours are discussed in a later chapter. Encoding is a key step in the SIP model and has increasingly become a subject of debate due to the variation in measurement (e.g., from vignettes to eye-tracking methodologies) and is often being mistaken for the second stage in the SIP model, i.e. attribution of intent (Horsley et al., 2010).

2. The second stage consists of ***interpretation of cues***. At this stage, individuals were expected to attribute intent and causality to inter-personal cues previously encoded during the social interaction, interpreting their relevance to personal well-being and safety (Fontaine et al., 2010). In line with tenets of the SIP model, aggressive individuals have been shown to interpret socially ambiguous cues and stimuli as hostile and harmful, a tendency which has been termed a *hostile attribution bias* (Fontaine, 2010). Increased hostile attribution has been linked to a history of antisocial and aggressive behaviour (Schönenberg & Jusyte, 2014), including severe ASB (Dodge et al., 1990; Wu et al., 2014). Hostile attribution biases have been shown to mediate the relationship between maltreatment and aggression (Zhu et al, 2020), and have been linked to anger rumination (Quan et al., 2019a), violent beliefs (Quan et al., 2019b), reactive and proactive aggression (Helfritz-Sinville & Stanford, 2015), psychopathy (Carroll et al., 2021) and narcissism (Law & Falkenbach, 2018).

3. Choice and clarification of goals follows attribution of intent. The SIP model posited that competent responders should be able to assess potential outcomes and generate appropriate, benign goals for the given social interactions. Conversely, aggressive individuals were believed to increasingly expect potentially negative outcomes particularly in relation to ambiguous social situations and consequently to formulate dominant or retributive goals (Dodge et al., 2013; Erdley & Asher 1996). Endorsement of antisocial or aggressive social goals has been linked to deficient social skills and aggressive behaviour (Crowe & Wilkowsky, 2013; Harper et al. ,2010; Wilkowski & Robinson, 2010). Further, sustained endorsement of dominant and retributive goals throughout development has been associated with escalation in aggressive and ASB (McDonald & Lochman, 2012).

4. The next step in the SIP model, ***response construction or selection*** consists of either generation of a response based on the available information or access to previously constructed responses based on similarities in the social context (Fontaine et al., 2010). Whilst competent responders were expected to generate appropriate, benign behaviours during the course of social interactions, individuals with a history of aggression were expected to construct aggressive responses (Dodge et al., 2013). The relationship between generation of aggressive responses and antisocial traits and behaviours appears inconsistent during childhood and stabilizes during adolescence (Fontaine et al., 2009). Aggressive response generation has also been associated with increased attribution of hostile intent (Fontaine et al., 2010).

5. Response decision is the final stage in the SIP model. During this stage, competent responders were expected to evaluate the positive and negative consequences of the previously constructed responses and choose the response most likely to result in achievement of prosocial goals (Fontaine et al., 2010). Conversely, aggressive responders would either act impulsively, disregarding potentially negative consequences, or endorse beliefs that enacting an aggressive behavioural response would likely result in a favourable outcome (Verhoef et al., 2021).

Cognitive processes within this stage have been summarised under the response evaluation and decision sub-model (Fontaine & Dodge, 2006), according to which response decision is a function of the perceived efficacy and value of the generated responses. Following the assessment of previously constructed responses, the most favourable response would be selected and enacted (Fontaine & Dodge, 2006). Deficiencies in response selection within the response evaluation and decision stage have been linked to ASB (Fontaine et al., 2009, Fontaine et al., 2010); however, due to refinement of the cognitive skills necessary for effective decision-making strategies this link becomes more apparent during adolescence (Fontaine et al., 2008).

The Crick and Dodge (1994) SIP model has become an influential framework, allowing the study of underlying cognitive processes leading to aggression (Teglasi et al., 2022). Numerous studies have highlighted links between cognitive biases and aggression, specifically between increased aggression and preferential encoding of hostility related stimuli (Manning, 2020), increased attribution of hostile intent (Bailey & Ostrov, 2008), preferential selection of instrumental over benign goals, and generation of fewer prosocial responses, expecting positive outcomes from aggressive behavioural outcomes (Dodge et al., 1990). Longitudinal research has shown that patterns of aggression stabilise throughout development and into adulthood (Boxer et al., 2005; Guerra & Slaby, 1990) with Eme (2009) describing life-course-persistent ASB as "the single most important child clinical problem" (p.349). Endorsement of aggressive beliefs and aggressive behaviours as likely to result in positive outcomes appear to be malleable during childhood but become increasingly more stable during adolescence and adulthood (Butler et al., 2015; Moffitt, 1993).

Although the SIP model has initially been applied to the study of aggression during development (Butler et al., 2015), the framework has also been applied in studies of aggression in adult populations (Bailey & Ostrov, 2008, Tuente et al., 2019). According to the latter, adults with life-course persistent aggression and violence problems have been shown to accumulate and draw on limited and rehearsed behavioural repertoires and scripts (Gilbert & Daffern, 2017). Moreover, increased hostile attribution has been associated with a history of aggression in individuals with intermittent explosive disorder (Coccaro et al., 2016) and positive attitudes toward violence (Quan et al., 2019a, Quan et al., 2019b). Further studies have also shown that, on a similar note to aggressive children and adolescents, adults with a history of aggression demonstrated cognitive and emotional processing biases such as attentional biases toward hostile words, increased hostile attribution bias, and a tendency to expect aggressive outcomes to ambiguous social situations (Coccaro et al., 2016; Coccaro et al., 2017).

Recent work under the SIP framework has also increasingly distinguished between aggression subtypes and distinct patterns of processing. The reactive/proactive aggression dichotomy was based on two types of behavioural aggregates falling under the aggression spectrum (Thomson & Centifanti, 2018). Proactive aggression (i.e., cold, premeditated, goal-directed behaviour) has been associated with achieving instrumental goals (Li et al., 2013). Within the response evaluation and decision stage of the SIP model, proactive aggression is expected to be associated with a careful assessment of goals, behavioural options, and likely outcomes (Fontaine & Dodge, 2006). Comparatively, reactive aggression was associated with increased attribution of hostile intent, interpretation of ambiguous social stimuli as hostile, and angry emotional responses that are typically elicited by provocation (Fontaine & Dodge, 2006). Reactive aggression has also been associated with an abbreviated script search, streamlining the generation of aggressive responses and increased endorsement of aggressive goals (McDonald & Lochman 2012).

1.2.6. The role of emotion in SIP

The impact of emotion on cognition is two-fold: individuals' own emotional states impact how information is processed and emotional stimuli are preferentially attended to and encoded (Ribeiro et al., 2019). The SIP model has been critiqued for failing to consider the role of emotion (Arsenio & Lemerise, 2004) and providing an overly parsimonious explanation of aggression by emphasizing social deficiencies in aggressive individuals. In their response to these critiques, Dodge and Rabiner (2004) proposed that emotions are fundamental to the processing of social information and therefore permeated all stages of the model. Further, Dodge and Rabiner (2004) claimed the SIP model was "meant to be entirely emotional" (p.1006) with emotion "an adjectival descriptor of mental operations" (p.1006). Studies have since shown that emotional states and arousal can impact social-information processing (Hatfield & Dula, 2014, Stavraki et al., 2021). A positive relationship has been found between hostile attribution of intent and trait anger (Li et al., 2013) with both predicting aggressive behaviour (Bondü & Richter, 2016, Gilbert et al., 2013). Further, induced anger has been shown to narrow the cognitive scope resulting in attentional biases toward hostile cues and hostile interpretations of ambiguous situations (Veenstra et al., 2018, Maoz et al., 2017). In turn, increased hostile attribution has been linked to biases in processing others' emotions (Choe et al., 2013, Satmarean et al., 2022). The relationship between anger and aggression has been explored in other cognitive models of anger. One such model is the integrative cognitive model of trait anger and reactive aggression, which posits that individuals displaying higher trait anger display attention and hostile attribution biases (Wilkowski et al., 2010), an assumption shared by the SIP model.

Anger is a frequently experienced human emotion; a propensity to experiencing anger intensely can become maladaptive and has been associated with aggression (Wilkowski & Robinson, 2008). Intense anger can often escalate to aggression and has often been associated with reactive aggression following provocation (Denson et al., 2011). Proneness to anger has been linked to cognitive biases during encoding and response decision stages of the SIP model (Davies et al., 2020) and to increased hostile attribution bias (Cougle et al., 2017). Using an anger induction task, Harper et al. (2010) showed that increased anger resulted in increased hostile attribution and increased endorsement of hostile goals in aggressive children. Similarly, inducing anger in aggressive individuals resulted in an increase in aggression (Chester & DeWall, 2017) and hostile attribution (Reidy et al., 2010).

Higher trait anger has been reliably associated with early cognitive biases, namely, with increased attentional biases to threat (see Owen, 2011 for a review). Studies using emotional Stroop tasks showed greater interference by angry facial expressions in individuals displaying higher revenge-planning (Crowe & Wilkowski, 2013). Individuals high in trait anger (Cohen et al., 1998) and offenders with a history of violence (Smith & Waterman, 2004) also showed increased difficulties ignoring hostile distractors during visual search. Maoz et al. (2017) found an increased attentional bias toward angry faces was associated with emotion dysregulation, that is, with the ability to control anger expression. The relationship between anger and attentional bias is not fully understood although it has been proposed that the link may reflect dominant goals, with individuals displaying higher trait anger being more likely to attend to hostile cues in their social environment due to these being perceived as a challenge (Van Honk & Schutter, 2007). This concept is supported by studies linking induced anger to attentional bias to rewarding information (Ford et al., 2010).

Anger has also been associated with increased attribution biases (Veenstra et al., 2018, Wang et al., 2018). Hazebroek et al. (2001) found that when watching videos depicting negative and ambiguous outcomes, individuals high in anger displayed an increased tendency to identify the person in the video as an antagonist, assigned more blame, and believed negative events to be more relevant to their interest compared to individuals with low trait anger.

Similarly, higher trait anger has been associated with an increased tendency to attribute hostile intent in ambiguous situations, even compared to anxious individuals (Wenzel & Lystad, 2005). This relationship extends to processing ambiguous facial expressions, with increased aggression and trait anger being linked to an increased tendency to mislabel ambiguous faces as hostile (Mellentin et al., 2015). The association between anger and biases in the SIP model appears bidirectional (Wang et al., 2018), as cognitive bias modification interventions have been shown to lead to lower self-reported trait anger (Cougle et al., 2017). Indeed, Tuente et al. (2019) proposed that "hostility is the cognitive component of aggressive behaviour whereas anger is an emotional component" (p.67).

Studies have shown that individuals higher in trait anger are also prone to experiencing state anger more intensely (Gilbert et al., 2013). Experimentally induced state anger has been associated with selective attention to rewarding information (Ford et al., 2010), increased focus on instrumental goals (Lench et al., 2015), and increased dominance and status-seeking (Cabral et al., 2019). It has also been shown to lead to increased attribution of hostile intent (de Castro et al., 2003), with attention believed to act as a bridge between anger and attribution biases (Peng et al., 2015). Induced anger also resulted in increased processing of complex social scenarios (Semmler & Hurst, 2017, Moons & Mackie, 2007), suggesting an increased attentional spotlight toward salient information, termed "the angry spotlight" (Ford et al., 2010). This spotlight appears to narrow attentional focus, increasing accuracy at the cost of processing speed (Grubert et al., 2012). That is, studies have found slower RTs to processing of emotional information in induced anger (Bertsch et al., 2009) but increased precision (Young et al., 2011).

With anger increasing attention toward goals (Lench et al., 2015), and individuals with a history of aggression displaying increased endorsement of aggressive goals and beliefs (Butler et al., 2014), the association between aggression and anger could easily become maladaptive. This is particularly the case in individuals who are less able to regulate their emotional state and its effects on behaviour. Emotion dysregulation, i.e. impaired emotional awareness and ability to pursue goals and constrain impulsive responses when distressed, has been linked to increased aggression (Preson & Anestis, 2019), especially in individuals prone to anger (Mancke et al., 2017) and displaying other antisocial traits such as psychopathy (Garofalo et al., 2020). Emotion dysregulation has been associated with increased hostile attribution (Gagnon et al., 2015) although some studies suggest this association requires increased state anger rather than trait anger (Calvete & Orue, 2012). Most recently, the interaction between increased emotion dysregulation and aggressive script rehearsal predicted a significant improvement in the prediction of aggression (Hosie et al., 2021). The link between anger, emotion dysregulation and aggression is intuitively clear. It should however be noted that given the dimensionality of emotion dysregulation mechanisms, that is, from difficulties in impulse control to inability to engage in goal directed behaviours effectively, emotion dysregulation is likely linked to a spectrum of aggressive behaviours (Garofalo et al., 2020).

1.2.7. The role of memory in SIP

Alongside emotion, memory mechanisms are considered fundamental to the SIP model. Latent knowledge structures such as aggressive schemas, scripts, and internal representations (or blueprints for interaction) have been linked to biases in social information-processing (Hosie et al., 2021). In line with Bowlby's attachment theory and learning theories, knowledge structures are believed to comprise working models (Fontaine & Dodge, 2006) and scripts which become internalized throughout development, through experiencing and witnessing of aggression and violence in social contexts (Dodge et al., 1990). Under social-cognitive paradigms, these structures are believed to be central to information processing (Huesmann, 1988). Specifically, social scripts are believed to be retrieved, enacted, and reinforced following outcomes to ambiguous social situations (Ferguson & Dyck, 2012). They are stored in memory and used to guide behaviour by making predictions about events and other people's likely behaviour in a given social context (Gilbert & Daffern, 2017). Violent individuals appear to have an increased and easily accessible repertoire of aggressive scripts (Smith & Waterman, 2004). An investigation of aggressive cognitions in individuals with a history of violence showed that almost half of participants (44%) engaged in aggressive script rehearsal on a daily basis (Meloy et al., 2001).

Knowledge structures held in memory are considered distal influences of behaviours and cognitive mechanisms by which their influence is exerted on immediate behaviour are not well understood. Correlational studies have found relationships between aggressive schemas and physical assault, verbal threats, and hostility (Loper, 2003). Ward (2000) argued that schemas or implicit belief systems ordered around a dominant theme, actively guide perception and interpretation and have a direct effect on behaviour. This guidance can become maladaptive when guidance draws upon aggressive beliefs or schemas. Given the high complexity and variability of social interactions, studies have found that more than one maladaptive schema can be activated at one time (Lobbestael et al., 2007). This suggests that although knowledge structures can be associated with aggression and violence, the enactment of a schema into behaviour is not necessarily a linear process but rather one with multiple sources of input which can either amplify or reduce their influence.

Recent studies suggested that emotion is a key component in the activation of aggressive schemas. Van Wijk et al. (2020) showed that effects of early maladaptive schemas on aggression were mediated by anger. Schemas associated with intense anger, rage, and impulsivity have been shown to dominate social cognitions in an aggressive offender sample (Dunne et al., 2018). This was believed to be due to anger restricting cognitive processing to access rehearsed and frequently used aggressive repertoires but also mediated by hostile attribution biases and difficulties with impulse control (Anderson & Bushman, 2002). In a recent study investigating relationships between emotion regulation and aggression in male offenders, Hosie et al. (2021) found that frequency of aggressive script rehearsal was significantly correlated to emotion regulation difficulties. Further, they found that prediction of aggression was significantly improved when considering interactions between aggressive script rehearsal and lack of emotional clarity, and between script rehearsal and emotional awareness. These findings were in line with previous studies highlighting an association between anger and aggression (Anderson & Bushman, 2002), providing further support for the concept of emotion permeating cognitive networks underlying aggression.

1.3. Attentional bias

1.3.1. Introduction to attentional biases

Examination of cognitive biases under a social-information processing paradigm is aimed at explaining how cognitive processes guide social interactions. In particular, distinguishing between relevant and distractor information is fundamental to effective interaction within a social environment (Mallett et al., 2020). A mixed pattern of empirical research into aggression emerging in the 1950s suggested somewhat contradictorily that people could both focus their attention on a given task and have their attentional focus disrupted by distractor stimuli. In other words, people are able to selectively focus their attention on an oncoming car to determine whether it is safe to cross a street. At the same time, they would be filtering distractors such as noise, lights, and other people. The processing of emotional stimuli is an additional layer of complexity to selective attention for social information (Gong & Smart, 2020). Given that abnormal processing of emotional information has been linked to aggressive and antisocial behaviours and traits (Manning, 2020), selective attention to emotional stimuli is an important mechanism in social information processing.

Attention and WM mechanisms prioritising processing emotional stimuli have been the focus of extensive research (Grecucci et al., 2010, Xie et al., 2017). Whilst attention focus is not required to process emotional stimuli, it is necessary for discriminating emotional content (Vuilleumier et al., 2001). Negative stimuli are particularly salient and have been shown to automatically capture and hold the focus of attention (Dickins & Lipp, 2014). Under the SIP framework, individuals with a history of aggression are expected to display an enhanced bias toward negative stimuli, that is, to disproportionately attend to and encode hostile stimuli (Crick & Dodge, 1994). Such attentional biases reflect preferential allocation of attention toward hostility-related information relative to baseline stimuli (Moriya et al., 2014). They can be observed during attentional shifts toward or away from stimuli, engagement and disengagement from targets or distractors (Cisler & Koster, 2010). Facilitated engagement has been observed when hostility-related stimuli are preferentially processed, whereas attentional bias during disengagement is indicated by difficulties in reallocating attention to new information and away from hostility-related information, once this has been engaged with (Cisler & Koster, 2010, Dong et al., 2017). Finally, avoidance refers to preferential allocation of attention to stimuli other than hostility-related stimuli (Bardeen et al., 2017). The terms "negative attentional bias", "hostile-, anger-, aggression-related attentional bias" are used interchangeably in research; however, they all indicate an attentional bias toward aversive or hostility-related stimuli.

1.3.2. Theories of attentional bias

It is impossible to process the overwhelming amount of information available in our environment. Selective processing such as detection of threat has therefore been deemed fundamental for survival (van Rooijen et al., 2017). Indeed, attentional biases toward threatening stimuli have been demonstrated in infants (see Fu & Pérez-Edgar, 2019 for a review) and throughout development (Jenness et al., 2021). However, hypervigilance to threat can be maladaptive when attention is consistently allocated toward hostile information at the expense of processing potentially relevant information in social environments. Studies have shown that individuals displaying attentional biases toward threat also displayed a tendency to interpret ambiguous social information as potentially hostile (Mellentin et al., 2015). Such biases have been linked to psychopathology such as anxiety (Bar-Haim et al., 2007), depression (Jenness et al., 2021), and ASPD (Schönenberg et al., 2012), increased anger and aggression (Smeijers et al., 2017).

Beyond the evolutionary advantage of preferential processing for threat, recent theories of attentional bias draw on information theory, computational principles, and Bayesian formulations (Friston & Kiebel, 2009) to frame attention-related questions in terms such as uncertainty, surprise, and precision (Feldman & Friston, 2010). In other words, the focus has shifted toward evaluation of attention as "more concerned with optimizing the uncertainty or precision of probabilistic representations, rather than what is being represented" (Feldman & Friston, 2010, p. 1). Conceptualizing attention and perception in such terms facilitates the study of attention mechanisms in larger physiological and psychological contexts and is applicable to the study of top-down (Fiebelkorn & Kastner, 2019) and bottom-up mechanisms of selective attention (Beck & Kastner, 2009).

1.3.3. Measurements of attentional bias

Attentional bias to threat has most commonly been investigated in relation to anxiety disorders (Cisler & Koster, 2010) although attentional bias paradigms have been applied to the study of aggression, and antisocial traits and behaviours (Mellentin et al., 2015). Attentional biases are typically measured using behavioural measures, most often reaction times (RT) to stimuli in cognitive tasks such as the dot-probe task (DPT) or the emotional Stroop task (van

Rooijen et al., 2017). In the emotional Stroop task, participants are shown negatively and positively valenced words in different colours and asked to identify the colours. Longer latencies are typically expected when naming the colour of negatively valenced words due to difficulties in disengaging from the emotional content of the word even though this is irrelevant to the task. Although this has been widely deployed in clinical and non-clinical populations, recent studies highlight potential validity concerns (Rodebaugh et al., 2016). This task has also come under criticism for failing to differentiate between an attentional bias to emotional words or a delay in motor response (Yiend, 2010).

In the DPT, a pair of stimuli (usually emotional vs neutral) are presented simultaneously. One of the stimuli is replaced by a probe and an attentional bias is indicated by faster responses to the probe replacing hostility-related compared to neutral stimuli. One of the strengths of this task is that it allows the study of attentional selectivity for a wide range of emotional stimuli, including naturalistic faces (Cooper & Langton, 2006). The DPT has been used extensively in the study of attentional biases to angry faces relating to anxiety and more recently aggression (Yiend, 2010). However, similar to the emotional Stroop task, the validity of the DPT has come under scrutiny in the last decade due to the low reliability of traditional bias score (BS) computations (Schmukle, 2005). Further, EEG studies using the DPT have found neural activity indicative of attentional bias toward hostility-related information but no behavioural indication of this bias (Thigpen et al., 2018, Chapman et al., 2017). Given the paucity of studies reporting the reliability of traditional BS, it has so far been difficult to determine whether the potentially low reliability of the task contributed to the inconsistency in findings in behavioural studies.

Visual search paradigms have also been used to examine attention allocation to emotional stimuli. In the last decade, studies have extended visual search paradigms to the examination of attentional selection in real-life environments. According to the first two tenets of Desimone and Duncan's (1995) biased competition model of attention, the visual system has a limited processing capacity, meaning that stimuli in our environment compete for representation; this competition was believed to be influenced by top-down goals. Although this model derived from research on single-cell neuronal recordings, research has since provided ample support for competition and top-down bias in the visual system (Carlisle et al., 2011, Carlisle, 2019). Attention focuses cognitive processing resources on salient information; however, the repetitive nature of the visual searches usually performed in real-life situations increases experience-driven efficiency (Peelen & Kastner, 2014). Similar to real-life visual searches, visual search paradigms allow the study of engagement with visual targets and of factors driving efficiency in visual search.

Although considerable advancements have been made in behavioural and neurophysiological studies of attention during visual searches, research has only recently extended this paradigm to naturalistic selection conditions (Peelen & Kastner, 2014). Whilst naturalistic visual searches differ considerably from previous research using mostly shapes and simple objects, they are surprisingly efficient (Peelen & Kastner, 2014). Studies investigating visual search for social stimuli such as faces have mostly used schematic representations (Huang et al., 2011); however, advancements have been made in this area of research as well, witha recent study by Burra et al. (2017) using naturalistic faces finding angry faces were preferentially attended to and encoded even when they were irrelevant to the task. Few studies have used visual search paradigms to examine visual biases in relation to aggression or antisocial behaviours and traits, and findings so far have been mixed. One such study used naturalistic faces as targets a visual search task and found no differences in attention allocation during visual search between a forensic-psychiatric sample and a non-clinical sample (Brugman et al., 2018). Evidence of bias toward aggressive (compared to neutral) words was however found in violent offenders (Smith & Waterman, 2004).

1.3.4. Attention guidance by WM

Attention plays an important role in encoding, retrieval, and restructuring of information in memory (Soto et al., 2012). Neural studies have shown that visual experience modifies how objects are represented in the visual cortex leading to distinct patterns of selectivity in single neurons (Gauthier & Logothetis, 2000). The third main tenet of Desimone and Duncan's (1995) biased competition model posited that competition should be integrated across systems. Bottom-up and top-down competitions for representation have typically been studied in isolation; however, the few studies integrating the two have shown evidence in favour of integration (Beck & Kastner, 2009). Influential theories of WM and attention propose that multiple items can be activated in WM; however, attention controls the activation of specific representations (Oberauer, 2009). WM can thus be conceptualised as the result of recruitment of multiple neural systems of attention (Postle, 2006). Indeed, studies have shown that neural resources of visual WM and attention overlap (D'Esposito & Postle, 2015).

Spatial attention has been shown to be oriented toward internal representations (Kiyonaga & Egner 2013) even in the absence of conscious awareness (Pan et al., 2014). That is, WM guides attention toward stimuli in the environment matching the internal representations (Soto et al., 2008). This guidance appears to be mediated by task goals (Woodman et al., 2013) and WM capacity (Dowd et al., 2015), although the latter is not increased by lowering the quality of WM representations (Zhang & Luck, 2011). Importantly, WM only biases visual perception when information held in WM is perceived as important to the task (Bahle et al., 2018).

Most studies have examined feature-based guidance of attention by WM (see for example Soto et al., 2007, Soto et al., 2012); however, recent experiments showed this is different to object-based guidance. Multiple features can guide attention but only one object-based template can guide attention at any given time (Berggren & Eimer, 2018). WM biases of attention (i.e., greater attentional capture of information matching WM templates) have been predicted by impairments in attentional control mechanisms in autism and attention deficit hyperactivity disorder (Geurts et al., 2004, Manly et al., 2001). Most recently, WM biases of visual attention to angry faces have been found in individuals with higher anxiety (Yao et al., 2019). However, WM guidance of attention has not been studied in relation to aggression or ASB.

1.3.5. Attentional biases: links to aggression, antisocial traits and behaviours

Biased processing of facial emotions (i.e., disproportionate allocation of attention toward or away from social threat) has been reliably associated with internalizing disorders such as depression and anxiety (Harms et al., 2010). Comparatively few studies have examined attentional bias in relation to anger, aggression, or ASB (Kuin et al. 2017). Further, it is unclear whether biased processing of emotional faces is linked to later biases in social information-processing and behavioural outcomes. A systematic review (Mellentin et al. 2015) showed that individuals prone to aggression and anger also showed increased attentional biases toward angry faces across attentional bias paradigms. Additionally, participants with higher aggression (Smeijers et al., 2017) and psychopathic traits (Newman et al., 2010) displayed deficits in emotion perception, suggesting that antisocial traits may predict different types of visual processing biases. For instance, CU traits have been associated with psychopathy and persistent and severe patterns of ASB and violence (Frick & White, 2008) and have been used as proxy measures of psychopathy in non-clinical samples (Kimonis et al., 2012, Strickland et al., 2015). CU traits have also been associated with differential patterns of emotion recognition, especially for fearful faces (Ciucci et al., 2018). Individuals with ASPD have shown impaired face detection and categorisation (Lijffijt et al. 2012). More specifically, individuals with overlapping psychopathy and ASPD showed impaired affect recognition compared to individuals with ASPD and controls with a history of violence but no diagnosis (Kosson et al., 2006).

Psychopathic traits have also been linked to dysfunctional processing of negative emotional words, facial expressions (Hastings et al., 2008), and enhanced distractor suppression (Carolan et al., 2020). Similarly, higher CU traits predicted decreased recognition accuracy for fearful facial expressions (Brislin et al., 2018). Domes et al. (2013) examined attentional bias for violence-related words using an Emotional Stroop Task in a sample of offenders with ASPD. They found that offenders with ASPD showed an attentional bias toward violence-related words; however, within this sample, offenders with a history of childhood maltreatment showed a significantly higher violence-related attentional bias compared to offenders without a history of maltreatment. These findings underscore the importance of considering conceptual and measurement overlap in aggression and ASB research. Further these studies suggest that whilst attentional biases are associated with the spectrum of antisocial traits and behaviours, there are fine differences based on individual differences and experiences linked to specific cognitive biases toward hostile stimuli.

1.3.6. Attentional biases in the SIP

Attentional bias toward faces, particularly toward emotions such as fear or anger is an intrinsically social phenomenon; however, it is seldom investigated using a social information-processing paradigm (Wilkowski & Robinson, 2008). Although within the SIP model, aggression is expected to be associated with attentional bias to threat during encoding of social-processing information, this step is not well understood. This may be due to a wide range of tools used to measure social-information processing patterns (ranging from eye-tracking methodologies to self-report) which has resulted in inconsistent findings (Horsley et al., 2010). Social-information processing is typically investigated using vignettes (Fraser et al. 2005.

These, albeit attempting to emulate real-life ambiguous social situations, are likely to be processed differently and elicit a much lower emotional response (Li et al. 2013). Consequently, self-reported behavioural outcomes may be different to responses enacted in real-life situations. Moreover, there is considerable variation across studies in measurements of the different stages in the SIP model (Baurain & Nader-Grosbois, 2013; Fontaine et al., 2010) which has led to difficulties in identifying patterns of bias and links to antisocial traits or aggression.

The SIP model postulated that the most salient cues are typically attended to and encoded during social interactions (Crick & Dodge, 1994); it was therefore expected within this framework that aggressive children would preferentially attend to hostility-related and that this initial bias would increase the probability of an aggressive response being constructed and enacted. Given the vast amount of information in a social environment, this initial step was believed to be automatic and selective therefore not adequately measured through self-report measures (Coccaro et al., 2017). Using eye-tracking and drawings of social interactions, Horsley et al. (2010) measured encoding in relation to later stages of the SIP in an aggressive sample. They found that aggressive children did not attend more to hostile cues compared to controls, but showed increased attention toward non-hostile cues and a higher hostile attribution bias, although they recalled little of the non-hostile information presented. The authors concluded this provided evidence of top-down guidance of attention by aggressive internal representation toward schema-incongruent information. It is however also likely that aggressive children spent more time looking for potentially hostile cues in non-hostile drawings. These findings suggest that measurement of encoding using vignettes (Dodge et al., 2003) or sets of pictures measuring encoding of hostile social stimuli (van Goozen et al., 2002) can provide ambiguous data. The findings may be confounded by memory, motor control, and ability to provide verbal descriptions. Comparatively, RT based measures of attentional bias during encoding could be used to indicate attentional bias toward threat, difficulties disengaging from hostile cues, and attentional bias away from threat (van Goozen et al., 2002, Wilkowski et al., 2007).

1.4. Attribution bias

1.4.1. Introduction to attribution bias

Cognitive processes such as expectations, interpretation, and appraisal of cues have been shown to mediate the relationship between negative stimuli and aggressive outcomes (Gasse et al., 2020). These processes are believed to be guided by aggressive schemas and beliefs which are acquired early and reinforced through exposure during development. Social learning is believed to be an important mechanism in the development of aggression, with negative internal representations forming blueprints for social interaction which are drawn upon in ambiguous social situations (Blair, 2001), using past experience to guide interpretation of events and construct responses (Dodge et al., 1990). Aggressive schemas have been linked to violent offending (Milner & Webster, 2005) and ASB (Pasion et al., 2018). A history of ASB has also been linked to biased representations of self and others such as blame denial, justification, blaming others, attributions of hostile intent and rationalization of aggression (Chambers et al., 2008, Chereji et al., 2012). Methodologies used to investigate cognitive biases in social-information processes following encoding have been fairly consistent across the literature: participants are presented with vignettes describing ambiguous provocation scenarios (e.g., finding an abandoned backpack, picking it up as an ambiguous provocateur neutrally states that the backpack belongs to them, Fontaine et al., 2010). Participants are typically required to immerse themselves in the scripts and then answer questions designed to evaluate hostile attribution styles and response evaluation and decision processes (Dodge et al., 2015). Attribution and interpretation biases, referring to the tendency to attribute hostile intent in ambiguous social situations, have been used interchangeably in research (Hawkins & Cougle, 2013); this tendency is referred to as attribution bias throughout this research.

1.4.2. Attribution bias in the SIP model

Dodge (1980) found that aggressive children displayed a tendency to attribute hostile intent in ambiguous social situations. This attribution bias has been linked to various psychopathologies, including depression (Gasse et al., 2020); however, it was also believed to be instrumental to the development and maintenance of aggression (Chen et al., 2012). Whilst benign interpretations of ambiguous social information have been shown to lead to de-escalation of conflict, hostile attributions have been linked to increased likelihood of conflict (Dodge et al., 2015). Further, an increased hostile attribution bias has been identified as a risk factor in the aetiology and maintenance of aggressive (Coccaro et al., 2017) and ASB (Chereji et al., 2012) and linked to violent behaviour (James & Seager, 2006). Given the extensive relationships between hostile attribution biases and externalizing problems, the SIP model has been revised to embed "behavior in a widening circle of ecological contexts" (Dodge et al., 2015, p. 9311). Ecological contexts, i.e. family, community, culture, are believed to exert an influence on attention and attribution processes potentially leading to increased aggression.

1.4.3. Attribution bias and aggression

Attentional biases to threat, hostile attribution biases, and deviant response evaluation and decision processes, are all expected to contribute to the likelihood of engaging in aggressive behavioural outcomes and reinforce existing aggressive schemas and beliefs (Fanning et al., 2014; Fontaine et al., 2008). With few exceptions, attentional biases have usually been studied separately to attribution and response biases (Horsley et al., 2010). For example, Sadek et al. (2021) found that increased attentional bias to negative stimuli was linked to aggression-supportive cognitions. However, studies have identified relationships between later stages of the SIP; for instance, deviant response evaluation and decision processes have been shown to mediate the relationship between hostile attribution bias during the interpretation stage and ASB (Fontaine et al., 2010). These findings suggest that investigations of cognitive biases would benefit from the application of a unifying framework such as the SIP.

Attribution biases occur relatively early in the social-information processing timeline (MacLeod & Holmes, 2012). Increased hostile attribution biases are believed to be predicted by higher accessibility to hostile beliefs, scripts and schemas (Wilkowski & Robinson, 2008). It has therefore been proposed within the SIP model that an increased hostile attribution bias has a dual role, one in generation of aggressive responses and another in reinforcement of pre-existing aggressive internal representations stored in memory (Dodge et al., 2003). This relationship has been found in correlational designs but also in studies examining the effect of cognitive bias modification procedures on attribution biases (AlMoghrabi et al., 2018). Attribution bias training has resulted in reduced hostile attribution toward faces (Hiemstra et al., 2019) and ambiguous social scenarios (Hawkins & Cougle, 2013). Although hostile attribution bias early in social-information processing has been linked to aggression and antisocial traits and behaviours (Arad & Bar-Haim, 2021), this relationship may be mediated by later stages in the SIP model such as response evaluation and decision processes (Fontaine et al., 2010). The SIP model originally posited that cognitive biases contributed individually to predictions of aggression and ASB (Crick & Dodge, 1994, Crick & Dodge, 1996); however, it is possible biases in social-information processing have a cumulative effect on aggressive outcomes.

1.5. Context for empirical work

Social-cognitive theories of aggression have focused on cognitive structures and their proximal and distal effects of behaviour. However, research so far has examined the relationships between cognitive distortions and aggressive behavioural outcomes by focusing on individual associations rather than how they would interact in real-world, complex social interactions. This approach, as well as variation in task designs and methodologies used to investigate links between cognitive biases and aggression, has resulted in a disjointed areas of focus in the aggression literature, and lack of understanding as to how social-information processes and biases interact or whether they compound to predict aggressive outcomes. The conceptual separation of SIP mechanisms and their effects on behaviour has also resulted in snapshots- like findings which are unlikely to be representative of real-world interactions and therefore poor models of aggression as a social behaviour. Taking a systemic approach to the study of the relationships between cognitive biases and antisocial traits and behaviours is expected to build on these findings as well as highlight the complexity of relationships between antisocial traits and cognitive biases at different stages in the SIP.

Further, given the considerable overlap in measurements of aggression and antisocial traits and behaviours, the specificity of these biases to aggression warrants further investigation. To date, a robust association has been found between aggression and attentional biases to hostility-related stimuli. Aggression has also been linked to later stages in social-information processing such as increased hostile attribution biases and availability of aggressive response schemas. This thesis will therefore investigate the distinct relationships between antisocial traits and behaviours and cognitive biases.

The SIP model posited that stages in social-information processing are influenced by emotion and core knowledge structures and a few recent studies have provided initial evidence of the influence of emotion on processing of social information. Such studies have highlighted that processing is impacted by the valence of emotional states, proneness to experience certain emotional states more intensely (e.g., higher proneness to experiencing intense anger), and ability to regulate such emotional states to achieve desired goals in a social interaction. These relationships will be examined in this thesis, both in terms of their individual contributions to the development of aggression and of their role as mediators between stages of the SIP and aggression.

Recent studies have questioned the validity and reliability of behavioural studies using RT based methodologies. Particularly regarding the DPT, it is unclear whether the spuriousness of the associations between attentional bias to threat and aggression or antisocial traits such as psychopathy and CU traits is due to variation in task designs (i.e., stimulus presentation duration, stimulus type: word vs faces) or low validity of traditional BS computations. Given that the DPT is considered the gold standard in attentional bias research (Kappenman et al., 2014), this thesis will investigate reliability concerns, using novel computational techniques to improve the reliability of BS computations. Links of novel bias computations to measures of antisocial traits and behaviours and stages in the SIP model will also be examined.

Theories and models of social cognition propose unitary frameworks of guidance of behaviour through cognitive bias mechanisms. This research review has highlighted that although research on cognitive biases is disconnected, the relationship between cognitive distortions and aggression is present across areas of research. There is considerable evidence suggesting that behaviour follows predictions of likely outcomes and that such predictions are systemic, shaped by experience and guided through top-down biases. Moreover, multiple cognitive systems appear to be drawn upon to form these predictions and reinforce them through mechanisms of attention and WM. Finally, as posited in the SIP model, these processes are shaped by emotion. The literature suggests a holistic approach, one which refers to a complete system of social-information processing, is instrumental to understanding the role of cognitive biases in generation and enactment of aggression. This thesis will therefore investigate relationships between cognitive biases, emotion, and antisocial traits and behaviours. The focus of the research is to unify this enquiry under a social-information processing framework.

1.6. Thesis aims and outline of studies

The overall aim of this thesis is to examine cognitive biases in social-information processing and how these relate to aggression and individual differences in antisocial traits and behaviours. Three empirical studies will be conducted with a specific focus on early (attention) and late (interpretation) biases and links to emotion and ASB. The first study will examine the validity and reliability of the DPT and whether increased reliability of BS computation indicates an attentional bias to threat in response to brief compared to prolonged exposure to hostile faces. This study will also examine links of bias to later stages in the SIP as well as measurements of aggression, CU traits, and ASB. The second study will investigate whether WM guidance of visual attention (i.e., the extent to which emotionally valenced faces held in WM modulate visual search for emotional faces) links to later stages of the SIP model, proneness to anger, aggression, and CU traits. The third study will focus on a holistic approach to the study of biases in social-information processing by including an examination of the role of emotion across stages outlined in the SIP model. This study will examine the extent to which inducing an angry emotional state influences the accuracy of encoding emotional information in WM, whether state anger impacts WM modulation of visual perception, and the extent to which proneness to anger and ability to regulate emotional states predicted aggression through stages of the SIP model.

2. Chapter 2: Study 1- Attentional bias to threat: Links to antisocial traits and behaviours

2.1. Abstract

This study aimed to examine links between stages in social-information processing and antisocial traits and behaviours. Encoding was measured as visual attentional bias to angry faces using a DPT during which angry and neutral faces were presented simultaneously, for either 100 ms or 500 ms. The findings did not indicate an attentional bias toward angry compared to neutral faces for either presentation duration although participants were faster to identify probes replacing targets presented for 500 compared to 100 ms. Aggression and CU traits did not predict an attentional bias toward angry faces; however, an attentional bias was found in the low compared to high delinquency group and an attentional bias was predicted by self-reported likelihood of enacting an aggressive behavioural response. Further, aggression was linked to later stages of the SIP, including emotional response to ambiguous social situations. These results indicated that a history of antisocial behaviour was related to encoding whereas aggression was related to later stages in the SIP. Implications and directions for future research are discussed.

2.2. Introduction

As summarised in Chapter 1, ASB is a behavioural spectrum encompassing a wide range of behaviours such as lack of consideration for others, rule breaking, deviation from social norms whereas aggression is a behaviour enacted with the intention to cause harm (APA, 2021; Butler et al., 2015). Whilst conceptually distinct, these behaviours are often co-occurrent (Rudo-Hutt, 2015) and research has increasingly focused on cognitive biases (i.e. reliably distorted representations of some aspect of the world) as precursors of ASB and aggression (Chabrol et al., 2014, van Leeuwen et al., 2014). Drawing on Crick and Dodge's (1994) Social-

Information Processing (SIP) model, this study investigated the relationship between steps in information processing and measurements of ASB and aggression. These steps were expected to have a bidirectional functionality, that is, to reinforce or interfere with processing in other stages (Fontaine & Dodge 2006; Fontaine et al. 2010). This concept is supported by previous studies showing systemic changes following SIP training protocols, with significant improvements in problem solving and reasoning scores compared to controls as well as more accurate explanations for behaviours (Houssa et al., 2017) and lower attributions of hostile intent in ambiguous social situations (AlMoghrabi et al., 2018).

The link between biases in social-information processing and aggression has been replicated throughout development, in children (Jusyte & Schönenberg, 2017) and adult populations (Coccaro et al., 2016). Specifically, increased aggression has been associated with increased hostile attribution (Gagnon & Rochat, 2017), deviant response evaluation and decision processes (i.e., in favour of selection and enacting of aggressive outcomes, Coccaro et al., 2017). These findings support the socio-cognitive view that cognitive biases have a role in the development as well as maintenance of aggression (Martinelli et al., 2018). As the SIP framework is advantageous for intervention design, studies so far have targeted the association between aggression and hostile attribution bias through cognitive bias modification designs (Hawkins & Cougle, 2013, AlMoghrabi et al., 2018).

Early biases in social-information processing are less well understood due to inconsistencies in the conceptualization and measurement of the encoding stage of the SIP (Coccaro et al., 2017). The encoding stage is proposed to be fast and automatic (Horsley et al., 2010) and attentional biases to threat have indeed been linked to aggression and antisocial traits and behaviours (Baskin-Sommers et al., 2014). The DPT is the most frequently used paradigm in attentional bias research (Price et al., 2015) although studies have also used alternative tasks such as the emotional Stroop (Bertsch et al., 2009). The DPT has been used extensively across clinical and non-clinical populations (Wieser et al., 2018) and is suited to the study of social stimuli (here, facial expressions) as emotionally valenced stimuli are typically presented in pairs which enables the direct assessment of attentional preference (Crago et al., 2018). Most studies using the DPT to investigate attentional biases have used a stimulus presentation duration of 500 ms (Mathews & McLeod, 2005, Le Pelley, Nistal, & Luque, 2013); however, some studies have found that shorter presentation durations in the DPT yielded almost double the effect size compared to prolonged exposure (Bar-Haim et al., 2007), and others found that using the DPT, an attentional bias could be found for 100 ms but not for 500 ms (Cooper & Langton, 2006, Chapman et al., 2019).

A systematic review of attentional bias toward angry faces in aggressive and anger prone individuals showed that cognitive biases in this sub-population are not restricted to attention selection but indicate a broader pattern of deviant information processing whereby ambiguous information is perceived as hostile (Mellentin et al., 2015). These findings have been replicated in antisocial violent offenders (Schönenberg & Jusyte, 2014) and forensic psychiatric patients (Brugman et al., 2016). However, a recent systematic review of attentional bias to threat (Manning, 2020) suggested this relationship between aggression and attentional bias is not well understood, with the lack of clarity proposed to stem from inconsistent methodologies. For instance, some studies focused on studies measuring individuals' ability to distinguish facial expressions and how this related to aggression (Kuin et al., 2017). Others have investigated attentional bias as detection of and engagement with angry faces. The inconsistency in findings may be due to the tasks measuring different dimensions of attentional bias, underscoring the importance of using valid, established measurements. It is important to consider attentional biases in the wider context of social-information processing, where they may play an important role in reinforcing and perpetuating hostility-related knowledge structures, therefore influencing social-information processing (Doherty et al., 2017). A better understanding of cognitive biases is therefore fundamental to linking negative outcomes to specific cognitive mechanisms, enabling the development of targeted interventions.

This study aimed to examine information processing patterns and identify cognitive biases associated with aggression and ASB under the SIP framework in an adult sample. The DPT was used to examine the first step of the model, i.e. encoding by investigating patterns of attention allocation to stimuli presented either briefly (100 ms) or for a prolonged interval of time (500 ms). An attentional bias was expected toward briefly presented (100 ms) angry faces. Higher reactive aggression was expected to be associated with an increase in attentional bias to angry faces. Given that recent studies highlighted potential validity and reliability concerns with the DPT, temporal dynamics of attention allocation were examined by comparing the reliabilities of traditional BS computations with alternative computation models from the anxiety literature. This was followed by exploratory analyses of the associations between attention allocation to angry faces, CU traits, and self-reported delinquency. Finally, cognitive biases in later stages in social-information processing were expected to predict increased attentional bias to angry faces during encoding.

2.3. Method

The method, data preparation protocol, and data analysis plan were pre-registered with the Open Science Framework (OSF, see this link for full pre-registration). This study was approved by the Department of Psychology University of Sheffield Research Ethics Committee. Written informed consent was obtained and the data were analysed anonymously.

2.3.1. Participants

Power and alpha were set at 1-β= 0.80 and α= 0.05. The sample size required for a medium effect size f= 0.25 (Cohen 1988) for the pre-registered analysis of covariance was 128.

For the regression analyses, the sample size required to find a medium effect size f= 0.15, for α= 0.05 and 1-β= 0.80 was 103. This study over-recruited beyond the target sample (N= 137). Purposeful sampling was selected deliberately to recruit participants with a history of aggression or antisocial behaviours and traits. As such, the study was advertised on social media (N=52, Facebook, Reddit). Specifically, the study was advertised in social media groups such as the "Antisocial Personality Disorder", "Covert Narcissism Group", and "Borderline Personality Disorder Awareness". Participants were also recruited through Prolific Academic (N= 55) and filtered to select participants who answered 'yes' to a screening question regarding prison sentences ("Have you ever been in prison for committing a crime?"). Finally, in order to achieve the required sample size, 30 undergraduate students were recruited from the University of Sheffield. All participants were rewarded for their time: undergraduate students were offered credits, those recruited through social media were given the opportunity to enter a prize draw for a £50 Amazon Voucher, and participants recruited through Prolific Academic were remunerated according to the guidance on the online platform. After exclusions due to missing data (N= 17) and exclusion of participants with less than 80% valid trials on the DPT (N= 4), a final sample of 116 participants (64 males and 52 females) was used in the analyses. The mean age of the participants was 30.71 (SD= 12.10).

2.3.2 Measures

The Dot-Probe Task: Attentional bias to threat was inferred from RTs on a modified DPT using Inquisit software 5.0 (Millisecond Software, 2019). Task stimuli consisted of pairs of photographs (one angry and one neutral) of 20 individuals: photographs of 10 males and 9 females were taken from the NimStim stimulus set (Tottenham et al., 2009), and photographs of one female which were taken from a stimulus set reported by Matsumoto and Ekman (1989). Stimuli were presented on a white background with a continuously visible fixation cross placed in the centre of the screen. In each trial, a stimulus pair was presented above and below the fixation cross for either 100 ms or 500 ms. Due to the task being completed online, each image was placed 15% of the given screen height away from the fixation cross. Following the offset of the stimulus-pair, a target probe consisting of an arrow head pointing either left or right (“<” or “>”) was presented until a response was made. The participants were instructed to press ‘E’ when presented with the arrow pointing left and ‘I’ when it pointed right. Each response was followed by an inter-trial interval consisting of the fixation cross displayed for 1000 ms. Attentional bias to threat was inferred from faster reaction times to probes replacing the angry faces relative to probes replacing the neutral faces. Across trials, each expression was equally likely to be presented for 100 ms or 500 ms, each photograph was equally likely to be on the top or bottom position, and the probe was equally likely to point left or right. The protocol consisted of 10 practice trials during which participants were given feedback if they responded incorrectly to the probe, and 1 block of 120 angry-neutral trials. Angry-face location, probe location and type, and actor were counterbalanced in presentation although each pair of stimuli consisted of photos from the same actor. The DPT stimuli and protocol were sourced from the Tel-Aviv University/National Institute of Mental Health Attentional bias Measurement Toolbox (TAU/NIMH, Abend, Pine, & Bar-Haim, 2014). Participants were instructed to focus their attention on the fixation point throughout the task and to respond to the probes as quickly and accurately as possible. Traditional BS were calculated as the difference between reaction times to probes replacing neutral faces (neutral trials) and reaction times to probes replacing angry faces (angry trials).

Social-Emotional Information Processing

Social- Emotional Information Processing patterns were measured using the Social-Emotional Information Processing Questionnaire (SEIP-Q). SEIP-Q variables were assessed in relation to 8 vignettes comprising ambiguous scenarios in which Person A was the victim of an adverse action enacted by Person B. The vignettes were designed to contain either direct aggression (e.g., physical aggression) or relational aggression (e.g., rejection). Participants were required to identify with Person A and answer the questions immediately following each vignette. The Attribution variable of the SEIP-Q was assessed using four Likert-scale questions measuring: direct hostile intent (e.g., "This person wanted to damage my car"), indirect hostile intent (e.g., "This person wanted me to feel unimportant"), instrumental non-hostile intent (E.g., "This person was in a hurry to get in to work"), and benign intent (e.g., "This person scratched my car by accident and didn't notice"). Two items measuring Negative Emotional Response using Likert-scaled questions (e.g., How likely is it that you would be angry if this happened to you") follow the attribution items.

The SEIP-Q then offered three potential behavioural responses to each vignette (a) socially appropriate, b) overtly aggressive, and c) relationally aggressive) and assessed the desirability of each response. Each potential response was followed by seven Likert-scaled questions, assessing response evaluation and decision variables: "Response Valuation" (i.e., whether the considered response is an appropriate one), "Outcome Expectation" (i.e., how this response would impact your and others' opinion of you), "Response Efficacy" (i.e.., how easy it would be to choose this response), and "Response Enactment" (i.e., the likelihood of choosing this response). Response evaluation and decision variables were calculated as an overall score including each of the three scenarios in the vignette. Scores for the four separate components in the response evaluation and decision variable were also calculated as an overall mean score for all three scenarios. Individual SEIP-Q scores for each variable were averaged on the same 0-3 scale. The SEIP-Q has been validated in community and clinical samples and findings have indicated good to excellent psychometric properties, as well as good reliability and validity (Coccaro et al., 2017). The scale assesses all stages detailed in the SIP model by Crick and Dodge (1994) with the exception of encoding. In this sample, alpha coefficients were good to excellent for Hostile Attribution (HA: α= 0.90) and Negative Emotional Response (NER: α= 0.88). Alpha coefficients were lower for Benign Attribution (BA: α= 0.68) and Instrumental Attribution (IA: α= 0.74). The coefficients for HA, NER, and BA were similar to previous reports (i.e. HA: α= 0.88, NER: α= 0.87, BA: α= 0.68; Coccaro et al., 2017) whereas the alpha coefficient for IA was higher in this sample than coefficients reported in previous samples (IA: α= 0.63, Coccaro et al., 2017 and α= 0.57, Coccaro et al., 2009).

Reactive and Proactive Aggression were measured using the Reactive-Proactive Aggression Questionnaire (RPQ, Raine et al. 2006). The questionnaire consists of 23 items measuring proactive (12 items) and reactive aggression (11 items) based on incidence of behaviours on a 3-point Likert scale which ranges from 0 (never) to 2 (often). Example reactive aggression items include "Gotten angry when frustrated" and "Reacted angrily when provoked by others", whereas proactive aggression items include "Vandalized something for fun" and "Hurt others to win a game". The RPQ has been shown to be applicable in an adolescent and adult populations (Brugman et al., 2017), has demonstrated high reliability and internal consistency (Cronbach's alpha values of 0.81 and higher, Cenkseven-Önder et al., 2016) and differentiates between a community and a forensic sample (Brugman et al., 2017). Reactive aggression has been linked to hypervigilance to threat and hostile attribution bias (Dodge et al., 2015; Lobbestael et al., 2016). Comparatively, proactive aggression has been linked to later stages of the SIP model (Dodge 1990) and has been associated with threat avoidance (Lobbestael et al., 2016) and instrumental goals (Muñoz Centifanti et al., 2013).

Callous-Unemotional (CU) traits were measured using the Inventory of Callous-Unemotional traits (ICU, Frick 2004), a self-report measure designed to assess CU traits using 24 items (e.g., "I do not show my emotions to others", "I do not care whom I hurt to get what I want") rated on a 4-point Likert scale ranging from 0 ("Not at all true") to 3 ("Definitely true"). CU traits have been associated with reactive as well as proactive aggression in children and adolescent samples, and are considered a strong predictor of antisocial behaviour in adulthood (Frick et al., 2014; Thomson & Centifanti, 2018). Whilst the ICU was initially developed for children and adolescents, it has been validated in adult samples (Byrd et al., 2013) and CU traits (and each subscale of the ICU) have significantly predicted adult antisocial behaviour beyond well-established predictors of offending (Kahn et al., 2013). In the present study, Chronbach's α was 0.69 for the CU traits scale. The unemotional subscale has previously been reported as having low reliability (Cardinale & Marsh, 2020); however, it was included in this study so as to allow comparison with previous findings.

The International Self-Report Delinquency instrument *(ISRD,* Short Version, Enzmann 2013) consists of 22 items assessing self-reported delinquency. Each question about the lifetime prevalence of a delinquent behaviour is immediately followed by an open-ended question about the frequency during the past month (for drug offences) or year. Whilst the overall lifetime prevalence score was used in this study to investigate links to attentional bias, the prevalence and specificity of antisocial behaviours in the present sample was also examined. The offenses were grouped into three categories: *property related offences* (vandalism, shoplifting, burglary, bicycle theft, car theft, and car breaking-i.e. stealing from a car), *violent offences* (carrying a weapon, group fights, snatching, extortion, assault), and *drug related offences* (drug use, selling drugs). Past month prevalence was recorded for drug related offences, whereas past year prevalences were recorded for property-related and violent offences.

2.3.3. Analysis Plan

A cross-sectional study design was used to investigate the associations between attentional bias to threat, social information-processing, and self-report measures of aggression, delinquency, and CU traits. A 2 x 2 repeated measures design was used with within-participants factors of emotion-type (angry and neutral) and stimulus presentation duration (100 and 500 ms) as independent variables (IVs), RTs as the dependent variable (DV), and reactive and proactive aggression scores and delinquency scores introduced as covariates. Social information-processing, aggression, delinquency, and CU traits were introduced as predictors in regression models with BS as DV.

I. Is an attentional bias to angry faces displayed when attention allocation is measured at an earlier (100 ms) compared to a later (500 ms) time and is it related to self-reported aggression and delinquency?

Individuals displaying higher reactive aggression were expected to display an attentional bias to angry faces when these were presented for 100 ms. A two-way repeated measures ANOVA with RTs as DV and emotional valence of the target (angry vs neutral) and target presentation duration (100 ms vs 500 ms) as IVs was expected to show an interaction effect with faster RTs to probes replacing angry faces displayed for 100 ms but not at 500 ms. The addition of reactive aggression as a covariate was expected to result in a significant effect of reactive aggression and no interaction effect between the two IVs. Exploratory analyses were planned with proactive aggression and self-reported delinquency as covariates. Finally, hierarchical regressions were planned to evaluate the prediction of attentional bias from aggression subfactors, delinquency, and CU traits. Regression models predicting BS added age, gender, and education at step 1, reactive aggression at step 2, proactive aggression at step 3, and self-reported delinquency and CU scores at step 4.

II. Do later steps in social-information processing predict an attentional bias to angry faces presented for 100 ms?

Measures of later stages in social-information processing were expected to predict a unique part of the variance in attentional bias scores for targets presented for 100 ms. Multiple linear regression models were conducted to investigate the relationship between BS when targets were presented for 100 ms and: 1) attributional and emotional response variables (hostile attribution, instrumental attribution, and negative emotional response), 2) response evaluation and decision variables for directly and relationally aggressive responses. Within the latter, response enactment, outcome expectation, response efficacy, and response valuation for relationally and directly aggressive responses were significantly correlated (correlations between 0.56 and 0.70). Composite variables were created using score averages for directly and relationally aggressive responses for each of these measures. Standardized alpha coefficients for the composite variables were excellent: α= 0.85 for composite response enactment, α= 0.94 for outcome expectation, α= 0.87 for response valuation, and α= 0.88 for response efficacy. A multiple regression model was computed with BS as the dependent variable and the composite response evaluation and decision variables as SIP predictors.

2.3.4. Procedure

Participants were provided with an identical link to the Inquisit website which hosted the DPT (Millisecond Software, 2019). The participants were asked to read the information sheet and consent to taking part in the study. Participants then completed the DPT followed by the self-reported measures which were recorded online using the Qualtrics research software (Qualtrics, 2019). The DPT has been used to induce cognitive biases in previous research (i.e., repeated exposure to threatening faces has been associated with an increased bias toward hostility-related information, Bar-Haim et al., 2007); therefore, this order precluded priming effects. All participants completed the procedure above; however, the undergraduate students recruited from the University of Sheffield completed the study in quiet laboratory spaces whereas the remaining participants completed the study remotely.

2.3.5. Data Preparation

Outliers in the attentional bias and self-report measures were Winsorized as pre-registered; values outside 1.5 interquartile ranges from the Tukey Hinges (lower and upper hinge corresponding to the first and upper quartiles or 25th and 75th percentiles respectively) were rescaled to the last valid value within the range. This approach has demonstrated a consistent benefit on score reliability across a range of DPT modifications (Price et al., 2015). Analyses were performed using RStudio 1.2.1335 (RStudio, 2018). An initial check showed no missing values in the attentional bias data. Incorrect responses on the DPT were excluded from analysis. Participants with less than 80% valid neutral or angry trials were excluded (N= 4). Finally, participants missing self-report data were removed (N=3).

Participants taking part in this study answered questions regarding past aggression and delinquent behaviour, a blanket approach to handling missing data was therefore not deemed appropriate for this study. Instead, patterns of missingness in the data were inspected using missing values heat maps for each self-report measure. If more than 20% of questions within a self-reported measure were left unanswered, the participant was excluded from any analyses including the measure (N=14). Missing data maps were examined for the remaining participants to verify that any remaining missing data patterns appeared random. The final analysis was conducted on data from 116 participants.

Examination of histogram plots, box plots, skewness and kurtosis indicated that age, proactive aggression scores, and self-reported delinquency scores were significantly positively skewed. Whilst a square root transformation was applied to age scores, the distribution of scores remained positively skewed. Analyses using raw and transformed age scores were highly similar, therefore raw scores were used in the analyses. Proactive aggression scores were converted from a continuous to a binary variable which distinguished scores from values above 0 (coded “1”). A histogram of self-reported delinquency scores showed a significant positive skew in the data due to a large number of participants either reporting not engaging in any antisocial behaviour (N= 40) or one instance of antisocial behaviour (N= 66). In order to identify individuals with a disproportionately high number of offences, self-reported delinquency scores were converted from a continuous variable to a binary variable in which scores below (not including) the value of 3 (corresponding to the 3rd quartile) were assigned to the value of "0", whereas scores above (including) the value of 3 were assigned to the value of "1".

2.3.6. Computation and reliability of attentional bias data and self-report measures

Traditional BS were computed per presentation duration for each participant. The split-half reliability of traditional BS was measured for each condition (i.e., presentation duration and emotional valence of the target) using computational methods. That is, 5000 random splits of the BS were performed and an overall reliability score was computed based on the split average. The resulting estimates were corrected using the Spearman-Brown formula (*rSB*) and two pairs of stimuli were removed from the dataset due to very low split-half reliabilities. Low split-half reliabilities were found for traditional BS for both 100 (r_{SB}= 0.17) and 500 ms (r_{SB}= -0.27).

Trial-level BS (Zvielli et al., 2015) were computed as an alternative to traditional BS. The pattern of the trial-level BS scores differs from a static trait-like signal as it examines attentional bias expression in time and trial-level BS have demonstrated overall higher levels of reliability than traditional BS (Zvielli et al., 2015). As stimulus presentation duration and congruency were randomised and counterbalanced in this study, a weighted-trials computational method was used instead of the nearest-trial method proposed by Zvielli et al. (2015), in which each trial is compared to the most temporally proximal trial of the opposite type. The weighted-trials method calculates the weighted mean of all trials of the opposite type, with closer trials weighted more heavily than more distant trials. Trial-level BS are then calculated by subtracting each trial from the weighted mean of all trials of the opposite type. Whilst in practice weighted-trial and nearest-trial methods yield similar trial-level BS, the weighted method poses two further advantages: 1) if a trial is equidistant from two trials of opposite type, the weighted method takes into account the mean of the two trials whereas the nearest-trial method arbitrarily chooses one pair over another, and 2) the nearest-trial method produces double-counts of trials such as Incongruent-Incongruent-Congruent-Congruent which results in artefactual periods where trial-level BS time series are flat.

Studies have shown that mean trial-level BS provide good to excellent reliability (Rodebaugh 2016). As trial-level BS are bi-dimensional (values above and below 0), it is divided into a) the mean of trial-level BS scores that are **above 0** for each participant (Mean trial-level $BS_{POSITIVE}$), an indicator of individual differences in the degree to which attention moves toward target stimuli; or b) the mean of trial-level BS scores that are **below 0** for each participant (Mean trial-level $BS_{NEGATIVE}$), an indicator of individual differences in the degree to which attention moves away from target stimuli.

Using the weighted-trial method, a time-series of trial-level BS per participant was calculated for the 100 (Figure 2) and 500 (Figure 3) ms conditions by subtracting each trial in which the probe replaced an angry face from the weighted mean of all trials where the probe replaced a neutral face. Next, $BS_{POSITIVE}$ (mean of positive trial-level BS, indicating attention being drawn toward angry targets) and $BS_{NEGATIVE}$ (mean of negative trial-level BS, indicating attention being drawn away from angry targets) were calculated for each participant, for each presentation duration. Split-half reliabilities for trial-level BS were considerably higher (r_{SB}= 0.79 for the 100 ms condition and r_{SB}= 0.77 for the 500 ms condition) than traditional BS; trial-level BS parameters were henceforward used as measurement of attentional bias and were referred to as $BS_{POSITIVE/NEGATIVE}$.

Figure 2

Trial-level bias scores for the 100 ms presentation condition: time series for all participants

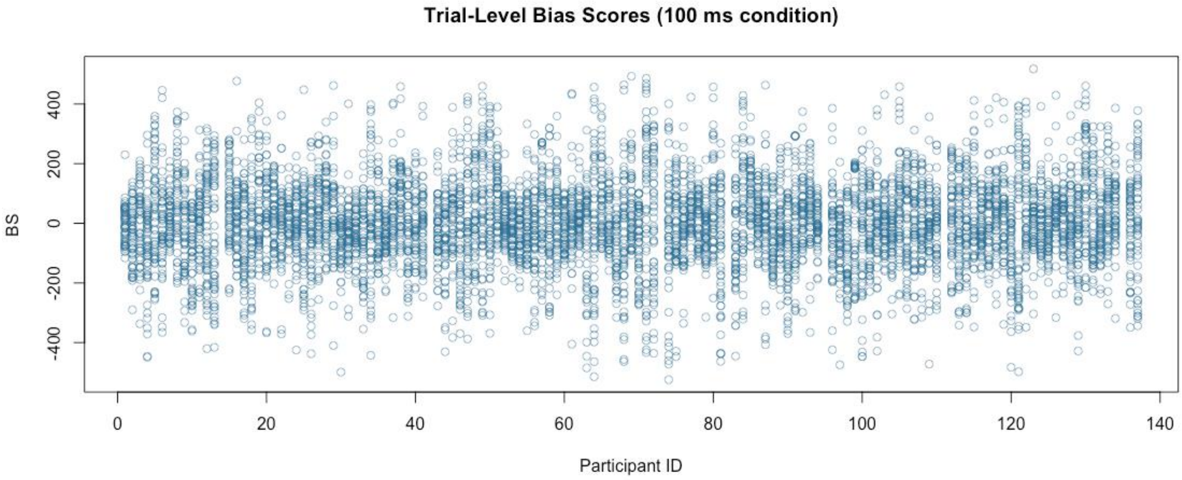
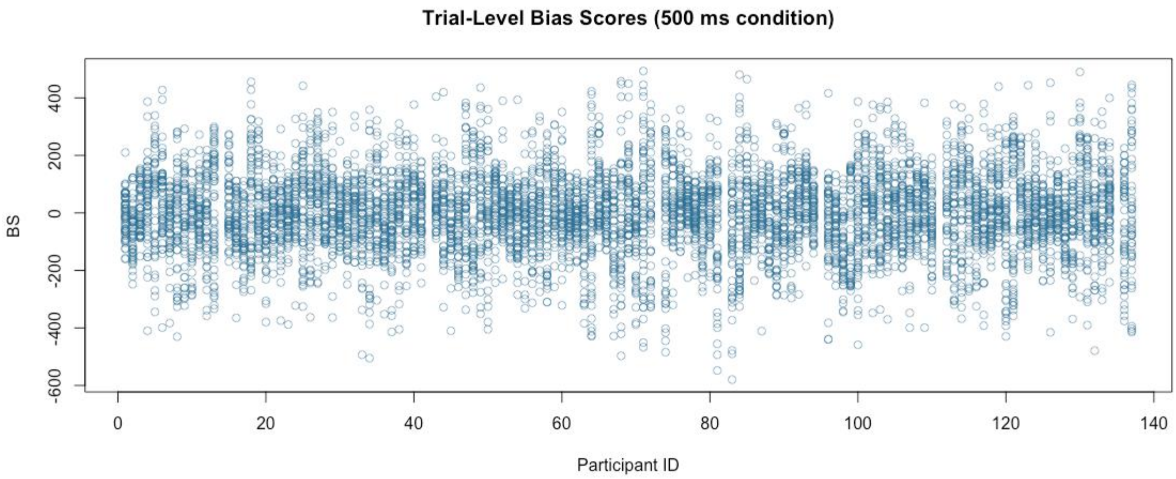


Figure 3

Trial-level bias scores for the 500 ms presentation condition: time series for all participants



2.4. Results

2.4.1. Sample characteristics: self-report measures

Table 1 gives the means and standard deviations for BS, RPQ and reactive aggression, CU traits, self-reported delinquency, and attributional and emotional response variables (hostile attribution, intent attribution, negative evaluation and response, and benign attribution). Table 2 summarises the correlations between BS and self-reported measures. No significant correlations were found between RPQ and CU traits or between BS and self-reported measures. There was a significant positive correlation between CU traits and reactive aggression (r= 0.21, p<0.05). Self-report data for aggression, delinquency, and CU traits are briefly summarized in the section below and revisited in the discussion.

Table 1

Means and standard deviations for aggression, CU traits, self-reported delinquency, SEIP-Q variables

	Mean	Standard Deviation
Reactive Proactive Aggression Questionnaire	10.71	5.79
Reactive Aggression	8.13	3.66
Proactive Aggression	2.51	2.5
CU Traits	29.19	6.07
Inventory of Self-reported Delinquency	2.28	3.01
Hostile attribution	1.92	0.48
Intent attribution	3.2	0.38
Negative evaluation and response	2.61	0.52
Benign attribution	2.71	0.48

Table 2

Correlation table for bias scores, aggression, CU traits

	$BS_{POSITIVE}$ (100 ms)	$BS_{NEGATIVE}$ (100 ms)	$BS_{POSITIVE}$ (500 ms)	$BS_{NEGATIVE}$ (500 ms)	Reactive Aggression	CU Traits
$BS_{POSITIVE}$ (100 ms)	-					
$BS_{NEGATIVE}$ (100 ms)	*-.35****	-				
$BS_{POSITIVE}$ (500 ms)	*.77****	*-.42****	-			
$BS_{NEGATIVE}$ (500 ms)	*-.36****	*.79****	*-.45****	-		
Reactive Aggression	*-.05*	*.07*	*.01*	*.06*	-	
CU Traits	*-.04*	*.01*	*-.02*	*.05*	*.21**	-

Note. BS= bias scores

*p<0.05, **p<0.01, ***p< 0.001.

Self-reported delinquency data showed a high **lifetime** prevalence of drug offences in this sample (39.7%) with 37.1% of participants reporting cannabis use, 12.9% reporting speed/xtc use, and 19.8% reporting use of heroin/cocaine at least once in their lifetime. Comparatively, participants reported using cannabis (25%) followed by heroin/cocaine (9.5%), and speed/xtc (8.6%) at least once in **the month** prior to taking part in the study. In terms of **lifetime prevalence**, over half of the participants (56.9%) reported at least one property related offence and 17.2% of sample reported having committed at least one violent offence. Comparatively, in the **12 months prior** to taking part in the study, 37.9% of participants reported having committed at least one property related offence and 8.6% of participants committed at least one violent offence. Property related offences in the 12 months prior to taking part in the study most commonly consisted of illegal downloads (37.1%), break-ins, and theft (4.4%). Of the violent offences committed in the 12 months prior to taking part in the study, 5.2% comprised carrying a weapon, and, as the questions had an open-ended format, three of the participants reported carrying a weapon on a daily basis, whereas one reported carrying a weapon "whenever I feel unsafe or feel like I might need it". This was followed by 2.6% of participants reporting taking part in group fights in the past year.

2.4.2. Online attentional bias data collection- Exploratory analyses

An independent-samples t-test was conducted to compare reaction times in the DPT for laboratory conditions and online collected data. Reaction times in laboratory settings (*M*= 515.52, *SD*= 89.11) were significantly faster than online reaction times (*M*= 589.03, *SD*= 106.19) conditions, ($t(232)= 7.32$, $p<0.001$). However, given that attentional BS are calculated on the basis of within-participant variation, analyses using these as DVs should not be affected by differences between participants. Conversely, ANOVA and ANCOVA tests using latency (reaction times) as a DV are likely to be affected. Separate models run in laboratory and online subsamples will therefore be computed alongside pre-registered full sample models to include experimental conditions (i.e., online versus laboratory-based) as an IV.

2.4.3. Results relating to pre-registered hypotheses

I. Is an attentional bias to angry faces displayed when attention allocation is measured at an earlier (100 ms) compared to a later (500 ms) time and is it related to self-reported aggression and delinquency?

A two-way repeated measures ANOVA was conducted to investigate changes in RTs as a function of emotional valence of the target (angry vs neutral) and target presentation duration (100 ms vs 500 ms). The results indicated a main effect of target presentation duration, ($F(1, 115)= 37.37$, $p<.001$, $\eta^2= 0.25$) with slower RTs for stimuli presented for 100 ms ($M= 578.55$, $SD= 108.01$) compared to 500 ms ($M= 562.75$, $SD= 105.53$). Contrary to the hypothesis, the effect of the emotional valence of the target ($F(1, 115)= 2.65$, $p= 0.11$) and the interaction between emotional valence of the target and target duration ($F(1, 115) = 1.38$, $p= 0.24$) were not significant. The ANCOVA was not conducted as the expected interaction effect was not significant.

Planned exploratory analyses

Two 2x2x2 mixed design ANOVA was conducted with latency (dependent variable) as a function of two within-subject IVs (emotional valence of the target and target duration) and between-subject IVs (self-reported delinquency category and proactive aggression- high vs low scores). There was no significant main effect or interaction effects. There was a main effect of self-reported delinquency scores ($F(114)= 5.12$, $p= 0.02$, $\eta^2= 0.04$) with participants in the higher delinquency group displaying faster reaction times to stimuli ($M= 538.97$, $SD= 100.82$)

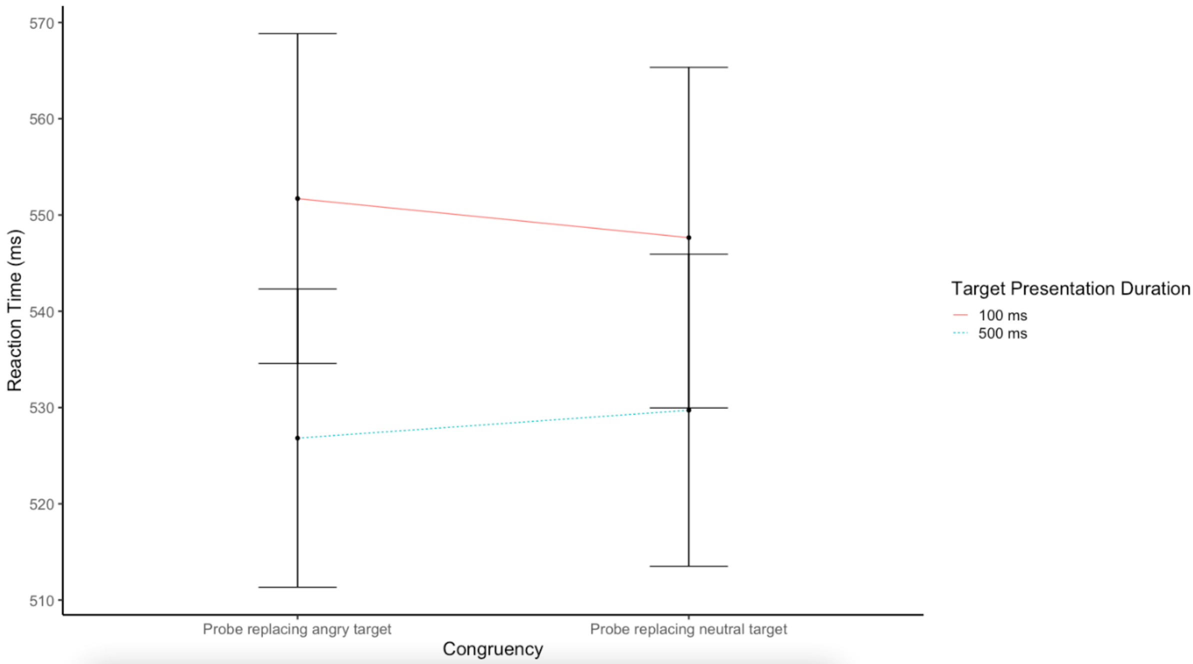
compared to participants in the lower delinquency group (M= 585.48, SD= 106.67). There was also a significant main effect of target duration (F(114)= 37.77, p<0.001, η^2= 0.25).

A statistically significant three-way interaction was found between target presentation duration, congruency, and self-reported delinquency categories (F(1, 114)= 3.80, p= 0.05, η_p^2= 0.03). For the simple two-way interactions and simple main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the p < 0.025 level. There were no statistically significant two-way interactions between target presentation duration and congruency for either low or high delinquency score categories.

Pairwise comparisons were conducted for each self-reported delinquency group. First, pairwise comparisons were run in the **high delinquency group** between RTs to probes replacing angry compared to neutral faces (congruency) across target presentation durations (100 and 500 ms) (see Fig. 4). There were no significant differences in comparisons.

Figure 4

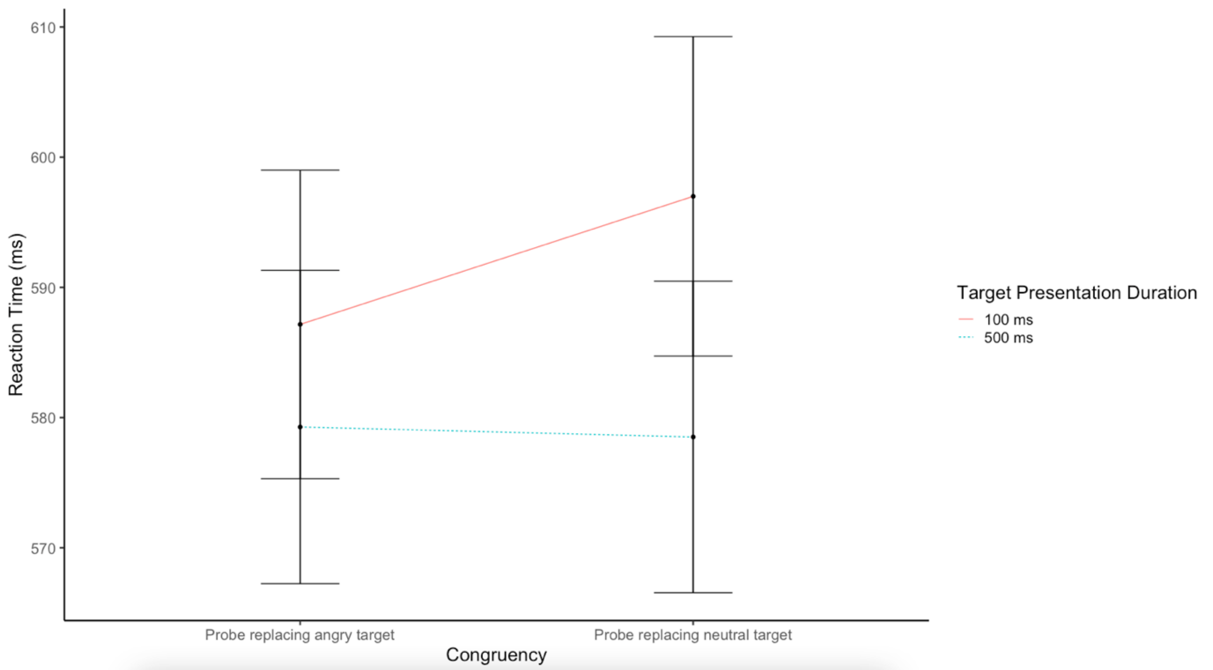
Interaction between congruency and target presentation duration in the high-delinquency group. Error bars show one standard error above and below mean.



Next, pairwise comparisons were run in the **low delinquency group** between RTs to probes replacing angry compared to neutral faces (congruency) across target presentation durations (100 and 500 ms) (see Fig. 5). A Bonferroni adjustment was applied (significance being accepted at p<0.025). When targets were presented for 100 ms, RTs were significantly faster for probes replacing angry faces (M= 587.16, SD= 105.33) compared to probes replacing neutral faces (M= 596.99, SD= 109.04, p= 0.01). No other comparisons were significant.

Figure 5

Interaction between congruency and target presentation duration in the low-delinquency group. Error bars show one standard error above and below mean.



Hierarchical Regression Models

The 2x2 ANOVA (see Hypothesis 1 above) showed no evidence of an attentional bias toward threat across participants; the 2x2x2 ANOVA with high and low self-reported delinquency as IV showed a relationship between self-reported delinquency and attentional bias to threat in the 100 ms presentation condition. Hierarchical models were computed as pre-registered with $BS_{POSITIVE}$ and $BS_{NEGATIVE}$ in the 100 ms presentation condition as the DV.

Hierarchical regression models were conducted with mean $BS_{POSITIVE}$ (see Table 3) and $BS_{NEGATIVE}$ (see Table 4) for targets presented for 100 ms as dependent variables. There were no significant predictors of BS and none of the steps were a significant improvement in fit.

Table 3

Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces ($BS_{POSITIVE}$).

	β	*SE*	*p*	R^2 *Total/Change*
Step 1				0.02
Gender	0.16	0.10	0.12	
Education	-0.01	0.09	0.94	
Age	-0.08	0.10	0.41	
Step 2				0.01/0.61
Gender	0.16	0.10	0.12	
Education	-0.01	0.09	0.52	
Age	-0.08	0.10	0.42	
Reactive aggression	- 0.05	0.09	0.61	
Step 3				0.01/0.22
Gender	0.18	0.10	0.07	

Education	-0.02	0.09	0.85	
Age	-0.07	0.10	0.46	
Reactive aggression	-0.11	0.11	0.30	
Proactive Aggression	0.14	0.10	0.22	
Step 4				0.002/ 0.70
Gender	0.16	0.10	0.14	
Education	-0.01	0.09	0.88	
Age	-0.09	0.11	0.37	
Reactive aggression	-0.09	0.11	0.44	
Proactive Aggression	0.15	0.11	0.17	
Self-Reported Delinquency	-0.09	0.11	0.40	
CU Traits	-0.003	0.02	0.84	

Note. *β*= standardized regression coefficient, SE= Standard Error.

Table 4

Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces ($BS_{NEGATIVE}$).

	β	***SE***	***p***	***R^2 Total/Change***
Step 1				0.01
Gender	-0.08	0.10	0.42	
Education	-0.02	0.10	0.85	
Age	-0.07	0.10	0.49	
Step 2				0.02/0.43
Gender	-0.08	0.10	0.43	
Education	-0.006	0.10	0.95	

Age	-0.07	0.10	0.47	
Reactive aggression	0.08	0.10	0.42	
Step 3				0.04/0.63
Gender	-0.09	0.11	0.39	
Education	-0.004	0.10	0.96	
Age	-0.08	0.10	0.46	
Reactive aggression	0.10	0.11	0.35	
Proactive Aggression	-0.05	0.11	0.62	
Step 4				0.05/ 0.88
Gender	-0.11	0.11	0.34	
Education	-0.001	0.10	0.99	
Age	-0.09	0.11	0.41	
Reactive aggression	0.11	0.11	0.31	
Proactive Aggression	-0.04	0.11	0.71	
Self-Reported Delinquency	-0.05	0.11	0.63	
CU Traits	-0.002	0.02	0.87	

Note. β= standardized regression coefficient, SE= Standard Error

II. Do later steps in social-information processing predict an attentional bias to angry faces presented for 100 ms?

Hostile attribution, instrumental attribution, and negative emotional response did not significantly predict bias toward or away targets presented for 100 ms. Results of the multiple linear regression with $BS_{POSITIVE}$ as DV and outcome expectation, response efficacy, response valuation, and response enactment composite scores for directly and relationally aggression as IVs showed that only response enactment was a significant negative predictor of bias toward probes replacing angry faces (β= -0.36, p= 0.01). There were no significant response evaluation and decision predictors of bias away from probes replacing angry faces.

Links between SEIP-Q scores, self-reported aggression, delinquency, and CU traits.

A visual inspection of the correlation matrix (see Table 5) showed significant correlations between SEIP-Q scores for the response evaluation and decision variables within each response type (socially appropriate, overtly, and relationally aggressive responses) as well as between relationally and directly aggressive responses. However, response efficacy ("*How easy would it be for you to act like this?*") was significantly negatively correlated with all other variables within the respective response type. Response evaluation and decision variables for aggressive responses were positively correlated to reactive aggression but not to CU traits. Similarly, there was a significant positive correlation between negative emotional response and reactive aggression but no correlation with CU traits. The latter were significantly correlated with response evaluation and decision variables for the socially appropriate response. There was no correlation between hostile attribution and either reactive aggression or CU traits.

Table 5

Correlation matrix for self-report measures of Reactive Aggression, CU traits, and scores on the SEIP-Q

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Reactive Aggression	-																	
2	CU traits	*.21**	-																
3	Hostile Attribution	*.17*	*-.04*	-															
4	Negative Emotional Response	*.35****	*.11*	*.57****	-														
5	Instrumental Attribution	*.35****	*.34***	*.16*	*.23**	-													
6	Benign Attribution	*-.01*	*.12*	*-.40****	*-.16*	*.15*	-												
7	Response valuation- Appropriate response	*.07*	*.14*	*-.07*	*.18*	*.43****	*.10*	-											
8	Outcome Expectation- Appropriate response	*.09*	*.23**	*-.06*	*.17*	*.43****	*.20**	*.78****	-										
9	Response Efficacy- Appropriate response	*.06*	*-.21**	*.35****	*.27***	*-.23**	*-.25***	*-.44****	*-.45****	-									
10	Response Enactment- Appropriate response	*.11*	*.23**	*-.06*	*.14*	*.43****	*.24**	*.68*****	*.69****	*-.59****	-								
11	Response valuation- Aggressive response	*.19**	*.01*	*.38****	*.29***	*-.12*	*-.23**	*-.21**	*-.09*	*.20**	*-.10*	-							
12	Outcome Expectation- Aggressive response	*.22**	*.09*	*.39****	*.25***	*-.02*	*-.23**	*-.10*	*-.00*	*.17*	*.02*	*.79****	-						
13	Response Efficacy- Aggressive response	*-.22**	*.02*	*-.20**	*-.14*	*.09*	*.10*	*.21**	*.16*	*-.00*	*.16*	*-.53****	*-.46****	-					

14	Response Enactment- Aggressive response	.35***	.03	.51***	.43***	.07	-.24**	-.02	.01	.26**	-.06	.69***	.61***	-.67***	-				
15	Response valuation- Relational aggression	.04	-.04	.42***	.26**	-.17	-.27**	-.05	-.01	.19*	-.07	.64***	.51***	-.23*	.48***	-			
16	Outcome Expectation- Relational aggression	.05	.08	.48***	.26**	-.07	-.34***	-.00	.11	.19*	.03	.60***	.70***	-.23*	.46***	.78***	-		
17	Response Efficacy- Relational aggression	-.04	.07	-.20*	-.15	.06	.19*	.08	.13	-.10	.21*	-.31***	-.29**	.57***	-.38***	-.44***	-.44***		
18	Response Enactment- Relational aggression	.11	.04	.49***	.41***	-.01	-.32***	.06	-.02	.39***	-.14	.42***	.39***	-.25**	.56***	.68***	.63***	-.66***	-

Note: *p<0.05, **p<0.01, ***p< 0.001

2.5. Discussion

This study set out to examine whether by using the DPT, an established measure of attention allocation, an attentional bias to threat would be found when varying the timeframe of attention allocation to angry faces. While the findings did not indicate an attentional bias toward angry faces for either presentation duration (100 vs 500 ms), participants were faster when responding to probes replacing targets presented for 500 ms compared to 100 ms. Further, self-reported measures of aggression, delinquency, and CU traits did not predict an attentional bias toward angry faces. However, SIP measurements showed a link between attentional bias and likelihood of enacting an aggressive behavioural response. A key strength of this study was that established, reliable instruments were used to measure constructs of interest. This research therefore not only adds to a growing body of literature on SIP biases, aggression, antisocial behaviour, and CU traits, but sheds additional light on previously unexplored associations between these constructs. These are canvassed below, beginning with a discussion of links between attentional bias and self-report measures in the context of each hypothesis and the wider literature, followed by an overview of the relationships between self-report measures and sample characteristics.

I. Is an attentional bias to angry faces displayed when attention allocation is measured at an earlier (100 ms) compared to a later (500 ms) time and is it related to self-reported aggression and delinquency?

RT data indicated that attention was deployed differently for the two target presentation durations. Although previous studies comparing RTs to targets presented for 100 and 500 ms found faster RTs for the former (Cooper & Langton, 2006), the opposite pattern of results was found here. That is, participants responded faster to stimuli presented for 500 ms than 100 ms. Previous studies have used between-participants designs (Staugaard, 2009) or grouped trials into blocks to compare RTs to different presentation times (Zvielli et al., 2015). These methods however did not take into consideration habituation to stimulus presentation durations which has been demonstrated to have an impact on RTs in the DPT (Staugaard, 2009). Consequently, the different pattern of attention allocation in the present study may therefore be due to stimuli being randomly presented for 100 or 500 ms thereby minimising trial-type adjustment.

Further, contrary to the first hypothesis, there was no evidence of an attentional bias to angry faces for either target presentation duration and no effect of reactive or proactive aggression on RTs. These findings are not in line with previous research using the DPT to investigate abnormal attention allocation to emotional stimuli in relation to externalizing traits and behaviours (Edalati et al., 2016); however, this does not necessarily indicate the absence of such an attentional bias. Recent reviews have highlighted inconsistencies in findings across studies using the DPT (Torrence & Troup, 2018, Chapman et al., 2019). The lack of studies addressing DPT reliability or using reliable attentional BS means that no meaningful conclusions can be drawn as to whether the DPT is a useful measure of attentional bias (Thigpen et al., 2018).

Previous studies have shown stimulus display (or exposure) time to have an important role in both behavioural and neural attentional bias findings (see Torrence & Troup, 2018 for a review). Display (or exposure) times of under 500 ms for threatening stimuli have been shown to yield almost double the attentional bias effect size compared to exposure times of 500 ms and above in anxious individuals (Bar-Haim et al., 2007). Studies using event-related potentials have also found effects of stimulus presentation duration, with some neural components (such as the P1 and P3) reflecting an attentional bias following brief stimulus presentations (Fox et al., 2008), compared to other components (such as the N170 and N2pc) which had enhanced amplitudes for stimuli presented for 500 ms (Kappenman et al., 2014). These results indicate that for longer stimulus presentation durations, attention is likely to disengage from angry faces before probe presentation, indicating a move away from stimulus-driven attention and into goal-directed attention. This is consistent with previous work on preparatory attention which is causally related to attentional selection in goal-driven tasks (Battistoni et al., 2017). In the present study, faster RTs to probes replacing stimuli presented for 500 ms may thus indicate that participants had sufficient time to disengage from the emotional stimuli and focus their attention on task requirements, therefore displaying increased efficiency in identifying the location of the probe.

Although participants in the high self-reported delinquency group displayed faster RTs to probes replacing both neutral and angry faces, the expected attentional bias toward angry faces presented for 100 ms was only observed in the low-delinquency group. Previous studies have identified a link between a history of criminal or antisocial behaviour and attentional bias toward threat (Sadek et al., 2021). Some studies have found a history of antisocial behaviour and antisocial traits predicted increased attentional bias to threat (Domes et al., 2013), others have found no evidence (Edalati et al., 2016), or even reduced responsiveness (Kimonis et al., 2006). For example, Smith and Waterman (2004) used a modified DPT in which threatening words were presented for 200 ms to investigate attentional bias to threat and found violent and non-violent offenders were slower to respond to probes replacing aggression-themed words compared to undergraduate students. They also found that violent offenders were slower than non-violent offenders and undergraduates to locate a neutral word surrounded by aggression-themed distractor words (presented for 500 ms). These studies suggest a distinct pattern of attention allocation to threat-related stimuli in individuals with a history of antisocial behaviours and traits; however, ambiguity in the results may be due to the mediating roles of psychopathic (Kimonis et al., 2020) or aggressive traits (Chan et al., 2010) or due to the variation and unreliability in DPT methodology (Puls & Rothermund, 2018).

Reactive aggression, self-reported delinquency, and CU traits did not significantly predict attentional bias toward angry faces presented for 100 ms. Previous associations between attentional bias to angry faces and self-reported externalizing measures have been tenuous, particularly in studies using the DPT. For instance, Schippell et al. (2003) found evidence for reactive (but not proactive) aggression predicting a bias away from threat. Other studies however failed to find evidence of a relationship between aggression and attentional bias to threat using the DPT (Kimonis et al., 2008) although one study found that higher aggression was predicted by higher CU traits at low levels of attentional bias (Kimonis et al., 2018). While these mixed findings could indicate an indirect effect of attentional bias on aggression, they may also reflect the low reliability of traditional computations of attentional BS in the DPT. In a recent paper, Miller and Johnston (2019) found increased attention orienting to threat during a DPT was positively associated with aggression although the split-half reliability of the BS (traditional computation as scores of difference) was as low as 0.14.

Within the SIP model, individuals high in reactive aggression are expected to exhibit early hypervigilance to threat (Li et al., 2013). A recent review of attentional biases to threat or hostile stimuli in relation to reactive aggression (Manning, 2020) found that across 12 eligible experiments, 6 different instruments were used to measure aggression and 5 experimental paradigms were used to measure attentional bias to threat. Most experiments found a relationship between reactive aggression and attentional bias to threat; however, variation in experimental designs and samples makes it difficult to ascertain which attentional processes are linked to aggression and whether they are specific to age groups and forensic samples or generalisable to wider populations. Although this association supports the SIP model which posits that attentional biases precede aggression (Maoz et al., 2017), steps in the SIP model following encoding and attention allocation are expected to contribute to the final outcome (de Castro et al., 2005).

II. Do steps in SIP following encoding predict an attentional bias to threat for early targets?

Contrary to the second hypothesis, attributional and emotional response variables of the SEIP-Q did not predict attentional bias to threat for 100 ms. However, a relationship was found between attentional bias and response enactment. That is, an increased self-reported likelihood of engaging in aggressive behaviours predicted a decrease in attentional bias toward probes replacing angry faces. Response enactment for aggressive responses was also significantly positively correlated with reactive aggression, indicating conceptual overlap between aggressive traits and increased likelihood of engaging in aggressive response. According to the SIP model, an increase in reactive aggression and response enactment for aggressive responses should be linked to an increase in attentional bias toward threat (Fontaine & Dodge, 2006); however, the opposite pattern of results was found. Similarly, previous research has shown a robust link between aggression and hostile attribution (Dodge et al., 2015). In this study, no significant correlations were found between hostile attribution and aggression, CU traits, or attentional bias.

The lack of a relationship between behavioural measures of attentional bias to angry faces and measures of aggression, CU traits, and later steps in the SIP model may be related to the unreliability of the DPT BS used in previous publications (Kappenman et al., 2014). Given that previous reports have relied on traditional BS computations, the lack of a significant association between aggression and encoding in this study may have been due to our use of trial-level scores which, albeit considerably more reliable, have not been used in relation to externalizing measures. Previous research has linked attentional biases as measured by various visual tasks to aggression (Manning, 2020). Further, studies using structural equation modelling have found evidence of indirect effects of early SIP stages on behavioural outcomes, mediated by response evaluation and generation stages (de Castro et al., 2005), providing further support for the relationship between encoding biases and aggressive behavioural outcomes being mediated by the SIP stages in between. It is therefore possible that the DPT is not a suitable measurement of attentional bias, meaning that the measure was unable to capture the relationship between attentional bias, later stages in social-information processing, and aggression.

In this sample, attentional bias to angry faces was found in the low-delinquency group and increased bias was predicted by lower aggressive response enactment scores. This suggests that participants with a history of antisocial behaviour or higher proneness toward engaging in aggressive behaviours are more efficient when completing the DPT. That is, they were faster to identify targets across conditions and displayed a pattern of avoidance of probes replacing angry targets. This may be due to an increased ability to disengage from the emotionally valenced stimuli to engage with task demands, resulting in a homogenous, improved performance across conditions. Avoidance of threat in order to achieve goals has been associated with externalizing traits such as proactive aggression (Lobbestael et al., 2016) whereas higher reactive aggression and trait anger have been linked to increased approach movements (Veenstra et al., 2018). Given that visual bias in this study was only linked to variables measuring antisocial behaviour or enactment of aggressive behavioural responses, the measurement of spatial attention allocation using a DPT with short presentation duration for angry stimuli reflects an enhanced visual performance in order to achieve instrumental goals.

The sampling strategy for this study was aimed at recruiting participants with a history of aggression and delinquency. However, recruiting across social media platforms, university recruitment platforms, and Prolific Academic resulted in a final sample from a wide range of backgrounds meaning the present pattern of results may not be specific to individuals with aggressive or antisocial tendencies. Measures of delinquency showed that nearly half of the participants had used illicit substances at least once in their lifetime (39.7%). In line with previous findings (see Peacock et al., 2018), past month prevalence of drug use was higher for cannabis (25%) than it was for hard drugs (18.1%) although both were considerably higher than the estimated annual prevalence. Conversely, lifetime prevalence of violent offences (17.2% of the participants reported at least one violent offence) was similar to previous findings in community adolescent samples (Enzmann, 2013, Muftić et al., 2014). Additionally, aggression scores in the present sample were similar to scores found in a large (N= 465) student sample (Preston & Anestis, 2019), lower than scores recorded in offender populations (Oostermeijer et al., 2016) although comparable to aggression scores in a sample of adult non-violent offenders (Kuin et al., 2017). In other words, with the exception of substance misuse measurements which are often prone to under-reporting (Gomes et al., 2019), there are no indicators in the present sample to suggest higher than average scores on measures of aggression and delinquency.

Whilst measurements of aggression and delinquency have been used extensively across sample types and age groups, CU traits have mostly been investigated in children and adolescent samples, as a precursor or predictor of risk for psychopathy in adulthood (Cardinale & Marsh, 2020). Nonetheless, within the last decade the ICU has been validated in adult samples and linked to delinquency (Byrd et al., 2013) and aggression (White et al., 2015, Muñoz et al., 2008). This study is the first to examine the relationship between CU traits and attentional biases to threat in adults. Although there was no evidence of a relationship between CU traits and attentional bias to threat using the DPT, positive correlations were found between CU traits, reactive aggression and SIP variables. Specifically, during the course of ambiguous social interactions, participants displaying higher CU traits were more likely to believe that other people behave in a way which enables them to achieve their own goals. Moreover, CU traits were only associated with appropriate responses (as opposed to aggressive or relationally aggressive responses) to ambiguous social interactions, with participants displaying higher CU traits reporting an increased likelihood of behaving in a socially appropriate manner, not because this response would be easy for them to enact but because they believed it would have the best outcome. This indicated that in ambiguous social interactions where there is no explicit threat, individuals with higher CU traits are likely to focus on others' goals and behave in a way that achieves their own. These findings are in line with previous research as Waschbusch et al. (2007) similarly found that children with conduct problems displaying higher CU traits generated fewer aggressive responses. This may however be due to study designs focusing on instrumental goals in ambiguous provocation scenarios. For instance, Helseth et al. (2015) used peer provocation scenarios and found that children displaying both conduct problems and higher CU traits generated significantly more aggressive responses than controls. Finally, there was no relationship between CU traits and hostile attribution which supports previous findings (Helseth et al., 2015).

In this sample, hostile attribution was not linked to reactive aggression. The relationship between aggression and hostile attribution in previous studies appears robust and has been demonstrated across age groups and population samples (see systematic reviews by Martinelli et al., 2018 and Klein Tuente et al., 2019). Further, a relationship between reactive aggression and hostile attribution is proposed within the SIP model (Crick & Dodge, 1994, Dodge et al., 1990). Specifically, individuals with a tendency to react aggressively when provoked were also expected to interpret the intentions of others as hostile in an ambiguous social interaction (Dodge et al., 1990). Some studies have indeed found evidence supporting a specific association between reactive aggression and hostile attribution (Lobbestael et al., 2013, Murray-Close et al., 2010). Conversely, other studies found contradicting results (Klein Tuente et al., 2019). For example, Helfritz-Sinville and Stanford (2015) found no association between reactive aggression and hostile attribution; however, groups of proactive and reactive aggressors displayed increased verbal and behavioural aggression compared to controls. Along with other studies revealing the relationship only in response to relational provocation (Murray-Close et al., 2010), these results suggest that associations between hostile attribution and reactive aggression may be contingent upon other factors such as impulsivity (Skeem et al., 2003) or impulsivity under negative affect (Gagnon & Rochat, 2017), anger rumination (Quan et al., 2019a), and psychopathy (Law & Falkenbach, 2018).

A positive relationship was found between reactive aggression scores and an increased emotional response, i.e. likelihood of feeling angry or embarrassed in ambiguous social situations. Additionally, individuals with higher reactive aggression scores perceived the behaviour of the antagonist as serving specific, instrumental goals, and self-reported being more likely to enact aggressive behavioural responses which were perceived as more acceptable and more likely to result in a positive outcome. State and trait anger have been linked to an increase in aggressive behaviour (Calvete & Orue, 2012). Moreover, difficulties in regulating emotion (or emotion dysregulation) have been related to increased engagement in aggressive (Preston & Anestis, 2020) and antisocial behaviour (Garofalo et al., 2020), psychopathy (Garofalo et al., 2018), and steps of the SIP (de Castro et al., 2005). In line with this research and the SIP model, it is possible that the endorsement of aggressive responses is mediated by the emotional state of the aggressor and their ability to control their emotional reactivity during the response generation and evaluation stages of their decision making.

2.5.1. Limitations

This study was the first to examine attentional bias expression over time using trial-level BS, computed as an alternative to the static, trait-like signal proposed by traditional BS in relation to measures of aggression, delinquency, and CU traits. A modified DPT was used, with exposure times randomly distributed throughout the task. Although these methods improved the reliability of the DPT as a measurement of attention allocation, they also limit comparisons with previous findings. Still, exploration of computational methods for the DPT has foregrounded the importance of questioning the reliability of reaction time tasks, which is seldom reported even in recent studies despite evidence to suggest reliability concerns apply to most methods of assessment (Rodebaugh et al., 2016). A cross-sectional design was used, meaning causal relationships could not be established, nor could the possibility of bidirectional associations between SIP stages and measures of aggression, delinquency and CU traits be dismissed. Finally, although the sample was heterogeneous with regards to age, race, and education, recruitment was aimed at individuals with a history of aggression which limits generalizability of findings. Whilst sample size was suitable for the planned analyses, exploratory analyses were likely underpowered and a comparison across community and forensic sample would help clarify the extent to which these findings are contingent upon a history of aggression.

2.5.2 Conclusions and future directions

Despite manipulating duration of exposure to angry and neutral faces, no attentional bias to threat was found in either 100 or 500 ms condition although the RT data indicates an increased engagement with briefly exposed stimuli. There was no evidence of a relationship between attentional bias as measured by the DPT and trial-level BS computations and measures of aggression and CU traits in this sample. Still, an attentional bias to angry faces was found in the low-delinquency group and a link of bias to response enactment for aggressive responses, suggesting later stages in the SIP model may have a mediating role in the relationship between encoding biases and behavioural aggression. Links between externalizing measures and social-information processing variables suggest that aggression and CU traits are linked to mechanisms processing information in social settings. Future research exploring the link between attentional biases (measured using the DPT) and externalizing measures should consider alternative BS computations and report reliabilities of measurements so as to allow comparisons. It is possible that the DPT is not a suitable measure of attentional bias during encoding; future studies should explore alternative tasks to measure encoding and preferential allocation of attention to threat using social stimuli. In this study links were found between aggression and stages of the SIP, including associations with emotional responses. Given that emotion is fundamental to the SIP model (Crick & Dodge, 1994) future studies should investigate the role of emotion processing and regulation in mediating the effect of SIP stages on aggression as a behavioural outcome.

3. Chapter 3: Study 2- Working memory guidance of visual attention to threat in offenders

3.1. Abstract

Aggression and trait anger have been linked to attentional biases toward angry faces and attribution of hostile intent in ambiguous social situations. Memory and emotion play a crucial role in social-cognitive models of aggression but their mechanisms of influence are not fully understood. Combining a memory task and a visual search task, this study investigated the guidance of attention allocation toward naturalistic face targets during visual search by visual working memory (WM) templates in 113 participants who self-reported having served a custodial sentence. Searches were faster when angry faces were held in working memory regardless of the emotional valence of the visual search target. Higher aggression and trait anger predicted increased working memory modulated attentional bias. These results are consistent with the Social-Information Processing model, demonstrating that internal representations bias attention allocation to threat and that the bias is linked to aggression and trait anger.

3.2. Introduction

The first study used the DPT, an established measure of attentional bias (Price et al., 2015), to investigate links between attentional bias to angry faces and measurements of antisocial traits and behaviours. There was no evidence of an attentional bias toward angry faces presented for either 100 or 500 ms. Further, no relationships were found between attentional bias toward angry faces and measures of aggression and CU traits. However, a relationship was found between attentional bias and delinquency, suggesting a link between a history of antisocial behaviour and early cognitive biases. This link is further explored in the second study, which investigated early and late social-information processing biases in a sample of former offenders. Besides the DPT, which is most commonly used to study attentional bias, visual search tasks have been used to investigate attentional bias to threatening targets placed amongst distractors (Wieser et al., 2018).

Visual search paradigms have been extended to real (as opposed to schematic faces) and studies have found faster RTs to angry (compared to happy) faces surrounded by distractors, termed the *"anger superiority effect"* (Pinkham et al., 2010). This effect is modulated by top-down mechanisms such as working memory (WM) and learned probability of events (Moriya et al., 2014). Further, information held in WM has been shown to act as a template, guiding visual attention toward stimuli matching the template (Soto et al., 2005). Drawing on WM and attention research, the second study used a visual search paradigm to investigate how WM guides visual attention to threat. This guidance was considered in the context of the SIP model and in relation to measurements of antisocial traits and behaviours. Drawing on previously reviewed literature on the relationship between social cognitive biases (attention and interpretation biases) and ASB, the focus shifted toward top-down (WM) guidance of visual perception and links to interpretation bias and aggressive outcomes.

Distinct patterns of information processing from encoding (Manning, 2020) to behavioural enactment (Repple et al., 2017) can be observed in individuals displaying higher aggression (Coccaro et al., 2017). Specific patterns may be associated with different forms of aggression (Elbert et al., 2017), associated behavioural outcomes (Wrangham, 2018), and distinct biases in social-information processing. For example, reactive aggression as a response to perceived provocation (Dodge et al., 2013) has been linked to hypervigilance to and difficulties in disengaging from threat during encoding and intent attribution. Proactive aggression, planned to achieve specific goals, has been linked to construction and selection of aggressive responses to ambiguous situations (Fontaine & Dodge, 2006, Lobbestael et al., 2016). Similarly, trait anger has been associated with cognitive biases (Maoz et al., 2017, Blair & Mitchell, 2009) and is likely to influence social-information processing.

In order to target interventions to specific information processing steps, it is necessary to understand the underlying cognitive mechanisms (Boxer & Dubow, 2002). Processing in the SIP model is hypothesised to be guided by emotion and a "database" encompassing memory, social schemas, acquired rules, and affect-event links (Fontaine & Dodge, 2006). Emotion and memory are therefore at the core of the SIP framework (Li et al., 2013). As a transition resource for mental structures (such as behavioural schemas and scripts), WM is directly involved in real-time processing (Fontaine & Dodge, 2006). However, the mechanisms through which WM influences information processing stages have not been elucidated. It is therefore necessary to examine WM biases of visual perception in the context of the SIP model.

WM templates as forms of internal representations or internal goals (Yao et al., 2019), precede information selection and processing (Battistoni et al., 2017). For example, everyday goal directed visual search tasks (e.g., searching for a face in a crowd) are typically preceded by a cue, such as knowing that the search target has dark hair or specific facial features. This cue in turn leads to the creation of an attentional template for the search goal (i.e., finding the face matching the description), which guides the visual search to match the attentional template (Soto & Humphreys, 2008). Such a template helps to manage the overwhelming amount of information available in the environment by providing top-down guidance of attention allocation to memory-matching, task-relevant stimuli (Carlisle, 2019). WM guidance of perception can become disadvantageous when cognitive resources are allocated toward threatening stimuli as distractors (Salahub & Emrich, 2020) matching WM templates (van Moorselaar et al., 2014).

Biased WM guidance of attention may contribute to a tendency to over allocate attention to threatening stimuli. Trait anxiety has been linked to attentional biases toward threat

(Cisler & Koster, 2010). Similarly, aggression has been linked to an increased attentional bias towards hostile stimuli (Manning, 2020). However, aggression related attentional bias research has addressed the possibility of attention allocation being guided by aggressive internal representations. Yao et al. (2019) found that highly anxious participants showed an increased attentional bias toward threatening faces when holding an angry face in mind. WM guidance of attention has not been explored in aggression.

Using a combined WM and visual search task, this study examines the extent to which WM modulated attentional bias in a former offender sample. There is a higher incidence of aggression-related knowledge structures (i.e., schemas, internal representations) in offender populations (Day et al., 2021) shown to account for almost a quarter of variance in aggression scores (Gilbert et al., 2013). Internal representations held in WM affect encoding, processing, and integration of information (Ogilvie et al., 2011). Given this guidance effect (Kumar et al., 2009, Carlisle & Woodman, 2013), social stimuli were expected to be held in WM as templates to constrain allocation of cognitive resources to stimuli matching internal representations. Additionally, participants were expected to preferentially encode (Burra & Kerzel, 2019) and maintain angry faces in WM (Jackson et al., 2014) thereby tending to allocate cognitive resources toward angry or hostile stimuli. Lastly, unmapped links were explored between WM modulations of visual attention and later stages in social-information processing. In line with the SIP model (Van Nieuwenhuijzen et al., 2017) an increased bias toward potentially threatening stimuli was expected to be predicted by higher aggression, trait anger, and hostile attribution bias.

3.3. Method

The method, data preparation protocol, and data analysis plan were pre-registered with the Open Science Framework (OSF, link available here). This study was approved by the

University of Sheffield Ethics Committee. Written informed consent was obtained and the data were analysed anonymously.

3.3.1. Participants

Power analysis assuming a small to medium effect size of d= 0.40 (for the expected main effect of congruency (H1)- in line with findings from Yao et al., 2019) showed that a sample size of 70 was recommended to achieve power of .95 (α= 0.05, one-tailed). As participants with less than 70% correct trials on the visual task were excluded from the final analysis, participants were recruited beyond the recommended sample size (N= 131) to compensate for potential exclusions. Following exclusions, data from 113 participants were analysed (M_{age}= 41.29, SD_{age}= 12.23; 33.63% female). In the final sample, 85% of the participants were White, 7% were Black, and the remaining 8% reported mixed, Asian, or other ethnic backgrounds. With RTs to the dual task as DV, the power levels afforded (α= 0.05, one-tailed) for the expected interaction (congruency and emotional valence of the WM template) in the current sample size for a small (f= 0.10), medium (f= 0.25), and large (f= 0.40) (Cohen, 1988) effect were .18, .75, and .99 respectively. The power level afforded to detect a medium effect size f= 0.15 for the composite reactive aggression and anger variable and hostile attribution in the current sample was .85. Participants were recruited via Prolific (www.prolific.co) on the basis of being English speakers, based in the United States or United Kingdom, and answered “Yes” to the custom screening: “Have you ever been in prison for committing a crime?” offered by Prolific.

3.3.2. Measures

Social-information processing patterns were measured using an abbreviated version of the Social Emotional Information Processing Questionnaire (SEIP-Q, Coccaro et al., 2017).

SEIP-Q variables were assessed in relation to vignettes comprising ambiguous scenarios in which Person A is the victim of Person B's adverse action. The vignettes were designed to contain either direct aggression (e.g., physical aggression) or relational aggression (e.g., rejection). Participants were required to identify with Person A. Attribution was assessed using four Likert-scale questions measuring: direct hostile intent (e.g., "This person wanted to damage my car"), indirect hostile intent (e.g., "This person wanted me to feel unimportant"), instrumental non-hostile intent (e.g., "This person was in a hurry to get in to work"), and benign intent (e.g., "This person scratched my car by accident and didn't notice"). Negative Emotional Response was measured using two Likert-scaled questions (e.g., How likely is it that you would be angry if this happened to you"). Participants were asked how likely they would be to enact a behavioural response that is either socially appropriate, overtly aggressive or relationally aggressive. SEIP-Q scores for each construct were averaged on the same 0-3 scale. In this sample, alpha coefficients were strong for hostile attribution (α= 0.87) and negative emotional response (α= 0.79) but lower for benign attribution (α= 0.47) and instrumental attribution (α= 0.56).

Reactive and Proactive aggression were measured using the Reactive-Proactive Aggression Questionnaire (RPQ) (Raine et al., 2006, see the Measures sub-section in study 1, Chapter 2 for a full description) comprising 23 items measuring proactive (12 items) and reactive aggression (11 items).

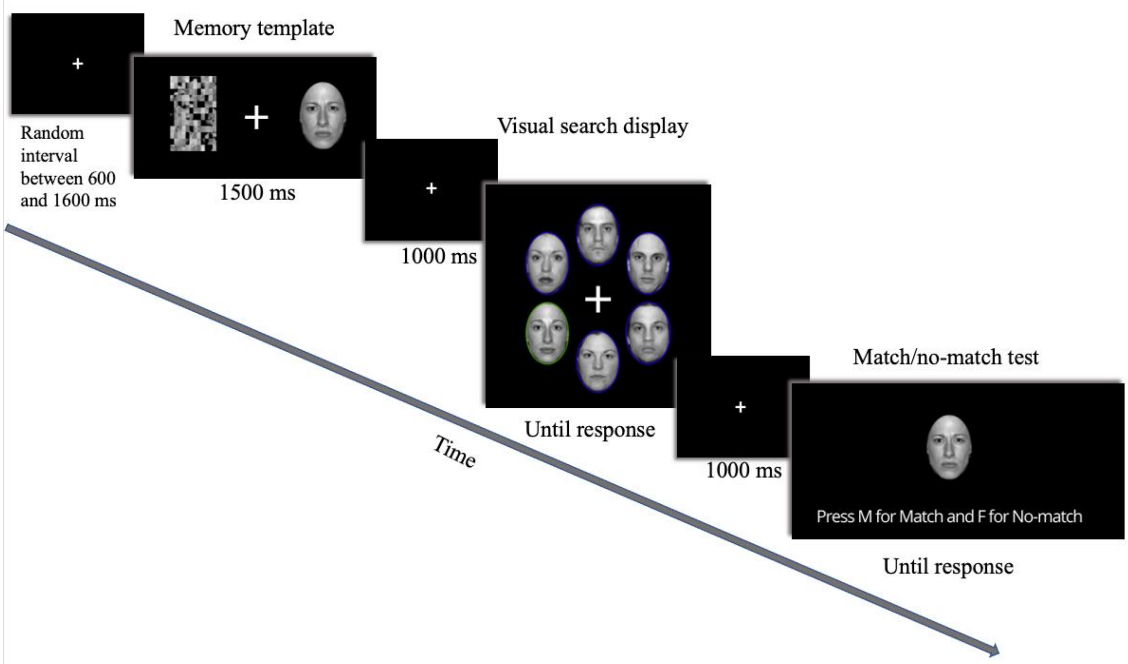
Trait Anger was measured using the 10-item Trait Anger subscale of the State-Trait Anger Expression Inventory–II (Spielberger, 1999). Items were scored from 1 ("Almost never") to 4 ("Almost always"). Trait anger correlates with behavioural aggression (Wang et al., 2018) and has demonstrated high reliability and validity (including concurrent validity) across clinical and non-clinical samples (Lievaart et al., 2016).

3.3.3 Dual task

The Gorilla Experiment Builder (www.gorilla.sc) was used to collect online reaction time data. Following Burra et al. (2017), attentional bias to threat was inferred from reaction times (RTs) on a dual task, consisting of a working memory task and a visual search task. As depicted in Figure 6, a memory template consisting of an angry or neutral face oval was initially presented lateral to a fixation point. The position opposite the face oval contained a scrambled version of the face oval, balancing the visual display (Salahub & Emrich, 2020). This was followed by a visual search array consisting of six faces (target and five other neutral faces, 3 male and 3 female faces, identities and gender randomly allocated in the display). Following the visual search array, participants were presented with a match/no-match test display and asked whether a face matched the memory template displayed at the beginning of the trial. The target was always presented laterally (never at vertical midline) and consisted of a face oval surrounded by an unfilled oval (coloured oval shape having a width of 20 pixels) whose colour was distinct from the remaining 5 ovals. The unfilled colour oval surrounding the target varied randomly between green (with the remaining neutral faces in the visual display being surrounded by blue) and blue (with the remaining neutral faces in the visual display being surrounded by green). The memory template was identical in valence and identity to the target for half of the trials (congruent condition) and had the same identity and a different emotional valence from the target for the other half (incongruent condition). The remaining neutral faces in each visual display were randomly extracted from the pool of neutral face ovals (excluding the target) and a random list generator assigned visual search display positions. Participants were required to memorize the face in the memory template, identify the gender of the target in the visual search display, and identify whether the face in the match/no-match test was the same as the memory cue. Further detail on the development and validation of the dual task can be found in Appendix A.

Figure 6

Dual task trial example



3.3.4. Analysis plan

A cross-sectional study design was used to investigate the associations between WM guidance of visual search, steps in the SIP model, and self-report measures of aggression, trait anger, and CU traits. A 2 x 2 repeated measures design was used with within-participants factors of: WM template and target congruence (whether the emotional valence of the WM template matched that of the target, i.e., congruent vs incongruent), and the emotional valence of the memory template (angry or neutral) as IVs and RTs as DV. Bias scores were used as DVs in hierarchical models with reactive and proactive aggression scores, trait anger, CU scores, and scores on the SEIP-Q as predictor variables.

I. Do WM templates guide visual search for naturalistic faces and is there an effect of the emotional valence of the WM templates?

A two-way repeated measures ANOVA with RTs (latency) to targets in visual search displays as the DV and trial congruency and emotional valence of the WM template (angry vs neutral) as IVs was expected to reveal an attentional bias toward congruent compared to incongruent stimuli. An interaction effect was also expected, whereby participants would display faster RTs to trials in which both WM templates and targets consist of angry faces (congruent & angry) compared to trials in which WM templates and targets consist of neutral faces (congruent & neutral).

II. Are WM visual selection biases predicted by self-reported aggression and trait anger?

Individuals displaying higher aggression and trait anger scores holding an angry face template in WM were expected to display an attentional bias toward a matching angry target and be slower to find (or demonstrate a bias away from) a neutral target preceded by an angry attentional template. This effect was expected to be specific to angry faces and therefore absent when attentional templates were neutral. Hierarchical regressions were planned to evaluate the prediction of attentional bias (BS1 and BS2 as DVs), from aggression subfactors and trait anger (as predictors). Reactive aggression and trait anger scores were significantly correlated ($r= 0.64$). Similarly, response enactment for relationally and directly aggressive behavioural responses were significantly correlated ($r= 0.63$). Composite variables were created by averaging reactive aggression and trait anger for the former and response enactment scores for the latter. Alpha coefficients for the composite variables were excellent ($\alpha= 0.88$ for the anger/reactive aggression composite and $\alpha= 0.87$ for the response enactment for aggressive responses composite). Significant correlations were found between anger, aggression, and SEIP-Q variables. Consequently, the pre-registered relationship between hostile attribution and BS was investigated separately, within the regression models with anger and aggression as predictor variables. This enabled the examination of the relationship between aggression, anger, and BS before controlling for hostile attribution. The hierarchical regression model predicting BS added age, gender, and education at step 1. Composite reactive aggression and trait anger scores were added at step 2. Proactive aggression was added at step 3 and hostile attribution was added at step 4.

3.3.5. Stimuli

Task stimuli for both WM and visual search tasks comprised angry and neutral faces (6 male, 6 female) selected from the NimStim stimulus set (Tottenham et al., 2009). Following previous visual search studies faces with visible teeth were discarded, selected face ovals were cropped at or near the hairline, and all pictures were converted to grey-scale (Burra et al., 2017). Given that pictures of faces are prone to low-level confounds, likely to influence visual search performance, the contrast and luminance histograms of the pictures were equalized using the SHINE Matlab toolbox (Willenbockel et al., 2010). Each face oval took up 20% of the screen height and 14% of width. Each face oval and scrambled counterpart in the memory template was placed 5% of the screen away from the fixation cross and face ovals in the visual search displays were placed 10% of the screen away from the fixation cross.

3.3.6 Procedure

Each trial began with a white fixation point on a black background, presented for a random interval between 600 and 1600 ms (Burra et al., 2017). Participants were presented with the memory template for 1500 ms which they were asked to memorise (Yao et al., 2019). This was followed by a fixation point displayed for 1000 ms which was subsequently replaced by the visual search array, displayed until response. Participants were required to identify the target, which was surrounded by a colour singleton distinct to the colour surrounding the distractors and report the target face gender as quickly and accurately as possible. Participants were instructed to use their left index finger to press “F” if the face oval is female and the right index finger to press “M” for male face ovals. In each trial, the visual search display was followed by a fixation cross for 1000 ms and a match/ no-match test whereby a face oval was presented at the centre of the screen and the participants were required to answer whether the face matched the memory template by pressing the “F” key for “no-match” and the “M” key for “match”. Before the experiment, participants completed 20 practice trials with on-screen feedback. If fewer than 14 face genders (70%) were correctly identified during practice (identification of the gender of the target), the participant was asked to read the instructions for the task again and complete another 15 practice trials. Each participant completed 192 trials in two blocks of 96. Participants completed the dual task followed by the self-report measures in the order listed above.

3.3.7. Data Preparation

Data preparation and planned analyses were pre-registered with the OSF. Callous-unemotional (CU) traits were included in the pre-registration and differential patterns of WM modulated attention allocation associated with callous-unemotional traits are addressed in Appendix B. High correlations were found between self-reported aggression, anger, and SEIP-Q response enactment variables (see Appendix B). As hypotheses were pre-registered for the relationships between hostile attribution, anger, aggression, and WM biases of visual perception, hostile attribution was added as a predictor in regression models with aggression and anger as predictors. The extended pre-registered exploration of links between WM modulations of visual attention and patterns of social-information processing is included in Appendix C.

Analyses were performed using RStudio 4.0.2 (RStudio, 2020). As planned, outliers were Winsorized; values outside 1.5 interquartile ranges from the Tukey Hinges (lower and upper hinge corresponding to the first and upper quartiles or 25th and 75th percentiles respectively) were rescaled to the last valid value within the range. This approach improves score reliability across a range of attentional bias task modifications (Price et al., 2015). There were no missing trials in the combined visual task. Following Burra and Kerzel (2019), participants with < 70% correct trials on either visual search or match/no-match test were excluded from analysis (N=18). Incorrect responses on both visual search and match/no-match were removed from the dataset analyses (Soto et al., 2012). Following Burra et al. (2017), RTs under 200 ms were deemed unlikely to reflect genuine responses and were therefore removed from the combined visual task dataset. RTs on the visual search task and the match/no-match test were extracted and analysed separately.

Missing values for the self-report measures were inspected using heat maps. No participants were missing > 20% of questions within a self-report measure. Examination of histograms plots, box plots, and z scores for non-normality indicated that proactive aggression was significantly positively skewed. These scores were converted to a binary variable which distinguished at 0 from scores above 0 (coded "1"). Response enactment scores for directly aggressive responses were positively skewed which was removed by square root transformation.

3.3.8. Computation and reliability of attentional bias data

BS were computed by subtracting mean RTs/ experimental condition of interest from mean RTs for congruent and neutral trials (see Table 6). Using 5000 random splits, Spearman-Brown corrected reliability estimates found low-to-medium split-half reliabilities of the BS. Alternative trial-level BS (Zvielli et al., 2015) were computed following the pre-registered analysis plan. Using the weighted trial method, a time-series of trial-level BS per participant was produced by subtracting the RTs for each trial from the weighted mean of all trials of the opposite type (i.e., baseline trials comprising congruent and neutral trials, see section 2.2.5. for more detail on trial-level BS computations). Table 6 shows that split-half reliabilities for trial-level BS were considerably higher than the reliabilities of the BS, trial-level BS were henceforward used as measurement of attentional bias, and were referred to as BS (bias scores). $BS_{POSITIVE}$ (mean of *positive* trial-level BS, indicating attentional bias toward target stimuli) and $BS_{NEGATIVE}$ (mean of *negative* trial-level BS, indicating attentional bias away from target stimuli) scores were calculated for each of the three conditions above.

Table 6

Bias scores: computation and reliability

Bias score and computation	Spearman-Brown (r_{SB}) for traditional BS	Spearman-Brown (r_{SB}) for Trial-Level BS
BS1: Comparing $RT_{congruent\&neutral}$ to $RT_{congruent\&angry}$	0.24	0.81
BS2: Comparing $RT_{congruent\&neutral}$ to $RT_{incongruent\&angry}$	0.43	0.86

Note. BS= Bias Scores.

BS for neutral targets identified whilst holding an angry face in WM ($BS2_{POSITIVE}$) were highly correlated to the remaining BS and were excluded from further analyses. Regression models having the excluded BS as DV did not have any significant predictors beyond the predictors for retained BS which are summarised below. BS1 scores, indicating attentional bias toward and away from angry targets identified whilst holding an angry face in WM (i.e., where the target was identical to the WM template, relevant to the second and third hypotheses) and the $BS2_{NEGATIVE}$ scores (i.e., where the target was different in emotional valence to the WM template, relevant to the second and third hypotheses), indicating an attentional bias away from neutral targets identified whilst holding an angry face in mind were retained.

3.4 Results

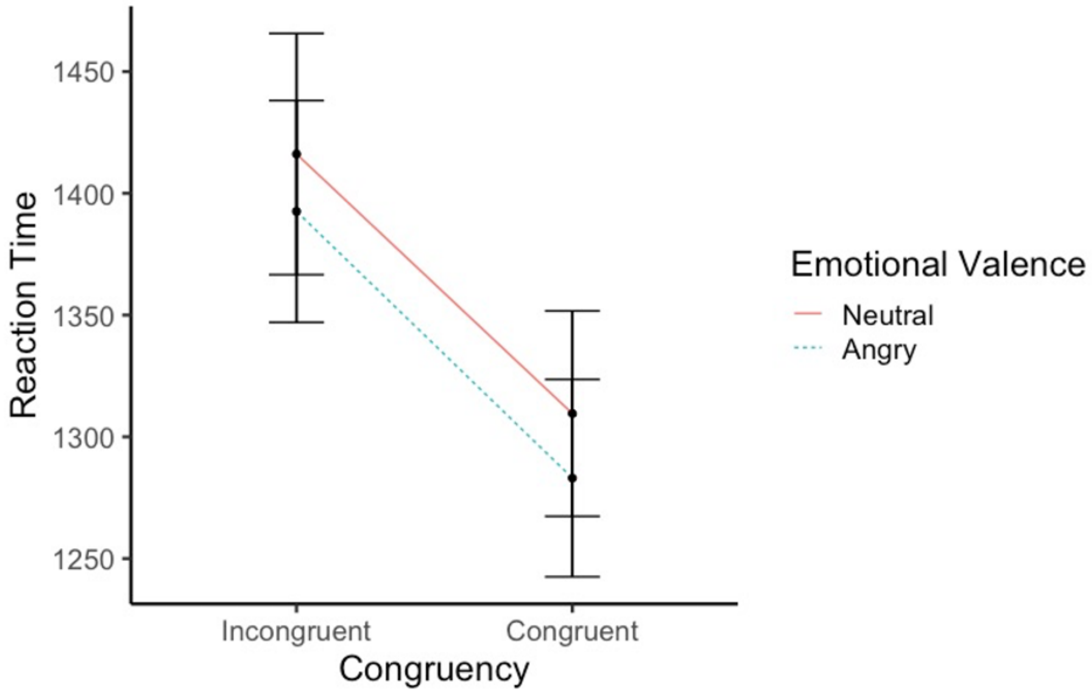
3.4.1. Results relating to pre-registered hypotheses

I. Do WM templates guide visual search for naturalistic faces and is there an effect of the emotional valence of the WM templates?

As expected, the ANOVA showed a main effect of congruency ($F(1, 112)= 75.7$, $p<0.001$, $\eta_p^2= 0.39$) with faster RTs to congruent (M= 1296.3, SD= 438.92) than incongruent trials (M= 1404.07, SD= 504.84). A further main effect of emotional valence of the memory template was found ($F(1, 112)= 6.68$, $p= 0.01$, $\eta_p^2= 0.05$) with faster RTs when visual search displays were preceded by angry memory templates (M= 1337.82, SD= 460.46) rather than neutral memory templates (M= 1362.85, SD= 490.96). Contrary to the hypothesis, there was no interaction between congruency and emotional valence of the WM template ($F(1, 112)= 0.02$, $p= 0.87$, see Figure 7).

Figure 7

Changes in reaction times as a function of congruency and emotional value. Error bars show one standard error above and below mean.



II. Are WM visual selection biases predicted by self-reported aggression, trait anger, and hostile attribution?

Hierarchical regression modelling is summarised in Tables 7-9. Composite reactive aggression and anger positively predicted attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$, β = 0.22, p= 0.03, model at step 2) and negatively predicted attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$, β = -0.19, p= 0.05). Reactive aggression and anger scores did not predict bias towards or away from targets after proactive aggression was added. There were no significant anger or aggression predictors of $BS2_{NEGATIVE}$ scores; however, hostile attribution scores were a significant negative predictor of attentional bias away from neutral faces identified whilst holding an angry face in mind ($BS2_{NEGATIVE}$, β= -0.21, p= 0.04, model at step 4).

Table 7

Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$).

	β	SE	p	R^2 Total/Change
Step 1				0.003
Gender	-0.08	0.10	0.41	
Education	0.06	0.10	0.53	
Age	-0.14	0.10	0.15	
Step 2				0.04/0.028*
Gender	-0.03	0.09	0.75	
Education	0.06	0.09	0.52	
Age	-0.07	0.10	0.48	
Reactive aggression and anger	0.22	0.10	0.03*	
Step 3				0.04/0.29
Gender	-0.03	0.10	0.71	
Education	0.06	0.09	0.51	
Age	-0.07	0.10	0.48	
Reactive aggression and anger	0.17	0.11	0.11	
Proactive Aggression	0.11	0.10	0.29	
Step 4				0.05/ 0.20
Gender	-0.03	0.09	0.77	
Education	0.04	0.09	0.72	
Age	-0.07	0.10	0.47	
Reactive aggression and anger	0.14	0.11	0.20	
Proactive Aggression	0.11	0.10	0.30	

Hostile Attribution	0.12	0.10	0.20

Note. β= standardized regression coefficient, SE= Standard Error

**p* < .05.

Table 8

Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$).

	β	SE	p	R^2 Total/Change
Step 1				0.03
Gender	0.15	0.09	0.10	
Education	-0.02	0.09	0.84	
Age	0.19	0.10	0.05	
Step 2				0.06/0.05*
Gender	0.11	0.10	0.25	
Education	-0.01	0.09	0.84	
Age	0.12	0.10	0.22	
Reactive aggression and anger	-0.19	0.10	0.05*	
Step 3				0.05/0.41
Gender	0.11	0.10	0.24	
Education	-0.006	0.09	0.95	
Age	0.13	0.10	0.21	
Reactive aggression and anger	-0.16	0.11	0.14	
Proactive Aggression	-0.09	0.10	0.41	
Step 4				0.06/0.17

Gender	0.12	0.10	0.23	
Education	-0.008	0.09	0.93	
Age	0.12	0.10	0.22	
Reactive aggression and anger	-0.17	0.11	0.12	
Proactive Aggression	-0.09	0.10	0.34	
Hostile Attribution	-0.13	0.10	0.17	

Note. β= standardized regression coefficient, SE= Standard Error

*$p < .05$.

Table 9

Hierarchical multiple linear regression analysis predicting attentional bias away from neutral faces whilst holding an angry face in WM ($BS2_{NEGATIVE}$).

	β	SE	p	R^2 Total/Change
Step 1				0.02
Gender	0.05	0.09	0.54	
Education	-0.06	0.09	0.50	
Age	0.04	0.09	0.64	
Step 2				-0.009/0.18
Gender	0.03	0.10	0.78	
Education	-0.06	0.10	0.50	
Age	0.001	0.10	0.99	
Reactive aggression and anger	-0.14	0.10	0.18	
Step 3				-0.006/0.24
Gender	0.03	0.10	0.76	

Education	-0.05	0.10	0.63	
Age	0.004	0.10	0.97	
Reactive aggression and anger	-0.09	0.11	0.44	
Proactive Aggression	-0.12	0.11	0.24	
Step 4				-0.01/0.03*
Gender	0.02	0.10	0.82	
Education	-0.03	0.10	0.75	
Age	0.00	0.10	0.99	
Reactive aggression and anger	-0.03	0.11	0.76	
Proactive Aggression	-0.12	0.10	0.25	
Hostile Attribution	-0.21	0.10	0.04*	

Note. β, standardized regression coefficient, SE= Standard Error

*p < .05.

3.4.2. Exploratory analysis

III. Match/no-match test accuracy: exploratory analysis

As a proxy measure of the quality of WM templates, match/no-match test accuracy (measured as percentage of correct trials) was introduced as a DV in a multiple linear regression having emotional valence of the memory template (angry vs neutral) and congruency (congruent vs incongruent) as predictors. Emotional valence positively predicted accuracy (β= 0.45, p< 0.01) with higher accuracy for angry (*M*= 82.72, *SD*= 13.64) compared to neutral faces (*M*= 76.49, *SD*= 15.7).

3.5. Discussion

This study examined how emotional faces as WM templates biased visual search and whether this bias was predicted by self-reported aggression, trait anger, and social-emotional information processing variables. As expected, WM templates guided attention during visual search for naturalistic face targets. Increased aggression and trait anger predicted increased WM attentional bias. Bias was further linked to self-report measures and key constructs in social-information processing, thus tapping into a previously unexplored link between WM templates and SIP stages. The combined visual task utilised in this study has not been employed in previous research; thus, the results build on the attention as well as the antisocial behaviour literature to spotlight specific relationships between cognitive biases and measures of aggression and trait anger.

Do WM templates guide visual search for naturalistic faces and is there an effect of the emotional valence of the WM templates?

This study examined the effect of emotional faces as WM templates on visual search. Consistent with predictions, faster reaction times were found for congruent visual search targets (i.e., targets identical to the templates held in WM), demonstrating a direct effect of emotional faces held in WM on visual search. Similar effects of congruency, whereby participants respond faster to targets sharing features with WM templates, have been found in previous studies examining WM biases of visual perception (Mazza et al., 2011), including in forensic populations (Krusemark et al., 2016). The present study was the first to extend this effect to naturalistic, emotional faces. Moving beyond low-level stimuli and feature based search (i.e., shapes and lines) (Pan & Soto, 2010), current findings provide further evidence that faces are likely to be encoded as integrated representations. That is, the faces presented in the visual search display and as WM templates were grayscale ovals and luminance and contrast were averaged across stimuli used in the task, thereby reducing the effectiveness of potential feature-based searches (Berggren & Eimer, 2018). Consequently, the WM template (or internal representation) was likely encoded and maintained in WM as a single object which effectively guided visual search.

Following the SIP model (Moriya et al., 2014), it was expected that visual search for an angry face whilst holding an angry face in WM (i.e., a negatively valenced internal representation) would be more efficient than visual search for a neutral face whilst holding a neutral face in mind. There was no supporting evidence for this prediction. There was however an effect of emotional valence of the WM template on visual search. Specifically, when participants held an angry face in WM, they responded to targets in visual search significantly faster than when holding a neutral face in WM, regardless of the emotional valence of the target. Participants were also significantly more likely to remember angry over neutral faces for the match/no-match test, suggesting a partial processing advantage for emotional compared to neutral faces.

Evidence of an anger superiority effect has been found across multiple populations including undergraduates (Fan et al., 2016, Moriya et al., 2014), community samples (Burra & Kerzel, 2019), anxious (Bar-Haim et al., 2007, Wieser et al., 2018) and aggressive populations (Taylor & Jose, 2014). In this study, the preferential encoding and maintenance of aggressive faces into WM was followed by an enhanced performance for all targets within the visual search task. Drawing on the SIP model, the findings demonstrate an overall WM bias of visual perception by emotional faces; however, the effect was not limited to angry faces as expected. Instead, an emotional (here negatively valenced) WM template appeared to broadly (and positively) bias visual search, increasing effectiveness and distractor suppression during visual search. In other words, holding a negative internal representation enabled participants to meet task goals effectively.

Are WM biases of visual selection predicted by self-reported aggression, trait anger, and hostile attribution?

As expected, an increase in reactive aggression and anger (composite score) predicted an increased attentional bias towards angry targets identified whilst holding an angry template in WM. Lower aggression and anger scores predicted increased avoidance of angry faces whilst holding an angry face in WM. Trait anger (Maoz et al., 2017) and aggression (Crago et al., 2019) have been linked to attentional biases in previous research. This study further showed that variation in aggression and trait anger also predicted WM modulated biases of visual search. In line with the SIP model (Dodge et al., 2013), reactive aggression and trait anger scores did not predict increased attentional bias to angry compared to neutral faces when proactive aggression was controlled. Although the SIP further postulated a link between reactive aggression and hostile attribution (Fontaine & Dodge, 2006) there was no evidence for this association in the present research. As previously suggested by Oostermeijer et al. (2016), these findings indicate SIP mechanisms may not be differentiated for aggression subtypes, which can exert an influence on both early and late steps in social-information processing. A lower hostile attribution bias was associated with avoidance of neutral targets when holding an angry face in WM. These findings are consistent with the SIP model (Dodge & Rabiner, 2004), specifically that increased aggressiveness and hostile attribution predicted increased WM modulated attentional bias towards emotional (here angry) and away from neutral faces respectively.

Behavioural studies have demonstrated robust WM biases of visual search using low-level stimuli (e.g., shapes and lines, Soto & Humphreys, 2007). However, in a social environment, social content must be held in WM to facilitate social exchanges (Yao et al., 2019). Moreover, WM is fundamental to effective communication and is particularly involved in social interactions when decoding and interpreting others' emotions and intentions (Maran et al., 2015). Angry faces are recognised more accurately (Jackson et al., 2009), preferentially encoded (Fan et al., 2016) and maintained in working memory (Jackson et al., 2014). This study constitutes a first step in understanding the effect emotional faces held in WM have on visual search, suggesting that the bias toward emotional faces is amplified by trait anger and aggression whereas lower aggression, anger, and hostile attribution predicted avoidance of targets. Angry templates were preferentially encoded and maintained in WM compared to neutral. In line with models conceptualising WM as a common resource, dynamically distributed according to prioritisation of salient stimuli (Bays et al., 2009), our findings indicate a greater proportion of cognitive resources was drawn to emotional faces. Further, holding an angry face in WM resulted in enhanced processing of both angry and neutral targets. This broad bias may be specific to goal-directed visual search for emotional naturalistic stimuli.

Threat detection is fundamental to survival. According to the threat capture hypothesis, threatening stimuli are automatically detected, employing early mechanisms which are independent of cognitive control (Öhman & Mineka, 2001). Soto et al. (2005) found that an attentional template held in WM can guide "early parts of the search process in an involuntary manner" (p. 260). The WM bias of visual attention found here may reflect a WM guided involuntary deployment of cognitive resources to the detection of an emotionally valenced target in a social environment which would match the WM template. However, Burra and Kerzel (2019) demonstrated that attentional biases to threat are only partially automatic and dependent upon context and task demands. Indeed, the main effect of congruency found in this study, with neutral targets found faster when holding a neutral attentional template in WM, supports the hypothesis that top-down attentional biases can extend to processing neutral stimuli when required by task demands. Furthermore, from a SIP perspective, it may be that holding an emotionally valenced internal representation- in this case, an angry face- makes detecting a matching emotion in a social environment task-relevant. Consequently, visual search would be increasingly efficient which may explain our observed enhanced visual search performance whilst holding an angry face in WM. This study focussed on angry and neutral WM templates and targets. Future research should investigate the effect of angry faces in relation to other emotions as well as neural events corresponding to these effects using the contralateral delay activity, an event-related potential component indicative of working memory maintenance (Salahub & Emrich, 2020).

Within the SIP model, higher aggression is associated with a higher prevalence of biased (negative) internal representations expected to guide attentional resources towards threatening stimuli and increase attribution of hostile intent during ambiguous social interactions (Fontaine & Dodge, 2006). In a recent paper, a neural activity boost was found during maintenance of self-associated (compared to other-associated) information in WM, indicative of prioritization by top-down attention (Yin et al., 2021). This supports the above link between aggression and bias within the SIP, namely, that individuals displaying higher aggression or trait anger are more likely to allocate disproportional amounts of cognitive resources to hostile stimuli which are viewed as threatening.

In this study, the WM modulated bias of visual perception was further associated with hostile attribution, linking top-down biases of attention to interpretation of intent. Hostile attribution was positively linked to anger but not to reactive aggression. Further links to anger and later stages in the SIP are detailed in Appendix B (Table B4). Hostility biases, referring to the tendency to attribute hostility in social interactions, have been linked to perceptual biases (Miller & Johnston, 2019) and aggression (Bushman, 2016), pointing toward a general hostility bias mechanism Smeijers et al. (2019). Our findings support the hypothesized links between hostile attribution and the proposed SIP model core that consists of acquired rules, memory, and social schemas. Within the model proposed by Smeijers et al. (2019), SIP stages are mapped onto the hierarchical Gaussian filter (Mathys et al., 2011) and tenets of the free-energy principle (Friston, 2010). In other words, memories, rules, and social schemas form complex blueprints of social interaction, optimized to reduce uncertainty about the world. For example, a person who has repeatedly experienced a certain ambiguous social situation as resulting in an aggressive outcome (e.g., a heated argument ending in a fight) has adjusted their belief and will display a general tendency to attend to and interpret specific cues as indicating similarities between the given situation and the blueprint.

This line of inquiry leads to two important considerations. The first refers to the acquisition and reinforcement of the information at the core of the SIP model. Individuals displaying higher anger and aggression also report higher anger-rumination tendencies (Gerin et al., 2006). These comprise rehearsing and dwelling upon hostile information which are believed to reinforce tendencies toward anger and aggression (Wilkowski & Robinson, 2010). In the present study I have addressed how an induced representation can bias visual selection. That increased aggression and trait anger predicted an increased bias toward emotional faces could also be explained by an increased tendency toward anger-rumination. The latter could facilitate encoding of hostile cues, leading to an increased focus on hostile information held in WM and consequently to a stronger modulation of visual selection. In the context of the SIP, increased rumination would also lead to increased overall selectivity for hostile cues and reinforcement of hostile biases.

The second consideration refers to the applicability of a general hostility tendency to other sensory modalities and stimulus categories. Previous research has linked aggression and trait anger to biases toward semantic threat. For instance, individuals displaying higher anger tendencies were slower to disengage from hostile words (Wilkowski et al., 2006) and exhibited difficulties in processing non-hostile information in ambiguously hostile visual scenes (Wilkowski et al., 2007). Using neuroimaging data and an emotional word Stroop task, Buades-

Rotger and Krämer (2018) found that attentional bias for antisocial words predicted aggression and that this relationship was fully mediated by amygdala reactivity to angry faces. The present study has also demonstrated a relationship between WM modulated attentional bias and hostile attribution bias. Overall, these findings provide support for a database of memory and schemas guiding cognitive processes across sensory modalities (Soto & Humphreys, 2008) whilst demonstrating the relative advantage of stimuli likely to be present and form part of a social interaction. That is, patterns of mediation may be completely different for neural activation to angry faces when using auditory semantic antisocial stimuli (i.e., direct provocation) compared to read words (i.e., inferred threat).

3.5.1 Limitations

The findings should be considered in the context of some limitations. Preparation of the task stimuli aimed to reduce low-level confounding effects (e.g., contrast, luminance, Willenbockel et al., 2010) which may have reduced ecological validity. Using natural faces as stimuli may have added noise due to physical differences. This concern was mitigated by presenting perceptually balanced real faces, demonstrating that WM modulated attentional bias was not due to low-level confounds, as proposed elsewhere (Coelho et al., 2010). Additionally, physical differences in naturalistic faces are present in daily interactions and their removal from experimental designs (e.g., by using schematic drawings of faces, Burra et al., 2016) results in a penurious summary of the allocation of cognitive resources during social interactions.

The present research further found significant relationships across variables of interest, good-to-excellent reliability for attentional BS, and self-reported measures showed reliability indices comparable to previous studies. Although later stages of the SIP model have typically been investigated using vignettes to illustrate ambiguous social scenarios and questionnaires to assess potential emotional and behavioural responses (Smeijers et al., 2019), future studies should seek to reproduce these findings in paradigms with higher ecological validity, i.e. measurement of responses to ongoing provocation. Composite scores were used in this study due to high correlations between trait anger and reactive aggression as well as between response enactment scores for relationally and directly aggressive responses (see Appendix B). Whilst the composite variables had high reliabilities, these findings raise the issue of considerable conceptual overlap between established measures of trait anger and aggression, and between aggression subtypes in the context of social-information processing stages.

Collecting reaction times data over the internet is a potential limitation of this study as the experiment was completed across a range of devices and visual display settings. Online data collection has increased considerably in the last decade and the quality of such data, albeit requiring careful examination (Armitage & Eerola, 2020), has been found to be acceptable (Miloff et al., 2015). Moreover, results were compatible with previous studies using similar paradigms. Online data collection enabled recruitment of a wide range of former offenders. However, the participants' criminal history and specifically whether they have a history of violence is unknown. Also, the lack of a control group means that no inferences can be made regarding whether these results are specific to a population with an incarceration history. Thus, future replication in distractor free laboratory-based experiments and extension to non-offender and offense-specific samples would be valuable.

The task used in this study, which combined a WM and a visual search task, the identity of the WM template and that of the target were always the same, providing an overlap between WM content and targets of visual attention. Given that RTs were measured in relation to participants' identification of the gender of the target, it was possible that participants reacted to the gender of the face held in WM, therefore bypassing the visual search aspect of the task. However, the task was cognitively demanding and there was no evidence of participants having bypassed the visual search component to react to the gender of the WM template (i.e., a pattern of learning in the RT data). RTs would become faster with practice as participants become reliant on the gender of the WM as being the correct response during the visual search array. No such pattern emerged from a visual inspection of RT plots for each participant indicated stable patterns throughout the task. Moreover, a bypass of the visual search component would imply no relationship between the emotional valence of the WM template and that of the target in the visual search display. However, an effect of congruency was found; faster RTs to targets in the visual search array matching the emotional valence of the WM template. This demonstrated identification of target and sufficient engagement to detect congruency.

Whilst there is evidence of engagement with both components of the task, a second issue to consider is whether participants engaged with the search aspect of the visual search task. The visual search and associated gender identification task were in line with previous work (Burra et al., 2017). The contrast and luminance histograms of the pictures were equalized, minimising feature-based search; instead, attention was drawn to the targets using colour singletons (i.e., bottom-up salience; Wolfe & Horowitz, 2017). Consequently, if the engagement with the visual search array observed in the data was not due to top-down modulation of visual search (i.e., search of the target and identification of gender, as proposed), it may be due to bottom-up salience of the colour singleton. Whilst this seems unlikely, as this pattern of attention allocation would not explain our findings, it would still be indicative of engagement in visual search by identifying a target among distractors.

Finally, if participants had engaged in a confirmatory form of visual search whereby they attended to the target merely to confirm that it corresponds to their WM template: a) a learning pattern as indicated by faster RTs would be expected, which is not present in the data and b) this would require a search for the target as indicated by the colour singleton and comparison to the representation held in WM. If this were the case, targets corresponding to WM template would strengthen (or reinforce) the representation (Souza & Oberauer, 2016) meaning participants would find it easier to identify a matching representation in a match/no-match test. However, I did not find an effect of congruency on percentage of correct responses, suggesting participants did not engage in such a form of visual search. Future research should further explore these issues by varying identity and emotional valences of WM templates and targets.

3.5.2. Conclusion

In summary, the results in this study showed evidence for a WM modulated attentional bias to naturalistic emotional faces. Higher antisocial traits predicted an increased WM bias of visual perception, meaning that participants displaying higher aggression and trait anger were more efficient in identifying emotional faces identical to the induced internal representations. Finally, a relationship was found between the WM modulated bias of visual perception and later stages in social-information processing including emotional response, hostile attribution bias, and response enactment variables. The present research demonstrated the role of emotional WM templates in social-information processing, from encoding social stimuli to behavioural outcomes. Moreover, these findings suggested that WM biases of perception may contribute to the development and maintenance of cognitive biases related to antisocial traits.

4. Chapter 4: Study 3- Cognitive biases associated with aggression and trait anger: the role of emotional state and emotion dysregulation

4.1. Abstract

Negative emotional states have been shown to have a considerable impact on visual processing. Using an anger induction procedure and the combined WM and visual search task developed in the previous chapter, this study examined the effect of state anger on WM biases of visual attention to emotional faces. Further, this study investigated links between proneness to anger, ability to regulate emotion and later stages in social-information processing and aggression. Whilst there was no effect of induced anger condition on accuracy of WM templates, participants in the induced anger condition were faster to respond to targets in the visual search display. The fastest RTs to targets in the visual search display were found in the mood congruent trials, in which participants in the anger induction condition holding an angry face in WM identified angry targets in the visual search display. Structural equation modelling showed that later stages in social-information processing mediated the effects of proneness to anger on aggression and ability to regulate emotional state was positively related to WM biases of attention and later stages in the SIP. The findings suggested that emotion is linked to all stages of the SIP although these links may be mediated by ability to regulate the experience of anger and effects of behavioural outcomes.

4.2. Introduction

The empirical chapters so far have investigated attentional bias to angry and neutral faces using the DPT (study 1) and WM guidance of visual attention to threat in an offender sample (study 2). Attentional bias measured using the DPT was not linked to aggression. However, increased aggression and trait anger predicted increased guidance of visual attention toward angry faces whilst holding angry faces in WM. This chapter extends the previously demonstrated WM guidance of visual attention to naturalistic stimuli to a community sample and explores the role of emotion in the SIP model of aggression. Building on the results from study 2, WM guidance of visual attention to threat was examined following induction of an angry emotional state. This study also examined links between emotion dysregulation and stages of the SIP model leading to aggression.

Anger is an emotion high in arousal and negatively valenced (Ford et al., 2012). Both anger and aggression are typical responses to provocation (Osgood, 2017) and there is a strong positive association between anger and aggression (Smith & Waterman, 2004, Osgood et al., 2021). Although anger can be expressed in a socially acceptable manner, maladaptive anger and aggression have been shown to impair functioning across cognitive and behavioural domains (Smith & Waterman, 2003) and come at a considerable cost to both aggressor and victim (Cohen et al., 2010). Threatening or aversive stimuli or events have been identified as precursors of angry emotional states (Repple et al., 2017). In turn, state anger has been associated in the cognitive domain with heightened attention (hypervigilance) to threatening stimuli (Smith & Waterman, 2004) and increased attribution of hostile intent in ambiguous situations (Quan et al., 2019a).

Processing negative cues does not directly result in aggression; however, cognitive theories propose that distorted processing of social information (i.e., cognitive biases) increase the likelihood of aggressive outcomes (Martinelli et al., 2018). The SIP model of aggression (Crick & Dodge, 1994) has been used to identify cognitive biases immediately preceding aggression. Emotionality, temperament, and the ability to regulate emotion are central to the SIP model (Dodge & Rabiner, 2004; Arsenio, 2010), with emotion believed to permeate all stages of the model (Dodge & Pettit, 2003). Consequently, aggressive individuals are expected to display a hypervigilance to potentially threatening emotional expressions, higher emotional reactivity, and poorer capacity to regulate emotion (Arsenio & Lemerise, 2004).

In a general emotion model, the rise of an emotion is accompanied by behavioural and physiological changes as well as changes in how the environment is appraised and encoded (Moors et al., 2013). Monitoring one's own and others' emotions shapes how a social interaction is perceived, remembered, and reacted to, and plays a considerable role in the encoding and revision of information (Trevors et al., 2017). Emotions and adaptive emotion regulation strategies can be stable, consistent, and have a positive role, facilitating communication (Bridges et al., 2004) and threat detection. Conversely, emotion processing deficits have been associated with less refined emotion regulation strategies and higher rates of externalizing problems and aggressive behaviour (Roberton et al., 2012). Emotion dysregulation has been shown to mediate the relationship between general negative affect and aggression (Donahue et al., 2014) and to correlate with aggression beyond maladaptive anger control Cohn et al., 2010).

The valence of an emotional state changes information processing patterns to influence behavioural output (Souza et al., 2021) and anger has frequently been identified as a precursor of aggression (Novaco & Taylor, 2015). Whilst expressions of anger can be evaluated as assertive (Thompson & Berenbaum, 2011), strong, and competitive (Sell et al., 2014), high levels of trait anger have consistently been associated with aggressive behavioural outcomes (Gilbert et al., 2013). There is an established effect of anger on social-information processing (Smeijers et al., 2020). Anger has also been associated with increased attentional bias toward angry faces (Maoz et al., 2017), reward (Ford et al., 2012), and hostile social cues (Veenstra et al., 2018). Additionally, experimentally induced (or state) anger has been linked to increased attribution of hostile intent (de Castro et al., 2003), choosing self-focused over others-focused goals (Harper et al., 2010), and higher levels of aggression, dominance, and status seeking (Cabral & de Almeida, 2019). The relationship between emotional states and systematic biases in social-information processing appears robust, underscoring the importance of studying associations between anger dysregulation and cognitive processes leading to aggressive behavioural outcomes.

Attentional and WM mechanisms regulate the extent to which anger impacts cognition. Namely, increased state anger has been linked to a narrow attentional scope (Harmon-Jones et al., 2013), decreased attention allocation to non-salient information (Peng et al., 2015), and increased visual attention to rewards (Ford et al., 2010). Moreover, a negative emotional state (or negative affect) has been linked to a decrease in WM capacity and an increase in precision of WM templates, potentially as a consequence of the narrowing in cognitive scope. Long et al. (2020) for example found that WM precision, measured using a precision index (circular standard deviation of a Von Mises distribution, inversely related to the visual WM precision), was significantly lower in the negative compared to neutral emotional state condition. Of the handful of recent studies investigating the link between WM precision and emotional state, one study found no evidence for an effect (Souza et al., 2021) while the general consensus suggested emotional state, albeit modulated by WM capacity and encoding time, enhanced the precision of WM representations (Spachtholz et al., 2014, Long et al., 2020). The narrowing of cognitive focus also impacts how facial expressions are processed and encoded into WM (Maran et al., 2015), disrupting holistic face processing (Curby et al., 2012) and enhancing processing of affect-congruent facial expressions (Quarto et al., 2014).

The ability to regulate emotion effectively draws on skills such as emotional awareness and employment of adaptive strategies to adjust one's emotional state and pursue goal-oriented behaviours (Gratz & Roemer, 2004). Emotion dysregulation is associated with both internalizing and externalizing problems (Neumann et al., 2010) including aggression (Garofalo et al., 2020) in both community and offender samples (Hosie et al., 2021) and has been associated with anger (Bertsch et al., 2020) and psychopathic traits (Garofalo et al., 2018). The relationship between emotion dysregulation and aggression is believed to be mediated by negative affect (i.e., disposition to experience negative emotional states, Donahue et al., 2014), with fluctuations in negative effect linked to increased likelihood of interpersonal conflict (Simons et al., 2016). Although emotion is modelled to permeate every step of the SIP model (Crick & Dodge, 1994), to date, only four studies have investigated emotion regulation in the context of the model (Smeijers et al., 2020).

These studies have identified associations between emotion dysregulation and hostile attribution in young adults (Gagnon et al., 2015) and children (de Castro et al., 2005). Emotion dysregulation was also shown to moderate the relationship between anger and aggression, with adaptive emotion regulation diminishing the relationship between anger and reactive aggression (Calvete & Orue, 2012). Helmsen et al. (2012) found a direct effect of emotion dysregulation on aggression in young children which was not mediated by SIP variables. Finally, emotion dysregulation has not been studied in relation to attention or WM processes within the SIP model although studies linking emotion dysregulation to later stages in the SIP suggested influences should be found at early stages (e.g., encoding) as well (Smeijers et al., 2020). This is consistent with theories of emotion regulation which posit that regulation strategies are impacted by availability of cognitive resources (such as WM capacity, Opitz et al., 2012) and previous research indicating a link between increased emotion dysregulation and increased attentional bias to threat (Bardeen et al., 2017).

Taken together, evidence indicates that negative emotional states (including induced anger) have a considerable effect on WM representations and face processing (Long et al., 2020). Increased anger and the associated narrowed attentional focus have been linked to an increasingly thorough evaluation of social information (Semmler & Hurst, 2017, Young et al., 2011). Mechanisms of WM and attention are intimately linked, playing crucial roles in processing social information (Xu et al., 2021). It has been shown that attention is preferentially allocated to stimuli matching WM templates (Soto et al., 2005). Therefore, it is possible that emotional states (e.g., anger) can lead to a narrowing attentional focus during visual search via enhancing the precision of WM templates. Using an anger induction procedure, the present study examined whether induced anger leads to enhanced precision in encoding of emotional faces as WM templates.

Previous studies have shown that angry faces are preferentially attended to and encoded (Salahub & Emrich, 2020); it was therefore expected that increased encoding accuracy would be observed for angry compared to neutral faces. The effect of state anger on the relationship between WM and visual perception has never been investigated. However, increased state anger has been associated with increased precision of WM representations (Long et al., 2020) and slower RTs to emotional faces (Davies et al., 2020). Hence, participants in the induced anger condition were expected to display slower RTs to targets in the visual search display. Guidance of attention by WM from our previous study (Chapter 3, Satmarean, et al., 2022) was expected to be replicated in the present sample. That is, a main effect of congruency was expected (i.e., faster RTs to targets matching the WM template) and participants holding an angry face in WM were expected to identify target faces among distractors more efficiently (i.e., faster RTs). Finally, structural equation modelling (SEM) was used to examine the role of emotion (i.e., emotion dysregulation and proneness to anger) in the relationships between stages of the SIP and aggression. A positive relationship was expected between emotion dysregulation, trait anger, and cognitive biases in the SIP leading to aggression.

4.3. Method

The method, data preparation protocol, and data analysis plan were pre-registered with the Open Science Framework (OSF, see link for full pre-registration). This study was approved by the University of Sheffield Ethics Committee. Written informed consent was obtained and the data were analysed anonymously.

4.3.1. Participants

Long et al. (2020) found a significantly higher precision of encoding in WM for participants in a negative compared to neutral emotional state (*Cohen's d*= 0.46). In line with these findings, power analysis assuming a medium effect size of Cohen's *d*= 0.46 for the pre-registered 2x2 ANOVA, namely the predicted main effect of emotion induction on WM encoding precision (G*Power, Faul et al., 2007) revealed that a sample size of 138 participants was recommended to achieve power of .85 (α= 0.05, one-tailed). Participants were recruited beyond the recommended sample (N= 220). Following the exclusion of participants with less than 70% correct trials on the visual task, data from 207 participants was used in the final analyses (M_{age}= 36.89, SD_{age}= 13.09; 55.56% female). For the pre-registered 2x2x2 ANOVA, the power levels afforded (α= 0.05) for the expected main effect of congruency (emotional value of the target matching emotional value of the WM template) in the current sample size for a small (f= 0.10), medium (f= 0.25), and large (f= 0.40, Cohen, 1988) effect were .18, .75, and .99 respectively. Participants were recruited via Prolific (www.prolific.co) on the basis of being English speakers and based in the United States or United Kingdom. The study was not made available to participants who had taken part in a previous study using similar tasks (Chapter 3, Satmarean et al., 2022).

4.3.2. Measures

Social-information processing patterns were measured using an abbreviated version of the Social-Emotional Information Processing Questionnaire (SEIP-Q, Coccaro et al., 2017, see the Measures sub-section in studies 1 and 2 for a full description). In this sample, Cronbach's alpha was α= .87 for hostile attribution, .78 for response enactment for behaviourally aggressive responses, and .80 for response enactment for relationally aggressive responses.

Emotion dysregulation was measured using the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004). The DERS is an established measure of emotion dysregulation, comprising 36 items rated on a 5-point Likert scale ranging from 1 (Almost never) to 5 (Almost always). Sub-scales address: 1) Lack of emotional awareness, 2) Lack of emotional clarity, 3) Difficulties controlling impulsive behaviours when distressed, 4) Difficulties engaging in goal-directed behaviours when distressed, 5) Nonacceptance of negative emotional responses, and 6) Limited access to effective emotion regulation strategies. Greater overall DERS scores suggest greater difficulties regulating emotion. The DERS demonstrated good psychometric properties (Garofalo et al., 2018) in both clinical and nonclinical populations (Gratz & Roemer, 2004). DERS factors correlate with externalizing and internalizing problems (Neumann et al., 2010). In this sample, Cronbach's alpha was $\alpha = .95$ for the overall emotion dysregulation score and varied between .81 and .93 for the subscales.

Trait Anger was measured using the 10-item Trait Anger subscale of the State-Trait Anger Expression Inventory–II (Spielberger, 1999, see the Measures sub-section in study 2 for a full description). In this sample, Cronbach's alpha was $\alpha = .81$ for the overall anger score, .63 for the reactive, and .86 for the temperament subscales.

Reactive and Proactive aggression were measured using the Reactive-Proactive Aggression Questionnaire (RPQ) (Raine et al., 2006, see the Measures subsection in study 1, Chapter 2 for a full description) comprising 23 items measuring proactive (12 items) and reactive aggression (11 items). In the current sample, Cronbach's alpha was $\alpha = .86$ for the overall aggression score.

Emotional state measurement: To assess the ***effectiveness of the mood induction procedure***, participants were asked to rate their current feelings of anger, frustration, irritation, and annoyance on Likert scale items ranging from 0 (Not at all) to 8 (Extremely). An average was produced for each participant (Ford et al., 2010, α = .87). Given that research has shown reappraisal (i.e., ratingemotional state immediately after emotion induction) to be an effective means of reducing anger (Keltner et al., 1993, Szasz et al., 2011), in this study anger was measured prior to the emotion induction procedure (to measure baseline emotional state) and following completion of the visual task, in line with Spachtholz et al. (2014). In this sample, Cronbach's alpha was α= .95 for emotional scores both before and after induction.

4.3.3 Emotion Induction

Emotion was induced using autobiographical recall. Participants were asked to recall a time when they felt little emotion (neutral condition; Ford et al., 2010) or a time when their (ex)partner or best friend acted unfairly toward them, resulting in an angry emotion (anger condition). They were instructed to provide a detailed description of this person's behaviour, how the situation came about, how it made them feel, and what the outcome was. Participants were required to spend at least 4 minutes writing about the event to ensure optimal reliving (Veenstra et al., 2017). Autobiographical recall has been used extensively in emotion induction research (Szasz et al., 2011) and is effective in inducing anger (see Joseph et al., 2020 for a review).

4.3.4. Dual task

The Gorilla Experiment Builder (www.gorilla.sc) was used to create and host the experiment and collect online reaction time data. Following Burra et al. (2017) attentional bias to threat was inferred from reaction times (RTs) on a dual task, consisting of a working memory task and a visual search task. The development and validation of the dual task are detailed in Satmarean et al. (2022) and the Methods subsection in study 2.

4.3.5. Analysis Plan

A cross-sectional study design was used to investigate the effect of induced emotional states on WM representations and visual search, as well as associations between WM guidance of visual search, social information-processing, and self-report measures of aggression, trait anger, and emotion dysregulation. A 2 x 2 x 2 mixed model design was used with within-participants factors of: 1) emotional valence of the WM representation (angry or neutral), 2) emotional valence of the memory template (angry or neutral) and a between-participants factor of induced emotional state (random allocation to angry versus neutral conditions) as IVs and RTs as DV. Structural equation modelling was used to examine the relationships between bias scores and self-reported measures (IVs) as predictors of aggression (DV).

I. Is induced anger associated with enhanced precision in encoding of faces?

A 2x2 mixed ANOVA with WM encoding accuracy as the dependent variable and induced emotion (between-subject factor) and emotional valence of the WM template (within-subject factor) as independent variables was expected to reveal a main effect of emotional valence of the WM template. This would indicate preferential allocation of attention to and encoding of negative emotional information. WM encoding accuracy was measured by percentage of correct responses on the match/no-match test and emotional valences of the WM template and of the target consisted of angry and neutral faces. A main effect of emotion induction was also expected with participants in the induced anger condition expected to display higher accuracy on the match/no-match test compared to participants in the neutral condition. An interaction effect between induced anger and the emotional valence of the memory representation was not expected as previous studies have demonstrated a broad influence of induced anger over WM performance for emotional facial expressions (Maran et al., 2015).

II. Does induced anger have an effect on visual search for faces? Is this effect modulated by a previous effect of induced anger on WM templates?

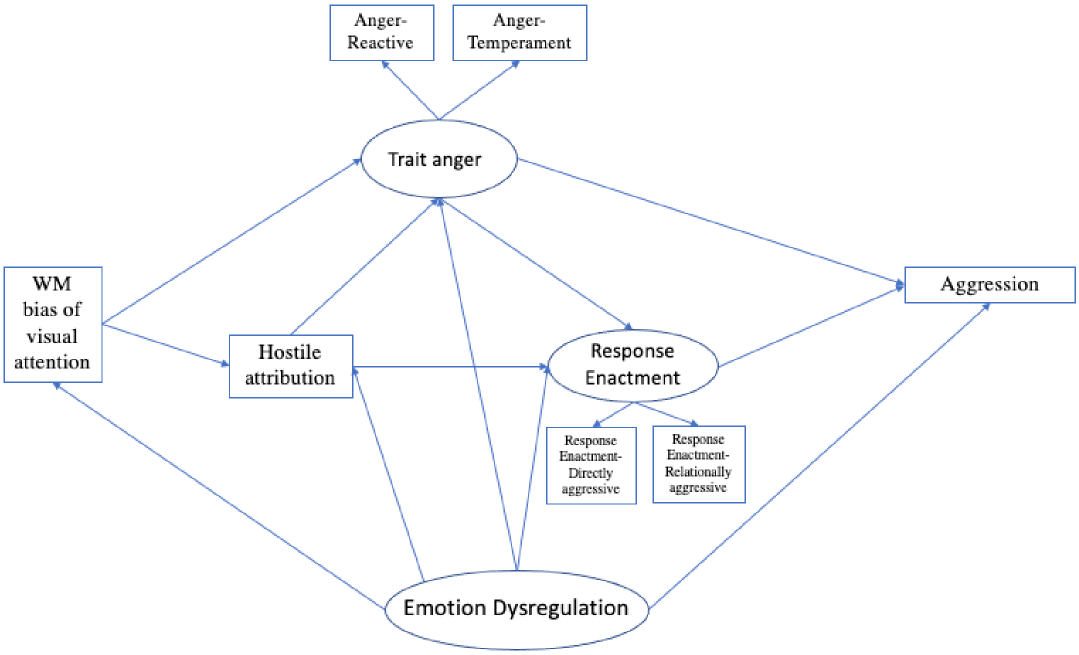
A 2x2x2 ANOVA with mean RTs (latency) to targets in visual search displays as the dependent variable and emotional valence of the WM template (angry vs neutral, within-subject factor), emotional valence of the visual search target (angry vs neutral; within-subject factor), and induced emotional state (angry versus neutral; between-subject factor) as independent variables was expected to reveal a main effect of induced emotion. That is, participants were expected to display slower RTs to targets in a visual search task following anger induction. Whilst a hypothesis was not pre-registered for a pattern of interactions, an interaction effect was expected between induced emotional state and visual search, with slower RTs to angry faces following anger induction compared to neutral faces in a neutral affective state (see Quarto et al., 2014). However, this effect may be modulated (i.e., adjusted, kept in proportion) by WM templates.

III. Does emotion regulation mediate the relationship between hostile attribution, anger, and aggression?

High correlations were expected between constructs related to aggressive traits and behaviours. Structural equation modelling (SEM) based on maximum likelihood estimation was used to examine the effect of emotion dysregulation and trait anger on aggression and steps of the SIP model (encoding, attribution of intent, response enactment). The hypothesized model of the relationships between these constructs is shown in Figure 8. Formal research hypotheses for the links in this model are detailed below.

Figure 8

Pre-registered structural equation model of the moderating role of emotion dysregulation in the relationship between WM guidance of attention, hostile attribution, anger, and aggression



WM bias of visual attention -> aggression

Within the SIP model, individuals with a history of aggression are expected to disproportionately attend to and encode stimuli perceived as threatening. Preferential encoding of threatening information is the first step in SIP and believed to precede hostile attribution of intent (Dodge, 2006) during social interaction. While some studies have found a direct link between attentional biases toward threat and aggression, others have not found a significant relationship (see Manning 2020 for a review). Additionally, Maoz et al. (2017) found a positive relationship between anger and attentional bias to threat, separate from measurements of attribution biases, suggesting distinct mechanisms for attention and attribution biases. Taken together, these findings indicate that the association between encoding and aggressive behavioural outcomes is likely to be indirect, mediated by subsequent steps in the SIP and trait anger. Therefore, *WM biases of visual attention were expected to be positively related to hostile attribution with indirect effects of WM biases of visual attention on aggression through anger and subsequent SIP steps.*

Anger, later stages in the SIP model -> aggression

Interpretation of social cues and intent attribution constitute the second SIP stage. When presented with an ambiguous social situation, individuals with a history of aggression are expected to display a tendency to interpret ambiguous social cues to hostility. Within the model, increased hostile attribution leads to instrumental or aggressive goals (Fontaine et al., 2010) which in turn precede aggression. There is evidence of a direct relationship between anger and aggression (Gilbert et al., 2013). Moreover, anger (Quan et al., 2019a, Wilkowski & Robinson, 2010) and later stages in social-information processing (Fontaine et al., 2010) have been found to mediate the effect of hostile attribution on aggression. Anger has also been shown to have an indirect effect on aggression through response evaluation and decision (including response enactment) stages of the SIP (de Castro et al., 2005). Consequently, *response enactment was expected to mediate the association between anger and aggression and response enactment and anger to mediate the association between hostile attribution and aggression.*

Emotion dysregulation -> aggression

Emotion dysregulation has been shown to be positively related to anger and to have both direct and indirect effects on aggression, through interpretation and response evaluation and decision steps of the SIP model (de Castro et al., 2005). Emotion regulation strategies act to control the effect of emotion at several steps in SIP, reducing the probability of aggression (Smeijers, 2020). Increased emotion dysregulation has been associated with aggression

(Calvete & Orue, 2012) and increased measures of psychopathy (Garofalo et al., 2018). Thus, *emotion dysregulation was expected to be positively related to aggression and for anger and later SIP stages to mediate the association between emotion dysregulation and aggression.*

4.3.6. Stimuli

Task stimuli for both WM and visual search tasks comprised angry and neutral faces (6 male, 6 female) selected from the NimStim stimulus set (Tottenham et al., 2009). A full description of stimuli preparation can be found in the Methods subsection in study 2.

4.3.7. Procedure

Participants completed the self-reported measures above (in the order listed above). After measuring current emotional state, participants were randomly allocated to one of two emotion induction conditions (angry and neutral) and completed the autobiographical recall task. This was followed by the dual visual task. A full description of the visual task can be found in the Methods subsection in study 2. The visual task was followed by a second measurement of emotional state and the debrief. This study was approved by the Department of Psychology, University of Sheffield research ethics committee.

4.3.8. Data Preparation

Analyses were performed using RStudio 4.0.2 (RStudio, 2020). As planned, outliers were Winsorized; values outside 1.5 interquartile ranges from the Tukey Hinges (lower and upper hinge corresponding to the first and upper quartiles or 25th and 75th percentiles respectively) were rescaled to the last valid value within the range. There were no missing trials in the combined visual task. Following Burra and Kerzel (2019), participants with <70% correct trials on either visual search or match/no-match test were excluded from analysis

(N=12). Incorrect responses on both visual search and match/no-match were removed from the dataset. Following Burra et al. (2017), RTs under 200 ms were deemed unlikely to reflect genuine responses and were therefore removed.

None of the participants were missing >20% of questions within a self-report measure. Histograms plots, box plots, and z scores for skewness and kurtosis indicated that the proactive aggression scores were significantly positively skewed. These scores were converted to a binary variable which distinguished 0 from scores above 0 (coded "1"). Response enactment scores for directly aggressive responses were positively skewed which was removed by square root transformation. Response enactment for relationally and directly aggressive behavioural responses were significantly correlated (r=.63). A composite variable was created by averaging response enactment scores. Alpha coefficients for the composite variables were excellent (α= .87 for the response enactment for aggressive responses composite). Accuracy of encoding scores (i.e., percentage of accurate match/no-match trials per participant) was negatively skewed which was removed by a cube transformation.

Examination of skewness and kurtosis for variables included in the SEM can be seen in Table 10. Age was positively skewed; following a square root transformation the skewness coefficient improved (0.48) however the kurtosis coefficient was lower than 2.5 (2.48). Transformations have been shown to only have a minor effect on moderate or slight univariate non-normality and transformed variables can lead to degradation in hypothesized SEM models, impacting model fit indices (Gao & Mokhtarian, 2008). Moreover, given that SEM model fit indices depend on kurtosis but not on skewness (Yuan et al., 2005), non-transformed age scores were included in SEM modelling.

Table 10

Skewness and kurtosis estimates for the variables included in the pre-registered structural equation model

Variable	Pearson's kurtosis	Skewness
Anger	2.89	0.39
Reactive Proactive Aggression	2.91	0.51
Hostile Attribution	2.97	-0.07
Response Enactment	2.63	0.34
Emotion Dysregulation	2.55	0.34
Age	3.11	0.80
WM bias of attention toward target	2.94	0.34

4.3.9. Computation and reliability of attentional bias data

As pre-registered, trial-level BS (Zvielli et al., 2015) were computed using the weighted trial method. A time-series of BS per participant was produced by subtracting the RTs for trials in which participants were asked to hold an angry face in WM from the weighted mean of all trials in which participants were asked to hold a neutral face in WM (see section 2.2.5. for more detail on trial-level BS computations). This was divided into $BS_{POSITIVE}$ (mean of *positive* BS, indicating attentional bias toward target stimuli) and $BS_{NEGATIVE}$ (mean of *negative* BS, indicating attentional bias away from target stimuli) scores. Split-half reliabilities for BS were high (Spearman Brown adjusted $r= 0.91$).

4.4. Results

4.4.1. Manipulation check

Emotional state was measured prior to the anger induction procedure and after completion of the visual task. Means and standard deviations for emotional states before and after angry and neutral emotion induction can be found in Table 11. Emotional state variables were positively skewed; therefore, a non- parametric test was used to examine emotional state in the anger induction condition. A paired Wilcoxon signed-rank test showed that the autobiographical recall of an anger inducing event resulted in a significantly higher median for state anger ($p < 0.001$, large effect size $r = 0.54$). The median emotional state score after neutral emotion induction was also significantly higher than the median score prior to emotion induction ($p < 0.001$, large effect size $r = 0.65$). The median score for emotional state after anger induction was significantly higher than the median score for emotional state after neutral emotion induction ($p = 0.02$, small effect size $r = 0.16$). Participants in the induced neutral emotional state displayed lower state anger than participants in the induced anger condition, both prior and after emotion induction.

The unexpected difference in baseline anger scores between the two groups could be explained by a randomisation failure, where individuals with higher state anger scores were grouped in one condition. Exploratory analyses were conducted to better understand the differences in baseline, increase in anger across conditions, and whether participants in the induced anger group showed increased proneness to anger; however, there was no evidence of difference in trait anger between emotion induction groups (see Appendix D for the full exploratory analyses). Given that state anger increased for both groups, the responses provided by the participants in the autobiographical recall task were examined. A proportion of the participants in the induced neutral emotion group (N=20) recalled an event where they mentioned experiencing feelings of anger or participating in a conflict or argument with another person. Conversely, a small number of participants in the induced anger condition stated they could not recall an anger inducing event (N=3). Exploratory analyses were conducted whereby these individuals were moved to the emotion induction condition corresponding to their response (so induced anger for participants recalling an anger inducing event and induced neutral for participants who could not recall an anger inducing event). Findings were highly similar to the analyses using the original randomised groups (below) and are detailed in Appendix D.

The pre-registered analyses are detailed below; however, there is some uncertainty in interpreting the difference in state anger between groups. It is possible that emotional state self-reported measures do not accurately capturing the potential dynamics of state anger over time or the dimensions of emotional state induced using the autobiographical recall task. Given that the results were highly similar once participants were separated based on recall of situations in which they experienced anger in a social context, it is also possible that remembering a negatively charged social event resulted in activation of aggressive schemas as well as an increased emotional response (a discussion can be found in Appendix D). As no measurement of emotional state was taken following the anger induction procedure, it is not possible to ascertain the effectiveness of the procedure and therefore whether the experimental manipulation was successful. The results below should be interpreted with these limitations in mind.

Table 11

Means and standard deviations for state anger across emotion induction conditions, before and after emotion induction

Emotion Induction	State Anger-Before (Mean/SD)	State Anger- After (Mean/SD)
Neutral	0.85/ 1.4	1.9/ 2.03
Angry	1.48/ 1.79	2.48/ 2.08

Note. SD= standard deviation

4.4.2. Results relating to pre-registered hypotheses

I. Do participants in the induced anger condition display enhanced precision in encoding of faces?

Participants were expected to display increased accuracy in encoding angry compared to neutral faces in WM and participants in the induced anger condition were expected to display a higher percentage of correct responses to the match/no-match test. A two-way mixed ANOVA was conducted with percentage of correct trials as the dependent variable and emotional valence of the WM template (angry vs neutral, within-participants) and induced emotional state (angry vs neutral, between-participants) as independent variables. As expected, there was a main effect of the emotional valence of the WM template ($F(1, 218)= 102.13$, $p< 0.001$, $\eta_p^2= 0.32$) with higher accuracy for angry ($M= 85.2$, $SD= 11.61$) compared to neutral ($M= 80.61$, $SD= 13.47$) faces encoded as WM templates. Contrary to the hypothesis, there was no evidence of a main effect of induced emotional state and no interaction was found between induced emotional state and emotional valence of the WM template.

II. Does induced anger have an effect on visual search for faces? Is this effect modulated by a previous effect of induced anger on WM templates?

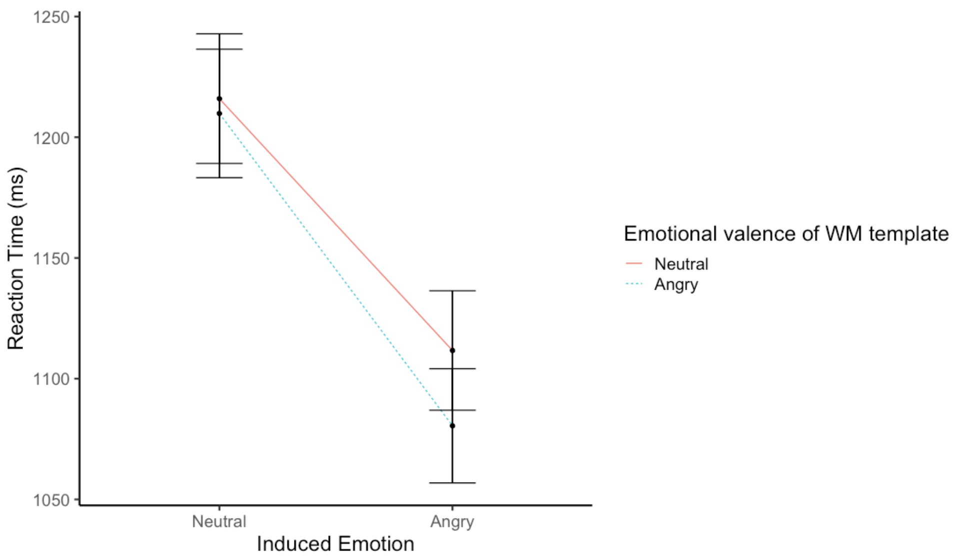
A 3-way mixed ANOVA was performed to evaluate the effects of induced emotional state, emotional valence of the WM template, and emotional valence of the target in the visual search display on RTs to targets in the visual search array. As hypothesized, a main effect of induced emotion was found ($F(1, 218)= 5.46$, $p< 0.05$, $\eta_p^2= 0.02$). However, this was not significant as the effect was found in the opposite direction to the pre-registered hypothesis. Slower RTs were expected in the induced anger compared to neutral condition. However, faster RTs were found in the angry ($M= 1096.08$, $SD= 360.47$) compared to neutral ($M= 1212.92$, $SD= 394.13$) induced emotion condition.

As expected, a main effect of the emotional valence of the WM template was found ($F(1, 218)= 12.45$, $p< 0.001$, $\eta_p^2= 0.05$) with faster RTs to targets identified whilst holding an angry ($M= 1144.58$, $SD= 378.16$) compared to neutral ($M= 1163.36$, $SD= 385.63$) template in WM. The main effect of the emotional valence of the target in the visual search display ($F(1,218)= 0.004$, $p= 0.95$) and the interaction between the former and induced emotion ($F(1,218)= 0.01$, $p= 0.91$) were non-significant.

A two-way interaction was found between emotional valence of the WM template and induced emotion condition ($F(1, 218)= 5.54$, $p< 0.05$, $\eta_p^2= 0.02$, see Figure 9). The Bonferroni adjusted p-value (significance being accepted at $p<0.025$), the simple main effect of emotional valence of the WM template was significant in the anger ($p< 0.001$) but not in the neutral ($p= 0.56$) emotion induction condition. In the induced anger condition, RTs were faster for targets identified whilst holding an angry ($M= 1080.48$, $SD= 352.42$) compared to a neutral ($M= 1111.68$, $SD= 368.48$) face in WM.

Figure 9

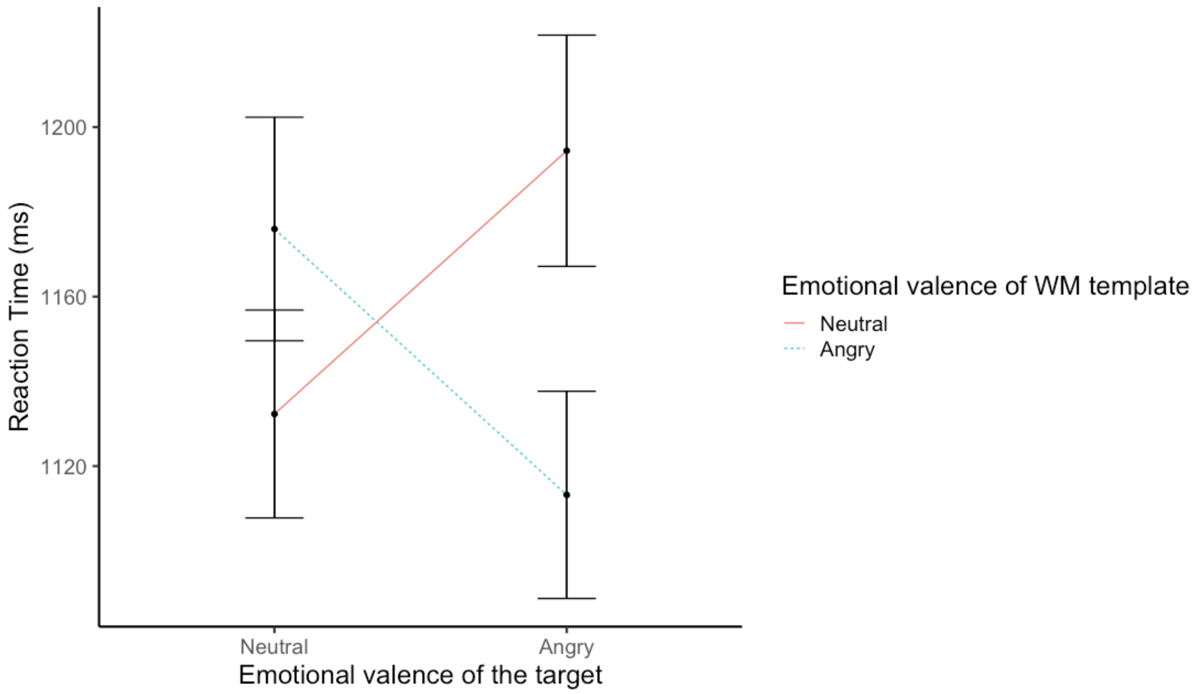
Interaction between emotional valences of WM template and induced emotion. Error bars show one standard error above and below mean



As expected, a main effect of WM guidance of visual attention was found; that is, a two-way interaction was found between the WM template emotional valence and that of the target in the visual search display ($F(1, 218)= 104.76$, $p< 0.001$, $\eta_p^2= 0.32$, see Figure 10). Considering the Bonferroni adjusted p-value, the simple main effect of emotional valence of the target was significant for both angry ($p< 0.001$) and neutral ($p< 0.001$) WM template conditions.

Figure 10

Interaction between emotional valences of WM template and target. Error bars show one standard error above and below mean



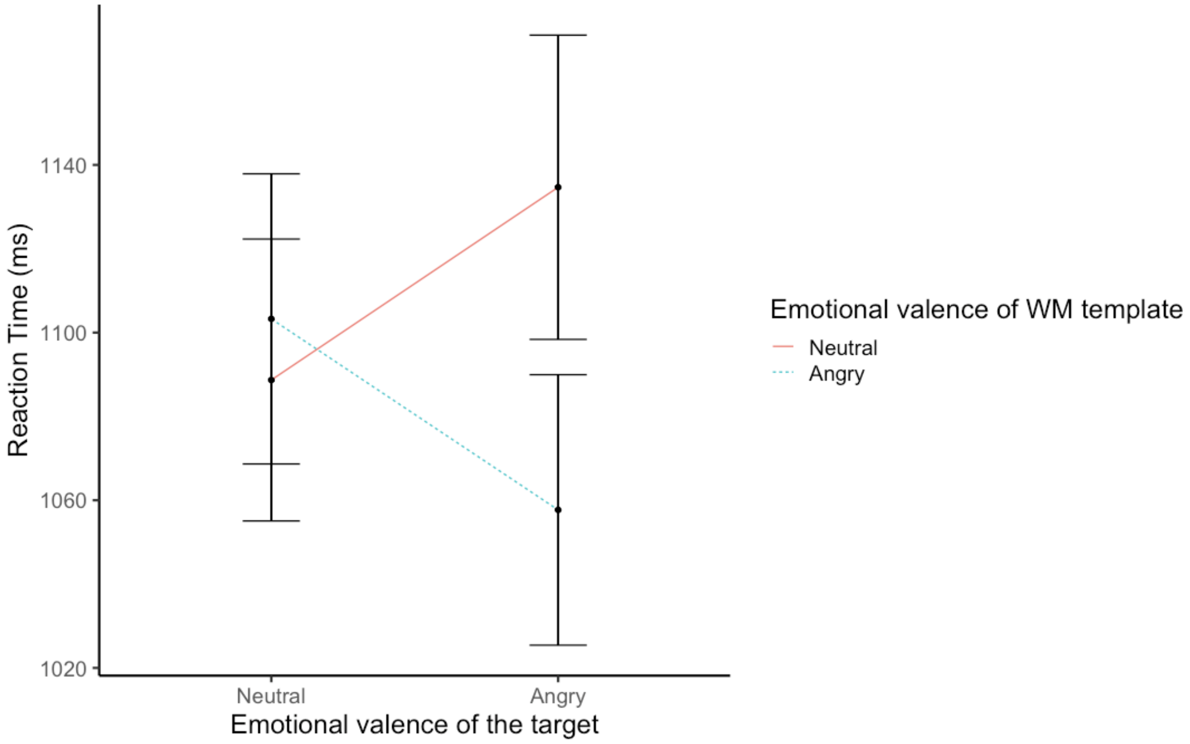
There was a statistically significant three-way interaction between induced emotion condition, emotional valence of the target, and emotional valence of the WM template ($F(1, 218)= 7.60$, $p< 0.01$, $\eta_p^2= 0.03$). For the simple two-way interactions and simple main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the $p < 0.025$ level. There were two statistically significant two-way interactions between emotional valence of the WM template and that of the target for both angry ($F(1, 110)= 38.8$, $p< 0.001$) and neutral ($F(1,108)= 66$, $p< 0.001$) emotion induction conditions.

Simple main effects were conducted next. Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of 0.0125. The simple main effects of the emotional valence of the target on RTs to the visual search display were statistically significant in the induced anger condition for both neutral: $F(1, 110)= 18.1$, $p< 0.0001$ and angry: $F(1,110)= 25.3$, $p< 0.0001$ WM templates. Similarly, simple main effects of the emotional valence of the target on RTs to the visual search display were statistically significant in the induced neutral condition for both neutral: $F(1, 108)= 43.9$, $p< 0.0001$ and angry: $F(1,108)= 44$, $p< 0.0001$ WM templates.

Pairwise comparisons were conducted for each emotion induction condition. First, pairwise comparisons were run in the **induced anger condition** between RTs to the two types of emotional valence for targets in the visual search display across emotional valences for WM templates (see Figure 11). A Bonferroni adjustment was applied. When preceded by a neutral WM template, mean RTs were significantly higher ($p< 0.0001$) for angry targets ($M= 1134.67$, $SD= 382.39$) compared to neutral targets ($M= 1088.69$, $SD= 354.27$). When preceded by an angry WM template, mean RTs were significantly lower ($p< 0.0001$) to angry targets ($M= 1057.69$, $SD= 339.9$) compared to neutral targets ($M= 1103.26$, $SD= 364.62$).

Figure 11

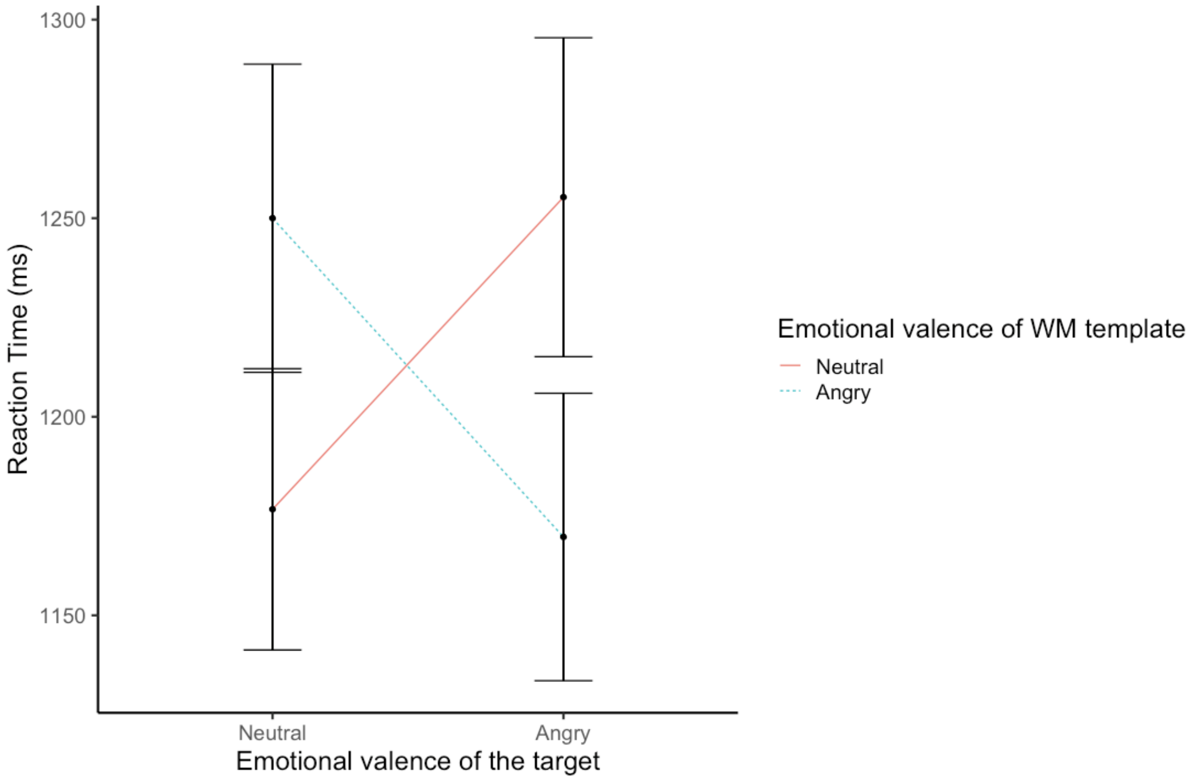
Interaction between emotional valences of WM and target in the induced anger condition. Error bars show one standard error above and below mean



Next, pairwise comparisons were run in the **induced neutral condition** between RTs to the two types of emotional valence for targets in the visual search display across emotional valences for WM templates (see Figure 12). A Bonferroni adjustment was applied. When preceded by a neutral WM template, mean RTs were significantly higher ($p < 0.0001$) to angry targets (M= 1255.29, SD= 419.21) compared to neutral targets (M= 1176.7, SD= 369.71). When preceded by an angry WM template, mean RTs were significantly lower ($p < 0.0001$) to angry targets (M= 1169.72, SD= 377.81) compared to neutral targets (M= 1249.99, SD= 405.11). The main effect of the emotional valence of the target in the visual search display and the interaction between the former and induced emotion were non-significant.

Figure 12

Interaction between emotional valences of WM and target in the induced neutral condition. Error bars show one standard error above and below mean



A paired t-test was conducted to demonstrate the expected difference in RTs to congruent compared to incongruent trials. Faster RTs were expected to congruent (emotional valence of the WM matching that of the target) compared to incongruent (emotional valence of the WM different to that of the target) trials. As expected, the t-test was significant (p<0.001, Cohen's d= 0.49, one-tailed) with faster RTs to congruent (M= 1122.74, SD= 363.01) compared to incongruent (M= 1185.19, SD= 397.7) trials.

H3: Does emotion regulation moderate the relationship between hostile attribution, anger, and aggression?

The expected relationships between SIP model data, emotion dysregulation, and trait anger were tested with SEM. The model was fitted across emotion induction groups. Correlations between variables are summarised in Table 12. The pre-registered model was a poor fit of the data: χ^2 (80, N= 207) = 202.86, $p < 0.001$, Comparative Fit Index (CFI) = 0.87, TLI = Tucker–Lewis Index (TLI)= 0.84, Root Mean Square Error of Approximation (RMSEA) = 0.08, and Standardized Root Mean Square Residual (SRMR) = 0.08) so an alternative path model was specified to test the same expected relationships. That is, observed variables were used for trait anger, response enactment for aggressive responses, and emotion dysregulation instead of latent variables having subscales as indicators (for example, anger-temperament and anger-reactive were previously used as indicators for anger as a latent variable). The path model was a better fit of the data, meeting the accepted criteria (χ^2 (8, N= 207) = 20.20, p= 0.01, CFI= 0.96, TLI= 0.87, RMSEA= 0.08, and SRMR= 0.04). Age and gender were not significant predictors of aggression and, for parsimony, were removed from the final model. Fixing the age and gender parameters to 0 resulted in the final model being just-identified (df= 0), meaning that chi-square, CFI, and TLI indices could not be used to assess the model fit. Significant parameter estimates, coefficients and confidence intervals for the final model are highlighted in Figure 13.

Table 12

Correlations between variables included in the pre-registered structural equation model

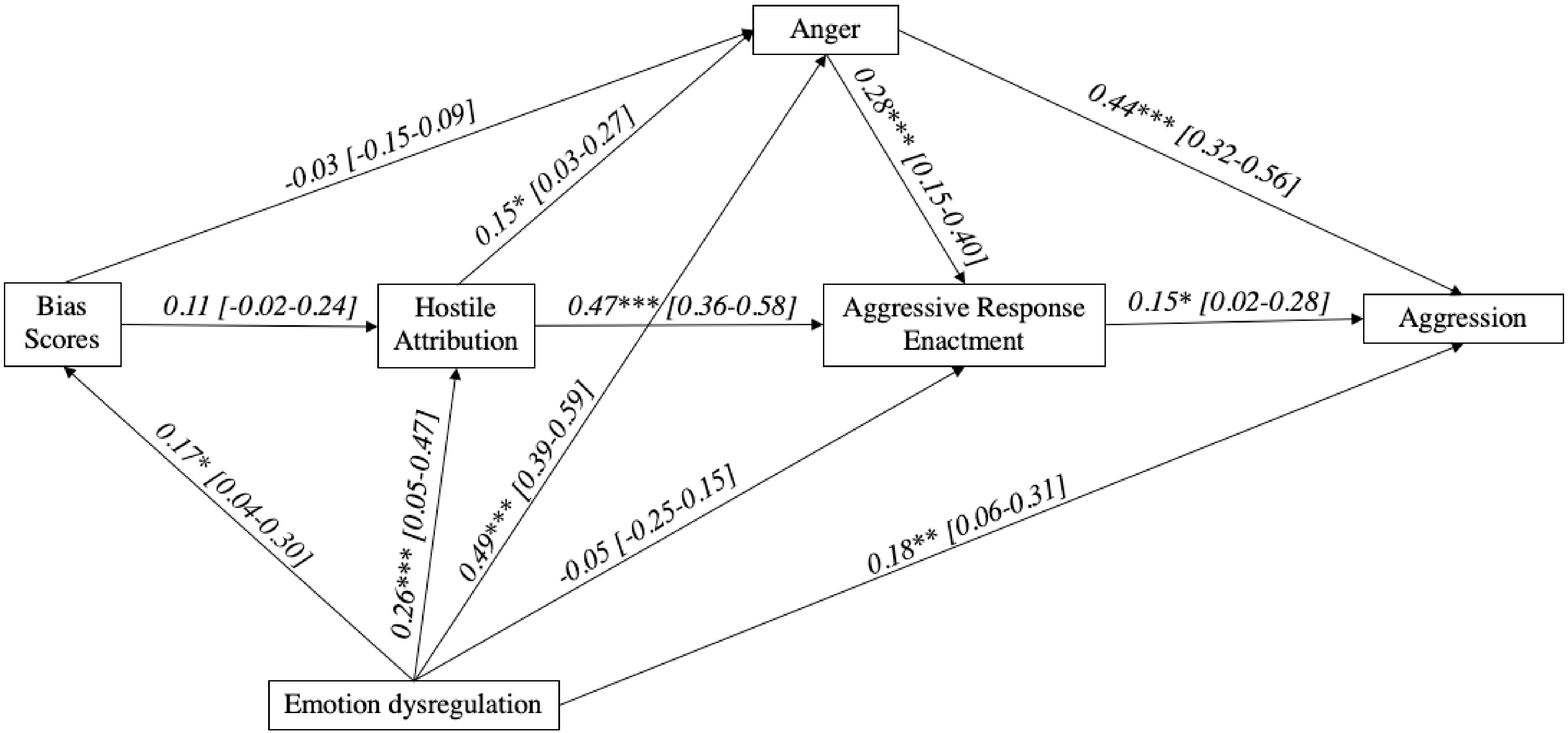
		1	2	3	4	5	6
1.	$BS_{POSITIVE}$	-					
2.	Anger	.07	-				
3.	Aggression	.00	.59***	-			
4.	Emotion Dysregulation	.17*	.53***	.45***	-		
5.	Hostile Attribution	.15*	.28***	.23***	.28***	-	
6.	Response Enactment	.15*	.38***	.34***	.22**	.54***	-

Note. BS= Bias Scores.

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

Figure 13

Path model of the moderating role of emotion dysregulation in the relationship between WM guidance of attention, hostile attribution, anger, and aggression, and standardized parameter estimates



* p< 0.05, ** p< 0.01, ***p< 0.001

WM bias of visual attention -> Aggression

Contrary to the pre-registered hypotheses, BS did not predict hostile attribution or anger. The total effect of BS on aggression through later stages of the SIP (hostile attribution, response enactment) and through trait anger was non-significant (β= -0.01, p= 0.82).

Hostile attribution and Response Enactment -> Aggression

In line with our hypotheses, response enactment significantly predicted aggression and both response enactment and anger mediated the relationship between hostile attribution and aggression. The results showed significant indirect effects of hostile attribution on aggression through anger (β= 0.06, p< 0.05) and response enactment (β= 0.07, p< 0.05). The total effect of HA on aggression through response enactment, through anger, and through anger and response enactment, was also significant (β= 0.14, p= 0.001).

Anger -> Aggression

Anger was positively related to aggression and there was an indirect effect of anger on aggression, mediated by response enactment (β= 0.04, p= 0.11).

Emotion dysregulation –> Aggression

Emotion dysregulation predicted aggression and increased bias toward target faces whilst holding angry faces in WM. However, the total effect of emotion dysregulation on aggression through hostile attribution and response enactment was non-significant (β= 0.02, p = 0.06).

4.5. Discussion

This study examined the effect of state anger on WM biases of visual attention to emotional faces and examine the role of emotion dysregulation and proneness to anger in the relationships between stages in social-information processing and aggression. As measurements of emotional state were taken prior to the anger induction procedure and following completion of the visual task and given the differences in emotional state at baseline, it was not possible to ascertain whether the experimental manipulation of emotional state was successful. There was no effect of emotion induction condition on accuracy of WM templates. However, participants in the induced anger condition displayed faster RTs to all targets in the visual search display and faster RTs to angry WM templates and targets in the visual search display. Later stages in the SIP model mediated the effects of trait anger on aggression. Emotion dysregulation was positively related to WM biases of visual perception but neither emotion dysregulation nor WM biases of visual perception predicted aggression through later stages in the SIP model. The study was the first to investigate the role of the core structures of the SIP model (WM, emotion) on stages of the SIP and aggression.

I. Is induced anger associated with enhanced precision in encoding of faces?

Participants showed higher precision (accuracy) when encoding angry compared to neutral faces. This was in line with previous research demonstrating preferential allocation of attention toward angry faces (Dickins & Lipp, 2014) and preferential encoding of angry faces in WM (Salahub & Emrich, 2020). Contrary to our hypothesis, participants in the induced anger condition did not display increased precision in facial encoding. Whilst previous research has demonstrated a link between induced negative affect and increased precision during encoding of low-level features (i.e., coloured dots, Spachtholz et al., 2014; Long et al., 2020) and objects (Xie & Zhang, 2017), this effect does not appear to extend to encoding faces.

The lack of an effect of induced anger condition on precision of encoding could be due to increased complexity of encoding faces compared to low-level features. It also may be due to differences in methodology- previous research has looked at the effect of negative affect which encompasses a range of negative emotional states whereas this study focused specifically on anger. It is worth noting that although participants in the induced anger group displayed significantly higher anger compared to participants in the neutral group, anger increased for both groups following completion of the autobiographical recall task. Beyond the effect of emotional state, it is possible that recall of an anger-inducing event activates anger or aggression related schemas which do not have a specific effect on precision of encoding.

A secondary explanation for the lack of a main effect of induced anger condition on precision of encoding could be the cognitively demanding nature of the dual task. The extent to which emotional states can influence WM processes depends on whether the emotional state is actively maintained and focused on- otherwise known as affective working memory (Mikels & Reuter-Lorenz, 2019). This suggests that individuals in different emotional states can encode faces with a similar precision by focusing on a demanding, top-down driven visual search which weakens their focus on their emotional state. This would be in line with a study by Gooding and Tallent (2003) who found that WM performances for facial identities and emotional expressions were highly correlated which suggested a reliance upon the same memory system. It also supports present findings for an overall increased precision of encoding for angry faces which suggests a cognitive bias toward encoding of angry faces (Gong & Smart, 2021).

II. Does induced anger have an effect on visual search for faces? Is this effect modulated by a previous effect of induced anger on WM templates?

All participants identified targets in the visual search display faster when holding an angry compared to a neutral face in WM demonstrating increased efficiency of visual search when holding an emotional (here angry) facial expression in WM. Additionally, a congruency effect was found, whereby participants responded faster to targets matching the representation held in WM. These effects have previously been identified in a sample of individuals who had served a custodial sentence (Chapter 3, Satmarean et al., 2022). Present findings therefore replicate guidance of visual attention by WM in a non-offender sample. Although participants in the induced anger condition were expected to be slower to identify targets in the visual search display compared to participants in the neutral emotion induction condition, the opposite was found. Participants in the induced anger condition identified targets in the visual search display faster than participants in the neutral emotion induction condition suggesting anincreased attentional focus toward task goals and away from distractors (van Steenbergen et al., 2011). Further, participants in the induced anger condition responded faster to targets when holding an emotional (here angry compared to neutral) face in WM. This suggested that participants were significantly more focused on task goals when the WM template was congruent with their emotional state.

Participants in the anger induction condition showed a better overall performance on the dual task. Further, they displayed similar RTs to neutral targets irrespective of the valence of the WM template. However, when identifying angry targets, holding an emotional (angry) face in WM augmented their performance compared to holding a neutral face in WM which resulted in a significant delay. Conversely, participants in the neutral emotional state condition showed distinct patterns of attention allocation to neutral targets, depending on the valence of the WM template. That is, neutral targets were identified faster than angry targets when holding a neutral (congruent) representation in WM. Patterns of attention allocation to angry targets were similar to those of participants in the anger induction condition. Taken together, these findings demonstrated that WM guidance of visual attention for naturalistic faces replicated across sample types. In addition, the findings suggested that emotional states may have an impact on the WM bias of visual attention. A link was found between emotional state and WM contents, that is, participants in the induced anger condition performed better on the dual task when the WM template matched their emotional state. This study also showed that WM modulation of visual attention had a stronger effect for angry compared to neutral targets in the induced anger condition. The congruency effect, whereby attention is drawn to targets matching WM contents was therefore reduced for neutral targets in the induced anger condition.

Emotional states have previously been linked to visual perception biases; however, this relationship is not clear, possibly due to the wide range of methods being used to induce affective states and study how they affect perception. Their effectiveness and reliability have been shown to vary across types of emotion induced, method, and population (see Joseph et al., 2020 for a review). The autobiographical recall method used in this study has been deemed most effective when inducing anger (Joseph et al., 2020); however, other methods such as Velten (reading elated or depressed self-referential statements, Velten, 1968), film, pictures, and music have been used extensively to induce negative affect and effects on visual processing have varied accordingly.

There is an element of ambiguity in emotion induction and effects on visual information processing. Broad concepts such as "positive/negative" emotional states encompass a wide range of emotions such as happiness, or relaxation for the positive emotional states and anger, sadness, anxiety, annoyance, for negative emotional states. Whilst there is some overlap between emotional states, the spectrum of "negative emotional states" is too broad. An increase in anger for example has been associated with increased fear and frustration but no change in emotions such as embarrassment, annoyance, loneliness, or impatience (Lobbestael et al., 2008). Indeed, states of sadness and anxiety typically have been shown to have distinctive effects on visual processing compared to induced anger (Shields et al., 2016). An anger induction procedure was used in this study, which may explain why the present findings partially diverge from patterns of processing associated with "negative emotional states". However, it is worth highlighting that this anger induction procedure may not have been successful in inducing the expected emotional states. This means that the present findings may diverge from previous studies due to variables which were not measured in this study (e.g., activation of negative/unpleasant schemas).

In line with the findings showing an increased focus on task goals in the induced anger condition, previous studies investigating specific associations between anger and visual processing have found that experimentally induced anger has been associated with an increased focus on instrumental (Harper et al., 2010) and dominant (Cabral et al., 2019) goals and rewards (Ford et al., 2010, Lench et al 2015). Induced anger has also been associated with later stages of the SIP such as increased attribution of hostile intent (de Castro et al., 2003), with attention mechanisms believed to act as a bridge between the emotional state and attribution (Peng et al., 2015). Finally, induced anger has resulted in enhanced processing of complex social scenarios (Semmler & Hurst, 2017, Young et al., 2011). High levels of anger are most influential in the development and maintenance of aggression (Gilbert et al., 2013). Indeed, induced aggression has similarly been shown to activate a complex, widespread neural network underlying negative emotions and decision-making (Repple et al., 2017), lasting beyond an aggressive encounter (Bertsch et al., 2009) which is consistent with findings that induced anger impacts multiple levels of processing. However, even though induced aggression has been linked to hypervigilance toward faces and particularly threat-related expressions (Bertsch et al., 2009) behavioural data indicated increased interference (i.e., slower RTs) for emotional faces in a Stroop task.

Studies examining reaction times following emotion induction have demonstrated the effect of induced emotion on information processing. It is important to note that the wide range of methods and ambiguous terms used to describe induced emotional states make it difficult to draw meaningful conclusions regarding which aspects of social processing are impacted and assign this effect directly to the emotional state manipulation. For instance, Quarto et al. (2014) used a gender identification task (NimStim facial expressions) and found significantly slower RTs to faces for participants in a negative compared to happy mood. Sound environments (therapeutic music/noise/silence) were used to continuously induce affective states whilst participants completed the task. Emotional sounds have been shown to influence visual perception; specifically, emotionally touching music has been associated with efficient encoding and faster recall of faces compared to rain noises and joyful music which interfered with encoding (Mado Proverbio et al., 2015). Whilst the effect observed by Quarto et al. (2014) was attributed to affective states, it may have been due to the type of music participants were required to listen to during the experiment. Grubert et al. (2013) found slower RTs in the negative compared to the neutral mood groups during visual search for low-level stimuli (bars). Emotions (happy/sad or negative/neutral) were induced by watching a mood-inducing film.

Few studies have examined the impact of emotional states on WM. Increased WM precision and lower probability of retrieving correct information from WM have been found following anger (Spachtholtz et al., 2014) and negative emotional state induction (Long et al., 2020). Other studies however found enhanced WM precision following negative mood induction with no cost to the quantity of information held in WM (Xie & Zhang, 2017) or no effect of emotion induction on WM precision (Souza et al., 2021). Figueira et al. (2017) used contralateral delay activity, an electrophysiological correlate of maintenance of representations in WM, and found significantly lower amplitudes in an induced negative emotional state, showing decreased WM capacity for task-relevant stimuli. Using behavioural measures, no evidence was found of enhanced or diminished precision of WM templates following emotion induction. Although a previous study by Grubert et al. (2012) found slower RTs during visual search following negative mood induction, this study found the opposite, i.e. faster RTs following anger induction and further enhanced visual search when holding an emotional (angry face in WM). It is possible that our findings are explained to some extent by the use of naturalistic faces which are preferentially attended to and encoded compared to low-level stimuli. However, Maran et al. (2015) found that emotional state induction through exposure to negatively arousing scenes was followed by an equalized performance for all faces whereas low arousal resulted in enhanced storage of angry faces in WM.

A more likely explanation for the present findings, which show an impact of induced emotion condition on both WM and WM modulation of visual attention, is in line with the SIP model of aggression in the context of the hierarchical Gaussian filter framework (HGF; Mathys et al., 2011). That is, emotion and memory being situated at the core of social-information processing means their impact on behaviour is likely a dynamic and holistic one rather than behaviour emerging as a summed output of their individual, isolated contributions. For instance, this study found guidance of attention by WM for naturalistic faces. This is consistent with tenets of the free-energy principle whereby attention is thought to be concerned with estimating uncertainty of probabilistic representations above optimizing what is being represented (Feldman & Friston, 2010). However, the results also showed that the induced emotion condition had an effect on WM modulation of visual perception. A processing advantage was found for angry targets identified whilst holding an angry representation in WM in the induced anger condition. This is in line with memory and emotion being part of a widespread network of cognitive resources employed in continuously generating models of the sensory inputs by reducing uncertainty (Friston, 2009). In line with the research above, this study is the first to demonstrate a potential effect of emotional state on WM and attention processes, and the effect of this three-way interaction on behavioural data.

It is worth noting that although the autobiographical recall task resulted in increased anger in the anger induction condition, higher state anger was also recorded in the induced neutral emotion condition. As can be seen in Appendix D, analyses conducted with participants grouped according to recollection of angry events (into induced anger/neutral groups) were highly similar to the analyses discussed above. Whilst the increase in anger across conditions suggests that anger induction may not be solely responsible for the observed effects, final levels of anger were significantly higher in the induced anger compared to neutral condition. Moreover, it is possible that the pattern of results for the first and second hypotheses reflects activation of anger-inducing internal representations (or schemas) which in turn biased visual processing in favour of stimuli matching participants' recollection of negatively valenced social events. Therefore, results suggest the autobiographical recall task impacted the emotional states of the participants; however, this change may not have been accurately captured by the measures of emotional state used in this study.

III. Does emotion regulation mediate the relationship between hostile attribution, anger, and aggression?

Although WM modulated attentional bias did not predict aggression, later stages of the SIP model such as hostile attribution and endorsement of aggressive responses significantly predicted aggression. The indirect effect of hostile attribution on aggression through trait anger showed that an increased tendency to attribute hostile intent was amplified by trait anger resulting in increased aggression. Similarly, higher trait anger was amplified by a tendency to endorse aggressive behavioural responses and predicted increased aggression. Increased aggression was also predicted by increased emotion dysregulation. Higher emotion dysregulation also predicted an increased WM modulated attentional bias; however, the indirect effect on aggression through the WM bias of attention was not significant. Finally, increased emotion dysregulation was amplified by trait anger and predicted increased aggression.

In line with the hypotheses based on the SIP model which states that emotion permeates each step of social-information processing, participants displaying increased trait anger and deficiencies in regulating their own emotional state wore more likely to report increased self-reported aggression. According to Gratz and Roemer (2004), emotion regulation is a multifaceted concept, encompassing emotional awareness, clarity, awareness, and the ability to engage in actions resulting in positive outcomes. Impairments across these domains are believed to be an indicator of emotion dysregulation and recent research has found robust links between dimensions of emotion dysregulation, aggression (Garofalo et al., 2020), and anger. The link between emotion dysregulation, anger, and aggression is intuitively clear. Being prone to angry emotional states while at the same time lacking the ability to regulate the experience of anger and its impact on actions and possible outcomes.

Previous studies found a relationship between emotion dysregulation, increased hostile attribution, and endorsement of aggressive behavioural responses (Garofalo et al., 2020). This study is the first to identify a link to another core structure of the SIP model, i.e. WM and its effect on visual processing. Although a relationship was found between emotion dysregulation and WM bias of attention, the latter did not predict anger. This suggests that it is not the propensity towards anger as an emotional state that is significantly related to early processing biases in social-information-processing, but rather a limited awareness and capacity to manage one's own emotional states. In comparison to early stages in processing, our findings demonstrate that the relationship between emotion and later stages of the SIP is impacted by propensity toward anger as well as the ability to manage it. The nature of these links suggests that although aggression is predicted by anger and dysregulation, an increase in the ability to regulate one's own emotion and related actions would weaken the relationship between anger and aggression and likely that between anger and later stages of the SIP. In other words, interventions targeting emotion dysregulation would likely impact its links to later stages of the SIP and aggression.

This study showed that although angry faces were preferentially encoded, encoding was not affected by the emotion induction condition participants were assigned to. Conversely, the emotion induction procedure was followed by distinct patterns of WM modulated bias of visual attention.. Of particular importance was the finding that angry targets were identified considerably faster in participants assigned to the anger induction condition holding angry representations in WM, suggesting congruency creates and reinforces participants' expectations. These outcomes demonstrated that emotional states may have an effect on mechanisms of emotion perception and emphasised the importance of studying them in ecologically valid study designs in which they are allowed to interact and model each other as they would in naturalistic social scenarios. Further, this research studied links between early and late stages in the SIP, propensity toward anger and links to emotion dysregulation. This examination revealed a distinction between early stages of social-information processing, i.e. WM modulation of visual perception, which are linked to emotion dysregulation but not to trait anger, and later stages of processing such as hostile attribution and endorsement of aggressive responses which were related to anger and emotion dysregulation. Overall, these findings suggest that induced anger may have an effect on WM modulation of visual perception; however, this effect may not be due to a tendency to experience anger but rather to individuals' ability to regulate this experience of anger and their actions.

4.5.1. Limitations

As this was an online study, it is possible that the anger induction procedure was less effective than a laboratory-based emotion induction procedure. However, previous research has shown anger manipulation to have a similar effectiveness in online compared to in-person research (Lutz, 2016). Although there was a significant increase in anger following the emotion induction procedure, measurements of state anger were taken following completion of the dual task. This could mean that state anger was higher at the beginning of the task and diminished throughout as focus on the cognitively demanding task increased. It is therefore likely that the influence of state anger on WM modulation of visual attention is subject to fluctuation over time. Future studies should investigate the temporal dynamics of current emotional state and interactions with attentional biases. There was an increase in state anger for both induced anger and induced neutral emotional conditions. This may be due to participants misunderstanding the instructions of the task as a few participants recalled anger-inducing events instead of times where they felt little emotion. Future studies using this task should clarify the induction of neutral emotion and the effectiveness of the autobiographical recall task in inducing anger. It is also possible that measuring state anger does not accurately capture the changes in emotional state induced by the autobiographical recall task and that this task induces a broader bias toward aggression related social information. Finally, recording state anger following the completion of the visual task (as opposed to immediately following emotion induction) may have introduced variation in the emotional state responses as emotional states vary over time.

4.5.2. Conclusion

The current research underscores the importance of examining cognitive biases under a unitary cognitive framework. This study investigated the effect of an anger induction procedure on the relationship between WM and visual perception rather than focus on these as separate mechanisms. This approach highlighted the interconnectedness of emotion, WM, and attention, by demonstrating a preferential allocation of cognitive resources toward WM templates and visual targets having an emotional valence which matched the induced emotion condition participants were assigned to. Moreover, the relationships between emotion dysregulation, propensity to anger, and stages of the SIP suggested that emotion is fundamental to social-information processing, impacting all cognitive stages preceding behavioural outcomes. In sum, these findings provide evidence supporting the SIP model as a framework for the study of aggression. This framework is particularly valuable when applied holistically, as core cognitive structures interact with all stages of social-information processing, and biases may be amplified by WM and emotion mechanisms.

5. Chapter 5: General Discussion

5.1. Review of research aims and objectives

The aim of this thesis was to investigate social-information processing biases associated with antisocial traits and behaviours within the framework of the SIP model of aggression. The experiments were designed to examine: 1) attentional bias to threatening faces and whether this is linked to later stages in the SIP model and measures of antisocial traits and behaviours, 2) WM guidance of visual perception, links of bias to later social-information processing stages and measures of antisocial traits in an offender sample, and 3) WM and emotional biases in the SIP model and links to aggression.

To address these lines of enquiry, the first study used an established measure of attentional bias, the DPT, to investigate attentional bias to angry faces. Although the reliability of the DPT was improved by alternative computation of the BS, the task was not a suitable measure of attentional bias as it did not indicate preferential attention allocation to emotional stimuli. The second and third studies used a WM guidance of visual attention to social stimuli paradigm. A combined WM and visual search task was designed to measure the extent to which naturalistic faces held in WM impacted visual search for target angry or neutral faces among neutral distractors. The key findings of the thesis are briefly outlined and discussed below.

- There was no evidence of attentional bias to angry faces as measured by the DPT, regardless of whether they were presented for 100 ms or 500 ms (Chapter 2).
- There was evidence of WM guidance of visual attention in offender and community samples (Chapter 3 and 4).
- WM biases of attention were linked to measurements of aggression and trait anger as well as later stages in the SIP (Chapter 3 and 4).
- Induced anger impacted WM guidance of visual search, suggesting emotional states have an effect on mechanisms of emotion perception (Chapter 4).

- Emotion dysregulation positively predicted increased WM bias of attention to angry faces, hostile attribution, and aggression (Chapter 4).

5.2. Limitations

Prior to outlining the findings across the three experimental studies in the present research, some limitations should be considered. Although the plans of analyses and pre-registered sample requirements resulted in having sufficient participants for well-powered (pre-planned) analyses, exploratory analyses may lack statistical power and care should be taken when interpreting the results. Further, data was collected online for all three studies; this may have resulted in different results compared to RT data collected in laboratory-based settings. However, ample evidence has shown high reliability of RT data collected over the internet (see for example Hilbig, 2016), and although it is possible that noise from the environment will have been captured in the data, this arguably increased the ecological validity of the findings as social-information processing naturally occurs in the presence of numerous distractors.

It is particularly important to note that the emotion induction procedure in the third study was also conducted online. This may have resulted in milder anger induction compared to laboratory-based studies and therefore a weaker effect on processing (Shields et al., 2016). However, Lutz (2015) conducted a laboratory-based experiment in which anger was induced using a frustrating task. They then replicated this study across five online platforms and found highly similar effect sizes across studies. In addition to the above, in the third study (Chapter 4) measurements of emotional state were taken prior to emotion induction and after completion of the visual task. This was designed to preserve the induced emotional state as appraisal of own emotional state has been shown to result in a decrease in intensity (Veenstra et al., 2017). However, as no measurements of emotional state were taken following the emotion induction procedure, it was not possible to determine whether the procedure was successful. These findings should therefore be considered bearing these limitations in mind.

In the first and second study I aimed to recruit participants with a history of antisocial behaviour. Although the first study was advertised on aggression-related pages on social media, we had limited response which resulted in recruitment of participants through university services and paid platforms. This resulted in a mixed sample which may have contributed to the lack of a relationship between attentional bias as measured by the DPT and measurements of antisocial traits and behaviours. However, the link found between WM guidance of attention to angry faces was replicated in former offender and community samples so the likely explanation for the findings in the first study refers back to the discussion on the suitability of the DPT in the measurement of attentional bias to emotional faces. Although the data from the second study was collected from participants who self-reported having served a custodial sentence, offence-related data was not collected meaning no claims can be made as to whether the relationships we have found in the data are related to violent or another type of offenses.

In the second and third studies, naturalistic faces were used for WM templates and targets in the search display. However, these faces were processed to avoid low level confounds. The experiments in the second and third study were designed as EEG studies, the aim being to measure ERPs as indicators of lateralized attention allocation to targets and quality of representation held in WM. As a result, targets were always displayed laterally to the fixation points, and the stimuli were processed (i.e., equalized the histograms and contrasts of the faces), and timed to accommodate requirements for EEG studies. It was not possible to use these designs for EEG studies due to the COVID-19 crisis, during which research laboratories were closed. In addition, this design has likely had an impact on the ecological validity of the task. Still, the present findings replicate existing findings in the fields of attention, WM, and aggression, which underscores the robustness of the novel findings.

For the dual visual task, a simple task design was used in order to maximize statistical power. Specifically, only angry and neutral faces were included in the dual task and the visual search display only included neutral distractors. Whilst the pre-registered hypotheses and proposed analyses were based on previous studies specifically investigating attentional bias to hostility-related information, it is possible that the patterns of preferential attention allocation toward angry faces found in this research reflect a bias toward emotional rather than angry faces. In a recent review of the role of emotional valence in processing of facial stimuli, Kauschke et al. (2019) found an equal distribution of studies finding a processing advantage for happy and angry faces. Further, studies using change detection paradigms found superior maintenance of fearful compared to neutral faces in WM for schematic (Simione et al., 2014) and naturalistic faces (Lee & Cho, 2019). Still a preference toward emotional faces would not explain the emotion induction congruency effect found in the third study (Chapter 4) where angry faces were processed more efficiently by participants assigned to the induced anger condition. Nonetheless, future research should replicate these findings and include further emotional valences such as happiness, disgust, fear, to clarify whether the effect found in the studies outlined here is specific to angry faces or can be generalised across a range of emotions.

Although individuals' proneness toward information matching their assigned emotion induction condition was examined, this research did not examine their ability to explore and adjust their performance based on distractors (i.e., either matching or inconsistent with their internal representations, Smeijers et al., 2019). Therefore, future research should expand the focus of the present enquiry to include WM guidance of visual perception in the presence of emotionally valenced distractors. Further, study designs should include targets and distractors displaying a range of emotions, to better understand whether WM guidance of visual attention differs as a function of emotions encoded or suppressed.

The third study investigated links between the SIP, emotion dysregulation, and how participants' own emotional state impacted their perception of others' emotions. Traditionally, studies of emotions have separated the study of emotion perception from that of emotional states. This is due to an expected influence of the latter on the former. In the real world however, processing of others' emotions is seldom unemotional therefore measuring them in isolation is unlikely to generate behavioural models which can be extrapolated to naturalistic social interactions. Although emotion dysregulation was measured prior to emotion induction and there was a control group for emotion induction, no other emotions were induced (e.g., happiness) and so it is possible that the results are due to an increased emotional experience rather than specific to anger. Still, as previously discussed, these findings extend previous research demonstrating specific cognitive effects associated with state anger.

Cross-sectional designs were used in this thesis which means that it is not possible to draw conclusions as to whether cognitive biases precede or are a consequence of aggression or antisocial behaviours. However, when considering the findings in the context of the SIP and the hierarchical Gaussian filter framework, I propose that cognitive biases and aggression are mechanisms in a reinforcement loop of aggression. Previous examinations of the link between cognitive biases and maladaptive traits and behaviours found evidence for bidirectional relationships (Van Bockstaele et al., 2014), supporting this conclusion. There appear to be changes in the direction of bias throughout development. Kauschke et al. (2019) found a bias toward positive information was predominant in children whereas this gradually changed to a bias toward hostility-related information in adulthood. Their review further supports the concept that aggressive internal representations are acquired and reinforced due to exposure and learning of deviant behavioural patterns. Whilst the causal nature of the link between cognitive biases and aggression is not known, applications of cognitive bias modification paradigms have resulted in reductions in both biases and aggression (AlMoghrabi et al., 2018).

This suggests that due to the bi-directionality of this link, interventions can have an effect on both components. Still, understanding the causality in the relationships between bias and behaviour, as well as how these relationships change during the course of development would be valuable and should be the focus of future research.

5.3. Attentional biases to social stimuli: links to aggression, delinquency, and CU traits.

Albeit subject to the limitations outlined above, this thesis makes a significant contribution to the literature on cognitive biases and links to antisocial traits and behaviours. In the first experimental study, attentional bias to angry faces was examined in relation to measurements of aggression, delinquency, CU traits, and social-emotional information processing. Attentional bias was measured using RTs to a modified DPT, randomly presenting angry and neutral faces for either 100 ms or 500 ms. Trial-level BS which take into consideration temporal dynamics of attention allocation (Zvielli et al., 2015) were computed for the attentional bias analyses and showed considerably higher reliabilities compared to traditional BS for the DPT. The DPT has been used extensively as a measure of attentional bias and previous research has demonstrated attentional biases to angry faces using the DPT in a range of samples (Bar-Haim et al., 2007, Van Rooijen et al., 2017). However, studies have increasingly called into question its validity (Rodebaugh, 2016) and suitability as a measurement of spatial attention allocation to emotional faces (Puls & Rothermund, 2018). Using the trial-level BS, no evidence was found of an attentional bias to angry faces for either target presentation duration and no links of bias to measurements of aggression and CU traits.

Due to significant positive skew in measurement of self-reported delinquency, participants in the first study (Chapter 2) were divided into high and low delinquency groups. The latter showed an attentional bias to angry faces when these were presented for 100 ms. Further, an examination of links of attentional bias to later stages in social-information processing showed that an increased attentional bias to angry faces was predicted by a lower self-reported likelihood of engagement in aggressive behavioural responses. The increased focus on task goals and decreased attention allocation to emotional valence of the targets suggested an enhanced, homogenous performance on the DPT for participants reporting high delinquency or increased propensity to behavioural aggression. According to the SIP model of aggression, individuals displaying higher aggression should show increased attentional bias toward potentially threatening or hostile stimuli (Dodge et al., 2013) although some studies have found patterns of avoidance associated with antisocial traits and behaviours (Lobbestael et al., 2016, Bacon et al., 2018).

The DPT is a simple cognitive task meaning that if the stimulus is presented for a sustained amount of time, participants are able to disengage from the targets (Isaacowitz et al., 2006) and focus on task demands, attend to other stimuli (distractors), or disengage from the task. Whilst stimulus presentation durations were varied and randomised throughout the task to increase engagement, behavioural results may not be sufficient to draw meaningful conclusions regarding participant engagement with the DPT and therefore of any attentional bias to threatening faces. Notwithstanding the increasing literature showing low validity of the DPT and low reliability of traditional computation of BS, this task is currently extensively used in the study of attentional biases (Ciucci et al., 2018, Jennes et al., 2021).

There was no evidence of an attentional bias in individuals displaying higher aggression and delinquency. An increased attentional bias toward angry faces was however found in the low-delinquency group and predicted by a decrease in aggressive behaviour. This added to existing literature showing no evidence of attentional bias to angry faces across visual tasks in relation to antisocial traits (Putman et al., 2004, Schwenck et al., 2014). However, these findings could also suggest that individuals in the higher-delinquency group and individuals self-reporting higher aggression either avoided engagement with the emotional faces or disengaged from them faster, resulting in a consistent performance across conditions in the DPT. Whilst it is possible that participants displaying higher antisocial behaviours in this sample disengaged with the emotional valenced stimuli to focus on task demands as they did not consider them to be real threats, RT data from the DPT does not offer sufficient information to draw meaningful conclusions.

5.4. Working memory biases of visual attention to social stimuli: links to anger and aggression

Attention and memory are fundamental in processing visual information and although their mechanisms are usually the object of independent enquiries, research has shown they are intimately linked (Olivers et al., 2011). There is considerable overlap in brain regions that are activated during encoding and maintenance of information and the two processes share similar capacity limits (see Olivers, 2008 for a review). Moreover, WM templates that are the focus of executive processes, i.e. directly relevant to perception or an ongoing task have been shown to bias visual processing (see Soto et al., 2008 for a review). Studies investigating WM guidance of visual perception have mostly used feature based visual search tasks (asking participants to search for features such as shapes or colours, Olivers et al., 2011); however, this paradigm has most recently been extended to naturalistic objects and scenes (Peelen & Kastner, 2014). Tenets of the SIP model share a number of similarities with existing literature on memory guidance of visual attention. Within the SIP model, memory is conceptualised as influencing all stages of social-information processing, from encoding to decision making to reinforcement of internal representations based on continuous learning in social interactions (Wilkowski & Robinson, 2008). Given this conceptual overlap and limitations of the DPT, the second study broadened the scope of the initial enquiry to investigate how induced emotional WM templates biased visual perceptions and how this related to subsequent steps in the processing of social information.

For the second experiment, a visual task was designed for the study of WM guidance of visual attention for naturalistic faces. Previous research has linked aggression to an increased attentional bias to threatening stimuli (Manning, 2020) and higher incidence of aggressive internal representations (Day et al., 2021). Given that recent attention and WM studies have highlighted a robust effect of representations actively held in WM on visual perception (Kumar et al., 2009), this study aimed to examine the extent to which emotionally valenced WM templates guided visual search. Further, this bias was examined in a former offender sample, in relation to measures of aggression, trait anger, and later stages in the SIP model. Building on the results from the first study, BS were created initially as scores of difference between conditions (i.e., neutral and angry WM templates, neutral and angry targets in the visual search display); however, similar to the DPT traditional BS computations, these had low reliabilities. Alternative trial-level BS were therefore used, taking into account the dynamic (as opposed to trait like) nature of attention allocation. These showed considerably higher reliability.

As expected, RT based analysis showed WM templates guidance of visual search extended to naturalistic faces. This was in line with previous studies demonstrating visual perception guidance by WM using shapes (Soto et al., 2005) and real-world objects (Berggren & Eimer, 2018). The primary role of visual attention is to focus cognitive resources on salient objects in our environment (Peelen & Kastner, 2014). As such, objects and contexts we are repeatedly exposed to are likely to fundamentally shape our perception of the environment through perceptual learning and top-down biases of visual perception. This has been demonstrated at neural level (de Beeck & Baker, 2010). Visual search has traditionally been studied in laboratories utilising study designs aimed at reducing distractors and isolating specific aspects of visual search which has resulted in drastically reduced complexity of visual searches compared to the real world. Although visual searches can be directed by simple features such as vertical lines or specific colours (Soto et al., 2007), individuals are unlikely to conduct most visual searches based on such features. Naturalistic visual scenes are cluttered and the targets of the search are typically familiar, acquired through learning processes (Woodman et al., 2013). Our second study has therefore extended the WM guidance of visual search paradigm to visual search for social stimuli, as part of the process of selecting relevant information from our social contexts, from among distractors.

In addition to the guidance of attention toward faces matching WM templates, the second and third studies both found an effect of the emotional valence of the WM template; that is, participants found all targets in the visual search display faster when holding an angry face in WM. They were also significantly more likely to remember angry faces encoded in WM compared to neutral faces. This showed preferential encoding and maintenance of faces in WM and a direct effect on visual perception across offender and community samples. Previous studies have found a bias toward faces displaying negative emotions (i.e., anger, fear, or disgust, Hansen & Hansen, 1988) have drawn on the evolutionary relevance of a processing advantage for negative facial expressions (Pinkham et al., 2010). However, most studies finding this advantage have used schematic faces (see for example Schlaghecken et al., 2017), meaning that the emotions may have been exaggerated compared to naturalistic faces. The present research found a processing advantage both in terms of RTs to the visual search task and accuracy in the WM task demonstrating a processing superiority for emotional (angry) faces. Evidence from EEG and fMRI research appears to support this bias toward negative stimuli (Norris, 2021).

In a recent review, Kauschke et al. (2019) further found that children showed a processing advantage for positive emotions and that this changed throughout development into a processing advantage for negative faces. Whilst this was more pronounced for emotionally valenced words compared to facial expressions (Bahn et al., 2017) these findings suggested biases toward negative stimuli can be learned through exposure. This concept supports the SIP model which posits that negative internal representations (i.e., schemas, memories acquired through experience of negative events) alongside emotions create blueprints for social interactions which can bias individuals' perceptions and increase the probability of aggressive responses to ambiguous scenarios. These in turn reinforce pre-existing aggressive schemas (Caramaschi et al., 2008). The enhanced performance on the visual search task found when an angry face was held in WM not only supports the studies summarised above, but also suggests that this negativity bias can result in hypervigilance. That is, the processing advantage applied to angry faces during encoding and maintenance in WM is extended to visual search for targets. These findings underscore the importance of addressing the close links between WM and attention in study designs.

The second study also extended the links of visual bias to measure associations with anger and aggression and found a significant correlation between reactive aggression and trait anger. Whilst the association between anger and reactive aggression has been found in previous studies (Veenstra et al., 2017), to include both measures in our analyses would have resulted in examination of their unique variance in our pre-planned regression models, whereas this research focused on the contribution of their overlap to the aggressive outcome as well. A composite reactive aggression and trait anger measure was created and results showed that an increase in the composite variable predicted a significant increase in attentional bias toward angry faces identified whilst holding an angry face in WM. This showed that higher aggression and trait anger predicted an increased bias of attention to angry faces whilst holding an angry face in WM.

Our findings support previous research showing preferential allocation of cognitive resources toward angry faces, and that this bias is increased with higher anger and aggression. That is, individuals displaying higher trait anger and reactive aggression increasingly allocated significantly more attention to angry targets in the visual search displays when required to hold angry faces in WM. Anger prone individuals have consistently shown biased patterns of visual perception (Mellentin et al., 2015) and increased attentional bias to negative stimuli has been shown to predict aggression (Brugman et al., 2016) across forensic and community samples (Smith & Waterman, 2003). Moreover, reactive aggression has been associated with a tendency to approach rather than avoid aggression related scenes (Lobbestael et al., 2016) and stimuli (Veenstra et al., 2018). In the context of the literature reviewed above, these findings suggest that in a social situation, individuals displaying higher trait anger are more likely to experience situational anger, to disproportionally allocate cognitive resources toward negatively valenced emotional stimuli, and to engage in behaviourally aggressive responses.

WM contents guide visual perception toward stimuli closely resembling the WM templates (Soto et al., 2012) and the present research has shown this extends to naturalistic, emotionally valenced stimuli. Further, an internal focus on implicit hostile cues has been shown to increase neural responsivity to angry faces (Buades-Rotger & Krämer, 2018) and ruminative attention, that is, increased internal attention to hostile information, amplifies feelings of anger and leads to increased aggression (Wilkowski & Robinson, 2008). Although the SIP model posited that memories and internal representations acquired through learning and reinforcement of existing social schemas impact stages in social-information processing as early as encoding, the present research provides evidence of this guidance. Moreover, it adds to a growing body of literature linking rumination to attentional biases, psychopathology, and increased likelihood of aggressive outcomes (Veenstra et al., 2018).

The immediate effect of internal representations on perception has been studied under various frameworks, with fMRI studies showing internal representations guided spatial attention (Lepsien et al., 2005). Further, EEG study demonstrated WM guidance of visual search (Kumar et al., 2009), and behavioural studies showed WM guidance of attention to threat in relation to anxiety (Yao et al., 2019). In contrast, the research addressing the relationship between internal representations, perception, and aggressive behaviour is scarce. Latent knowledge structures, (e.g., aggressive schemas) and their relationship with aggressive or antisocial behaviour have been studied with a specific focus on developing interventions for violent offenders (Day et al., 2021). Maladaptive schemas or sets of underlying cognitive styles have been directly linked to the development and maintenance of aggression (Tremblay & Dozois, 2009), increased anger and psychopathology (Dunne et al., 2019). However, recent research suggested this link is bidirectional, with deviant social-information processing patterns predicting increased aggression whilst the latter predicted worsening of deficiencies in social-information processing (Orue et al., 2021).

Whilst a correlation has been found between aggressive schemas and behaviour including in forensic samples (Dunne et al., 2018), less attention has been given to the mechanisms through which internal representations impact perception in social situations in real time. This is important because WM biases of visual perception are mediated by WM recruitment of relevant perceptual mechanisms. For example, occupying WM with a set of locations impaired visual search (Woodman & Luck, 2004) whereas maintaining colours in visual WM did not impair visual search (Woodman et al., 2001). Based on evidence suggesting overlap in sensory mechanisms across WM and attention results in interference (Chun, 2011), occupying WM with hostility related information may result in interference when processing hostile stimuli. This interference however appears to become guidance or increased capture of cognitive resources for stimuli closely matching WM templates, which is in line with findings from chapters 3 and 4. For instance, Seidl-Rathkopf et al. (2015) showed that attention can be guided to locations held in WM and Soto et al. (2005) demonstrated similar guidance toward object matching WM templates. Rather than discount the interference effect, these studies show that the effect of top-down guidance of attention is the result of neural mechanisms being unitarily employed in the preferential processing of information matching internal representations.

When processing real-world scenes, it is not sufficient that relevant information is enhanced during visual selection, mechanisms of distractor suppression must also be successfully engaged (Seidl et al., 2012). Both processes have been linked to task relevance with studies showing that information which is not actively maintained in WM or perceived as relevant to task goals does not guide perception (Carlisle & Woodman, 2011, Carlisle & Woodman, 2013). The second and third study demonstrated that WM biased visual perception for naturalistic faces for both negative and neutral targes surrounded by neutral distractors. Extrapolating the WM guidance of visual attention paradigm to emotionally valenced naturalistic social stimuli demonstrated the link between hostility-related WM templates and biases of visual perception for potentially threatening stimuli. This suggests that the effect of internal representations on behaviour could be measured in real time through interlinked mechanisms of attention and memory. Moreover, this paradigm could be applied to the study of real-world social scenarios where an active, hostile WM template could narrow the cognitive scope toward information closely matching this representation, resulting in increased focus on information perceived as hostile.

Increased aggression and anger have been shown to further enhance the narrowing of the attentional scope toward angry targets matching angry internal representations whilst successfully ignoring distractors. Peng et al. (2015) found that anger increased attribution bias through narrowing of attention scope toward salient information and away from distracting information. In the same vein, previous research has shown that higher aggression was linked to enhanced top-down disregard for information which did not match actively held interpretations of social scenarios (Horsley et al., 2010). In the third study, participants in the induced anger group showed faster RTs to all targets across emotional valences of WM templates and targets. In a real-world social interaction these findings may translate into increased attentional spotlight on information relevant to task goals, with increased aggression and trait anger being associated with an increased disregard for information that does not match a hostile interpretation of ambiguous social cues. In turn, such a pattern of bias toward cues perceived as hostile would likely increase the odds of an aggressive outcome, particularly in individuals with a predisposition to anger.

5.5. WM and emotion mechanisms in the SIP model of aggression in a community sample

Emotion is a core component of the SIP model of aggression (Crick & Dodge, 1994). The second study indicated an increased bias toward angry faces matching WM templates in relation to aggression and anger. However, real-world interactions in which individuals would display such cognitive bias toward hostility-related social information would likely be emotionally charged. Using an emotion induction paradigm and the visual task developed in the previous study, the third study examined the effect of induced anger on encoding and maintenance of WM templates as well as on the WM bias of visual perception. An angry emotional state has previously been shown to impact both WM (Maran et al., 2015) and attention processes (Gable et al., 2015). However, to date, no studies have addressed the effect of emotional state on the relationship between attention and WM. Hence, the third study investigated both whether induced anger led to enhanced precision in encoding of emotional faces in WM, and its effect on WM modulated visual search for angry and neutral faces among neutral distractors.

Although previous studies found that induced anger resulted in higher precision during encoding as well as increased attention to complex social scenarios (Semmler & Hurst, 2017), this precision appeared to come at the cost of processing speed (Grubert et al., 2013). Conclusions with regards to emotion induction are limited in the present research as caveated in the limitations sections above. Emotion induction condition did not have an effect on precision during encoding and maintenance of WM templates for either neutral or angry faces. However, participants in the induced anger condition identified targets in the visual search display faster than participants in the neutral condition, suggesting an increase in focus on task goals and away from distractors (van Steenbergen et al., 2011, Young et al., 2011). Similarly, Ford et al. (2012) found that following anger, neutral, and fear emotion induction, participants in the anger condition displayed selective attention to rewarding but not threatening images when compared to the other conditions. Induced anger has also been associated with increased endorsement of instrumental goals (Harper et al., 2010) and higher motivation to achieve goals (Lench et al., 2015).

Previous studies have also found slower RTs in negative mood conditions for emotionally valenced words (Cohen et al., 1998) and during visual search for objects (Grubert et al., 2013). In a study by Bertsch et al. (2009) slower RTs were found toward all (happy/angry/fearful/neutral) emotional expressions in the induced anger group, in participants high in trait anger; however, in the same study, EEG data showed an augmented neural response (i.e., hypervigilance) to angry faces following anger induction. This pattern of results may be explained by the nature of the task, as Bertsch et al. used an emotional Stroop task which required participants to attend to coloured emotional faces which were masked until participants could name the colour of the face. Hence, the behavioural and EEG results suggest that participants in an angry emotional state not only displayed hypervigilance to angry faces but also continuous engagement with emotional faces which was prioritised over encoding of colour, thereby becoming disadvantageous to task goals.

The results in the third study also showed that WM templates matching the assigned emotion induction condition of participants had an impact on visual search. That is, although participants in the induced anger condition were generally faster to identify targets, this effect was further augmented when holding an angry face in WM. However, few studies have investigated the interaction between emotional states and perception of others' emotions. In particular, positive emotional states have been shown to direct attention toward positive stimuli (Wadlinger & Isaacowitz, 2006) whereas negative affect mostly predicted poor performances on emotional Stroop tasks (Crocker et al., 2012, Hur et al., 2015). Notably, negative affect has been broadly used to encompass a range of emotional states such as depression, frustration, sadness and these have distinct neural and behavioural correlates compared to induced anger (Kreibig, 2010). The influence of anger on cognitive processes such as memory and executive function differs from other negative emotions. Comparisons of dimensions of negative affect have shown that anxiety impaired executive function (Shields et al., 2016) and decision making (Lerner & Keltner, 2001) whereas induced anger has not been linked to executive function or information processing impairments (Moons & Mackie, 2007).

In the third study, the most efficient performance on the visual task was found in the induced anger condition, for angry targets identified when holding a matching (i.e., angry) compared to a neutral face in WM. Targets were identified faster compared to all other conditions. For example, participants in the induced anger condition holding an angry face in WM were significantly slower to identify disconfirming (i.e., neutral) targets. These results demonstrate a top-down driven superiority effect for processing social-emotional stimuli even in the presence of distractors. Further, the present findings support previous studies indicating increased processing efficiency for anger-related information, especially when identifying others' emotions (Brennan & Baskin-Sommers, 2020).

5.6. WM biases of visual perception in the SIP model: anger enhances processing until it does not

Enhanced processing of information in participants assigned to the anger induction condition may be due to the "angry spotlight" (Ford et al., 2012) which directs cognitive processes toward rewarding information (Szymaniak & Zajenkowski, 2021). If induced anger results in superior processing of information in complex social scenarios (Semmler & Hurst, 2017) or for information which disconfirms previously formulated hypotheses (Young et al., 2011), it is important to consider when an angry emotional state becomes maladaptive. A robust link has been found between increased trait and state anger and aggression or violent behaviour (Wilkowski & Robinson, 2010). Gilbert et al. (2013) found that offenders showed an increased tendency to experience anger more intensely and frequently and that increased anger and endorsement of violent beliefs predicted aggression.

In the second study, increased trait anger and reactive aggression predicted enhanced processing of angry faces when holding an angry face in WM. The third study similarly found that participants in the induced anger condition displayed enhanced performance on the visual task, particularly when both WM templates and targets consisted of angry faces. To understand how measurements of trait anger and aggression increase the probability of an aggressive behavioural outcome, their links to later stages in the SIP model were investigated. In the second study, response enactment measurements, that is, likelihood of engaging in a given behavioural response, for relationally and directly aggressive responses were positively correlated to anger and reactive aggression. Hostile attribution was also positively correlated to trait anger. These relationships have been found in previous studies (Verhoef et al., 2021, Fontaine et al., 2009, Fontaine et al., 2010). However, the present research showed that response enactment for aggressive responses and hostile attribution predicted encoding biases, i.e. WM guidance of attention allocation to angry faces, providing evidence to support previously hypothesized links within the SIP model between WM, emotion, and steps in social-information processing.

Correlational analyses in the second study also showed that an increased negative emotional response to ambiguous social scenarios predicted bias away from angry faces. This variable measured how likely participants thought they were to feel angry or upset following an ambiguous social interaction. Associations between emotional responses and social-information processing have previously been studied in the literature. In particular, the ability (or lack thereof) to regulate one's emotional response has been explored in relation to regulation of anger and its associations with aggressive outcomes (Roberton et al., 2012). Emotion dysregulation has been associated with both externalizing and internalizing problems (Neumann et al., 2010). Although there is an established link between anger and aggression, to assume that increased anger leads to aggression is an overly parsimonious model (Roberton et al., 2012) as the former appears to become maladaptive when self-regulation fails (Simons et al., 2016).

In the third study, correlational analyses showed associations between response enactment for aggressive responses, hostile attribution, negative emotional responses, and WM bias of visual perception. Further, the same steps of the SIP model were linked to trait anger and aggression, replicating the links between externalizing measures, WM bias of attention, and later stages in social-information processing found in the previous study. Emotion dysregulation was included in the final study. This was positively correlated to both anger and aggression, and positively correlated to cognitive processes within the SIP model (i.e., encoding and attribution of hostile intent). Further, both trait anger and emotion dysregulation predicted aggression. Anger predicted aggression through later stages in the SIP model and emotion dysregulation predicted increased hostile attribution. This supported previous research highlighting links between emotion dysregulation, trait anger, and aggression (Garofalo et al., 2020).

Increased emotion dysregulation, i.e. inability to downregulate negative affect has been associated with poor behavioural control, aggression (Bertsch et al., 2020), violence (Tager et al., 2010), and psychopathy (Garofalo et al., 2018). Behaviour and emotion control are mediated by a broad, overlapping network of brain regions (Repple et al., 2017). These are involved in monitoring, selecting, mediating behavioural and emotional responses (Bertsch et al., 2020), meaning that poor emotion regulation plays an important role in enhancing the possibility of aggressive outcomes to potentially hostile situations. Whilst negative emotional states have an effect on cognition and can result in aggressive or violent outcomes (Bertsch, 2020), adaptive emotion regulation strategies are usually employed to prevent such outcomes. These strategies appear to be impaired in individuals with psychopathic traits (Garofalo et al., 2018) and in individuals with a history of externalizing problems and substance use (Simons et al., 2016). Only a handful of studies have investigated the role of emotion regulation strategies in SIP processes (Smeijers et al., 2019). Similar to our findings, these studies suggested that emotion dysregulation was positively related to hostile attribution, anger, aggressive response generation, and endorsement of aggressive beliefs.

5.6. Implications for the SIP framework

Social-information processing biases in the SIP appear to have a cumulative, consolidating effect, building up to aggressive outcomes. Firstly, this is illustrated by the extensive associations between externalizing traits and measurements of later stages in the SIP which was found in the present research. The first study found links between aggression, CU traits, and later stages of the SIP model. That is, aggression was associated with increased negative emotional response, attribution of instrumental goals, and a predisposition to behavioural aggression in ambiguous social situations. Although CU traits were not linked to attentional bias, they were linked to reactive aggression, attribution of instrumental goals and endorsement of socially appropriate behavioural responses. This pattern of results suggested CU traits are associated with a combination of instrumental and non-instrumental goals and forms of aggression (Fanti et al., 2009) as demonstrated by the increased focus on setting and achieving goals during the course of social interactions. The second and third studies found links between anger, aggression and response enactment variables for aggressive responses, instrumental attribution, and negative emotional response. These relationships showed that individual differences in antisocial traits and behaviours were related to cognitive process throughout the SIP model.

Externalizing measures were also associated with hostile attribution, i.e. attribution of hostile intent in ambiguous social situations. The first study did not find a relationship between hostile attribution and CU traits although small correlation was found between the two in the former offender sample (study 2). Further, hostile attribution was not correlated to reactive aggression in the first or second study but a positive correlation was found to reactive-proactive aggression scores in the third study. Although these results suggested a broad association between hostile attribution and aggression, previous research highlighted specific links to reactive aggression (Gagnon & Rochat, 2017) with reactive aggression decreasing following hostile attribution bias modification (Van Bockstaele et al., 2020). However, studies investigating the specificity of these associations have used a range of measures of reactive and proactive aggression, some of which show higher overlap. For example, Richey et al. (2016) found that hostile attribution mediated the association between a history of maltreatment and reactive (but not proactive) aggression. They used Dodge and Coie's (1987) measure of reactive-proactive aggression and the two dimensions of aggression only shared 20.5% of their variance. Comparatively, reactive and proactive dimensions of aggression as measured in our studies using the RPQ showed a considerably higher proportion of shared variance with correlations over 0.60. These findings show that hostile attribution relates to measures of aggression and antisocial traits; however, this relationship is mediated by measurements and sample types.

The cumulative effect of biases in the SIP is further evidenced in later social-information biases such attribution biases which follow perceptual biases, resulting in increased endorsement of aggressive behavioural responses (Smeijers, 2019). In our studies, hostile attribution did not predict attentional bias to angry faces as measured by the DPT. However, in the second study, an increase in hostile attribution predicted bias away from neutral faces when holding an angry face in WM, indicating a pattern of attention allocation away from stimuli which could not be reconciled with hostile internal representations. Finally, in our third study, a positive correlation was found between WM biases of visual perception and hostile attribution as well as indirect effects of the latter on aggression through anger and final stages of the SIP (i.e., response enactment for aggressive responses). Within the SIP model, attribution of intent is closely linked to the encoding stage (Zajenkowska & Rajchert, 2020). That is, attribution of intent must follow attention allocation to relevant social information (Horsley et al., 2010); however, eye-tracking studies have shown evidence of preparatory attention by pre-existing hostile or aggressive schemas (Wilkowski et al., 2007).

Some studies claim that once the encoding step of the SIP is biased, the following steps are affected as well (Smeijers et al., 2019). However, given the links between hostile attribution, trait anger (Zajenkowska & Rajchert, 2020), and emotion dysregulation, an increased likelihood of an aggressive behavioural outcome is predicted by more than a cascading effect of bias in the SIP model. Namely, the effect of hostile attribution on reactive aggression has been shown to be mediated by a tendency to act impulsively when experiencing intense negative emotions (Gagnon & Rochat, 2017). According to the Integrated Cognitive Model of anger and reactive aggression (Wilkowski & Robinson, 2008), individuals high in trait anger are biased toward hostile interpretation of ambiguous situations which leads to anger expression. Following the formulation of hostile interpretations, individuals are expected to engage in rumination and reappraisal whilst recruiting effortful control to downregulate anger or aggression. Given that rumination has been shown to amplify state anger and hostile attribution (Wang et al., 2019), a lack of effortful control results in increased risk of aggressive outcomes (Morales et al., 2016).

In relation to anger, effortful control refers to individuals' ability to downregulate their experience of anger, and to suppress anger-escalating thoughts and aggressive impulses (Denson, 2009). Low effortful control has been linked to increased risk and severity of aggressive outcomes (Denson et al., 2012). There are numerous similarities between the concepts of effortful control in the context of anger and aggression and emotion dysregulation. Specifically, they both involve awareness, acceptance, and control of emotions, as well as strategies to regulate emotional response to avoid unwanted outcomes (Preston & Anestis, 2019). It is therefore possible that following a hostile interpretation of an ambiguous social scenario, individuals prone to anger would engage in rumination which increases state anger. Focusing on one's internal state, particularly if this is an angry state has been shown to increase anger but also result in higher levels of processing (Stavraki et al., 2021). However, if these individuals also have difficulties controlling their emotional state and the extent to which it impacts their behaviour, they would be considerably more likely to engage in aggressive behaviours (White & Turner, 2014). In line with this evidence, positive correlations were found between hostile attribution and trait anger in the second and third studies, indicating a relationship in both offender and community samples. In the third study, links were also found between hostile attribution, emotion dysregulation, and aggression.

If we consider the SIP model in the context of the hierarchical Gaussian filter framework (Mathys et al., 2011), these findings suggest that emotion and memory are embedded in the SIP framework, having a holistic rather than segregated impact on stages of social-information processing. Under this framework, the brain is believed to continuously generate a predictive model based on sensory inputs, one which is constantly optimized to reduce uncertainty and increase accuracy of predictions (Mathys et al., 2014). This uncertainty can be reduced either by adjusting one's beliefs in light of new information from the environment, or by changing the environment to sample information to suit one's beliefs (Friston, 2010). The results from these studies can be mapped onto the SIP model and the hierarchical Gaussian filter. That is, the core of the SIP model consisting of a database comprising memories, acquired rules, social schemas, is believed to interact with steps in social-information through emotion (Crick & Dodge, 1994). This core is reinforced or adjusted depending on one's experiences; however, these are subject to perceptual and attribution biases.

What this means is that core beliefs and aggressive schemas can exert an influence over various stages in processing and typically this influence is to favour information congruent with existing beliefs. For example, aggressive internal representations (e.g., aggressive schemas) have been shown to guide attribution of intent (Wilkowski & Robinson, 2009), response formulation (Fontaine et al., 2010), and in the present research, visual perception. In this research, WM guidance of attention toward hostility-related information was predicted by increased aggression and trait anger, hostile attribution, and emotion dysregulation. Aggressive outcomes to ambiguous social situations in turn are believed to reinforce the pre-existing aggressive internal representations possibly strengthening their guidance of SIP processes so that it becomes increasingly automatic (Wilkowski & Robinson, 2010).

In line with the free-energy principle (Friston, 2009), internal representations should be constantly appraised and updated to increase the precision of the predictions individuals make about the world. According to this framework, sensory representations for social stimuli, which we have shown can be biased by induced internal representations, would be the first step in social interactions to be used to update social blueprints. The cumulative effect of biases in social cognition would have a similar effect in reinforcement of internal representations and result in patterns of escalation in aggressive behaviour for individuals who have a history of aggression and also display high aggressive traits, trait anger, and cognitive biases toward hostility-related information.

Individual differences in externalizing measures, proneness to anger and ability to regulate emotion can either have adaptive or maladaptive effects. That is, a situation in which an individual with a history of aggression interpreted someone's intentions as hostile in an ambiguous social situation, which increased their state anger, could be defused by adaptive emotion regulation strategies or a decreased proneness to experience state anger. However, as outlined above, research so far has found that proneness to aggression, anger, and higher emotion dysregulation amplify cognitive biases, escalating the probability of maladaptive emotional and behavioural responses in ambiguous social environments.

The present research underscores the value in examining the model as a whole, particularly accounting for the influence of variation in individual differences. This holistic examination showed that the influence of internal representations can be extensive, as early as visual perception and links to later stages of the SIP model. In the first study (Chapter 2), avoidance of angry faces was linked to delinquency and increased likelihood of engaging in aggressive behavioural responses in response to ambiguous social provocation. The second study (Chapter 3) showed that the link between attention and aggression was mediated by WM guidance and that increased hostile attribution predicted avoidance of targets which did not match angry faces held in WM. The third study (Chapter 4), highlighted the relationship between WM guidance of visual attention and emotion dysregulation, two core components of the SIP model of aggression.

Throughout the three studies, links were found between SIP measures and antisocial traits and behaviours. For example, CU traits were positively correlated to increased likelihood of enacting socially appropriate behaviours in response to ambiguous provocation in the first study, and positively correlated to increased likelihood of enacting aggressive behavioural responses in the former offender sample (study 2). In addition, hostile attribution was not correlated to aggression in studies 1 and 2 but predicted aggression through trait anger and increased likelihood of enacting aggressive behaviours in response to ambiguous social provocation (study 3). Finally, there was a significant positive correlation between aggression and increased negative emotional response following an ambiguous social provocation (study 1). This was supported by findings from study 3, where emotion dysregulation predicted not only an increase in aggression, but also increased WM guidance of visual attention to threat, increased hostile attribution, and increased trait anger. This showed that both emotional states and the ability to regulate their intensity and impact on behaviour significantly contributed to the likelihood of aggression.

5.7. Clinical and practical implications

Given these relationships within posited SIP mechanisms and between the SIP model and individual differences in proneness to anger, aggression, and emotion dysregulation, it is important to consider clinical and practical implications. Disentangling biases in social-information processing into smaller cognitive units can enable the study of the individual contribution of steps in the SIP to aggressive outcomes (Li et al., 2013). The SIP framework has previously been used in intervention research (AlMoghrabi et al., 2021) and the findings in the three experimental studies above highlight a number of potential intervention approaches.

For instance, the link between social cognition, hostile attribution, and aggression has been targeted in the last decade, in children (Houssa et al., 2017), adolescent (Merrill et al., 2017), and adult samples (AlMoghrabi et al., 2021). Although this association is less well understood in adults, a recent review by Klein Tuente et al. (2019) found the relationship was present across population types. These associations have also been found in the present research across sample types, suggesting that similar interventions can be used to target the relationship between cognitive biases and ASB.

SIP interventions have been aimed at reducing impulsiveness (Rabiner et al., 1990), and altering encoding and attribution biases leading to aggressive behaviour (Li et al., 2013). Dodge et al. (2013) investigated the effect of social-cognitive elements in preventive interventions for children with conduct problems. They found that reducing hostile attribution, decreasing endorsement of aggressive beliefs, and increasing benign response generation constituted 27% of the intervention impact. In an adolescent sample, hostile attribution bias training showed a decrease in hostile attribution as well as decreased reactive (but not proactive) aggression (Van Bockstaele et al., 2019). In adults, the SIP framework has been applied to trauma-related research (van Reemst et al., 2016), interventions to reduce impulsive aggression for individuals with intermittent explosive disorder (Coccaro et al., 2017), and interventions leading to a decrease in hostile interpretations and lower anger expression (Cougle et al., 2017). Similarly, associations between hostile attribution and aggression have been found throughout this research, suggesting that interventions targeting this relationship would be beneficial across demographics and sample types.

Although attentional biases and emotion dysregulation have been recognised as potential focuses of interventions in children and adolescents with aggression problems (Neumann et al., 2010), there is little applied research available (Santone et al., 2020). For example, Hiemstra et al. (2019) showed that cognitive bias modification reduced hostile interpretation of ambiguous facial expressions however this change was not reflected in measures of state anger and aggression. Emotion recognition training has recently been replicated in an adult sample of violent offenders (Kuin et al., 2020). Results showed that the training increased perception of happiness over anger although there were no changes in measures of aggression and hostility. Attribution/interpretation bias modification does not appear to have an effect on attentional biases to negative cues (AlMoghrabi et al., 2021). These findings demonstrate a pattern of results whereby interventions produce change; however, this change is specific to certain perception mechanisms or measurements of antisocial traits and behaviours. In line with these studies, the present research points toward separate mechanisms for top-down biases for visual perception and interpretation biases. Further, findings in the third study indicated specific links between emotion and visual processing/interpretation biases, suggesting that

The third study highlighted relationships between emotion dysregulation and WM guidance of visual attention as well as hostile attribution, proneness to anger, and aggression. This suggests interventions targeting emotion regulation strategies may have an effect on biases in social information processing as well as on aggressive outcomes. As previously outlined, maladaptive emotion regulation has been associated with a history of aggressive behaviour across population types (Verhoef et al., 2021), indicating the value of an intervention increasing individuals' ability to regulate their emotions adaptively (Roberton et al., 2014). To date, only a handful of studies have investigated emotion regulation interventions in aggression. Hesser et al. (2017) found that a cognitive behaviour therapy intervention comprising emotion regulation techniques reduced intimate partner violence. Although this change was partially attributed to emotion regulation training, it is unclear to what extent this contributed to the self-reported decrease in emotional abuse.

Nesset et al. (2021) conducted a randomised control trial whereby perpetrators of intimate partner violence were randomly assigned to cognitive-behavioural therapy or mindfulness-based stress reduction therapy. At 12 months' follow-up, they found a reduction in emotion dysregulation in both groups. Both types of therapy asked participants to reflect on their past experiences and consequences and were designed to increase participants' awareness of their own emotions and reframe these to avoid aggressive outcomes. These are dimensions of adaptive emotion regulation, and may explain why a small but significant decrease in emotion dysregulation was observed in both groups. Emotion regulation training has also been linked to a decrease in anger in an adolescent sample (Lotfali et al., 2017) and individuals with a history of substance abuse (Massah et al., 2016). These findings underscore the importance of targeting emotion dysregulation and whilst initial findings demonstrate effects on anger regulation, it is likely targeting these mechanisms would lead to lower aggression.

The present findings have wider implications for clinical research into anxiety disorders and related fields relying on attentional bias and attentional bias modification research. The DPT is the most frequently used paradigm in anxiety-related research (Wieser et al., 2018), has been used in hundreds of studies, and is widely used in clinical research (Price et al., 2015). However, a review of studies using the DPT to examine attention to emotional faces showed low validity and a lack of a significant interaction with manipulated factors, suggesting that the DPT may not be a suitable measure of attention to emotional faces (Puls & Rothermund, 2018). Even though traditional BS computations have been shown to have low reliabilities (Rodebaugh et al., 2016), increased attentional bias to angry faces as measured by the DPT has been linked to aggression (Kimonis et al., 2006), hostile attribution (Miller & Johnston, 2019), psychopathy and antisocial traits (Kimonis et al., 2020).

Whilst the lack of an attentional bias to angry faces measured using trial-level BS in our first study could be explained by the DPT failing to effectively disambiguate bias toward threat from disengagement from threat (Koster et al., 2004), it is also possible that the previously found relationship was due to publication bias (i.e., bias toward publication of statistically significant results in social sciences, Franco et al., 2014). The majority of studies using the DPT reviewed in the present research have not reported reliabilities of traditional BS computations. Conversely, studies which report such reliabilities do so in order to highlight low reliabilities (Price et al., 2015). Although in the first study, I implemented changes suggested by Price et al. (2015) to increase the reliability of traditional BS, very low reliabilities were found. In comparison, trial-level BS showed considerably higher reliabilities (Zvielli et al., 2015); however, such alternative, more stable computations are only mentioned but not used in subsequent research (Abend et al., 2018). Further research using valid and reliable methods is needed to examine links between anxiety and antisocial traits and attentional biases.

5.8. Directions for future research

Future research should seek to replicate the present findings in laboratory-based conditions, minimizing distractions and variation in how stimuli are presented in the RT task. Further, as the experiments in the second and third studies in this thesis were designed as EEG studies, future research could easily apply EEG measurements of cognitive timelines to these study designs. This would enable the examination of the quality and accuracy of the representation held in WM, whether these are different across emotional states and emotional valence of WM templates and targets. Moreover, lateralized indicators of attention allocation (such as the PD and N2pc) could be used to examine biases to emotional targets in the presence of neutral or emotional distractors. EEG studies have shown that behavioural measures may not always capture attentional bias, particularly to emotional faces (Thigpen et al., 2018); therefore, EEG data may provide useful further insights in the timelines of processes investigated in the present research.

Affective states have been shown to have a stronger effect on encoding when compared to retrieval (Mikels & Reuter-Lorenz, 2019). Future research should use ERP indicators of strength/accuracy of encoding such as the contralateral delay activity and distractor positivity to separate the influence of induced anger during encoding, maintenance, and retrieval of WM templates. Further, previous research has shown distinct patterns of encoding with slower encoding in angry emotional states (Davies et al., 2020). Future research measuring ERP indicators of encoding and maintenance in WM to investigate whether encoding speed and accuracy is contingent upon emotional state and the extent to which encoding and WM maintenance influences the processing timeline during visual search for naturalistic faces.

5.9. Concluding comments

Developments in understanding cognitive biases associated with antisocial traits and behaviours are fundamental to intervention design. The wide range of methods currently used to assess cognitive biases makes it difficult to summarise findings and apply them to the design of tools to be used in the diagnosis and treatment of externalizing disorders. Further, mixed findings point toward a lack of rigorous research, with numerous studies either using tools which have been shown to be have low validity or failing to report the reliability of behavioural measures. Finally, the literatures studying perceptual and attribution biases appear disjointed, providing snapshots of influences of bias on behaviour although these behaviours would seldom occur in isolation in real-world settings. In this thesis, the main aim was to apply the SIP model framework to the study of cognitive biases in relation to measures of aggression and constructs related to aggression such as anger, self-reported delinquency, and CU traits. This was accomplished by examining visual attention and WM guided visual attention to emotional faces and linking measurements of perceptual bias to later biases in the SIP model, emotion, and measurements of antisocial traits and behaviours.

The experiments presented in this thesis provide several novel findings. Firstly, this research showed that the reliability of BS in the DPT can be improved by trial-level BS computations which take into account temporal dynamics of attentional bias. However, BS with higher reliability were not related to later stages in the SIP model or antisocial traits and behaviours. For the second and third experiments, I designed a novel dual-task which allowed the measurement of WM biases of visual attention for naturalistic faces. Evidence was found of WM guidance of attention, that is, preferential allocation of attention to targets matching WM contents in both offender and community samples. Measurements of WM biases of perception toward angry faces were related to hostile attribution bias, increased endorsement of aggressive responses to ambiguous social provocation, and increased aggression. Additionally, as posited by the SIP model, emotion influenced both early and late stages of social-information processing. Participants in the induced anger condition displayed enhanced performance on the visual task and narrowed the attentional focus toward anger-related information. Trait anger was associated with increased bias toward angry faces and increased emotion dysregulation was associated with deviant patterns or processing throughout steps in social-information processing. This research demonstrated links between perception and attribution biases, and construction of aggressive behavioural responses, whilst highlighting the importance of experienced emotion and regulation strategies at every step. These studies emphasized the importance of unifying the cognitive bias literature under the SIP framework, using real-world stimuli and scenarios, this perspective providing multiple directions for future intervention work.

Appendix A: Dual Task- Development and validation

The development of the dual task was informed by the biased competition model of visual selection (Desimone & Duncan, 1995) according to which neural activity is biased in favour of object features held in WM. Facilitation of visual attention by WM is typically studied by asking participants to memorise content (a WM template) overlapping with the target of attention (Downing, 2000). Guidance of attention by WM has been observed at early stages during encoding (Soto et al., 2005) as well as during and after information is consolidated in WM (Olivers et al., 2006). Moreover, the guidance was absent when cues were primed but not encoded in WM for later report (Soto et al., 2005). This effect has been replicated in studies using functional magnetic resonance imaging (Soto et al., 2012) and event-related potentials (Kumar et al., 2009). The paradigm has been applied to visual search displays containing shapes and simple objects (Soto et al., 2005, Downing, 2000), and more recently to real-world objects and scenes (Seidl-Rathkopf et al., 2015) and facial expressions (Yao et al., 2019).

In line with previous research, participants in this study were required to complete the (dual) combined WM and visual search task. Only one face was displayed during encoding of the WM template due to limited capacity for face memory (Towler et al., 2016, van Moorselaar et al., 2014). To ensure that the encoding of the WM template was directly relevant to the task (Woodman et al., 2013) the match/no-match test was displayed at the end of each trial. The search array following the encoding display was a similar design to the visual search task using naturalistic face stimuli developed by Burra et al. (2017). As WM templates and targets had the same identity (and the same emotional valence in half of the trials), additional checks were conducted to exclude the possibility of participants responding to the gender of the target based on the memorised WM template. Specifically, RTs were plotted and inspected for each participant and revealed a stable RT pattern as opposed to a learning pattern, indicative of reliance upon information stored in WM.

References

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological science*, *11*(6), 467-473.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248-261. doi: 10.1037/0096-1523.31.2.248

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243-1265.

Soto, D., Greene, C. M., Kiyonaga, A., Rosenthal, C. R., & Egner, T. (2012). A parieto-medial temporal pathway for the strategic control over working memory biases in human visual attention. *Journal of Neuroscience*, *32*(49), 17563-17571. doi: 10.1523/JNEUROSCI.2647-12.2012.

Kumar, S., Soto, D., & Humphreys, G. W. (2009). Electrophysiological evidence for attentional guidance by the contents of working memory. *European Journal of Neuroscience*, *30*(2), 307-317. doi: 10.1111/j.1460-9568.2009.06805.x

Seidl-Rathkopf, K. N., Turk-Browne, N. B., & Kastner, S. (2015). Automatic guidance of attention during real-world visual search. *Attention, Perception, & Psychophysics*, *77*(6), 1881-1895. doi: 10.3758/s13414-015-0903-8.

Yao, N., Rodriguez, M. A., He, M., & Qian, M. (2019). The influence of visual working memory representations on attentional bias to threat in individuals with high trait anxiety. *Journal of Experimental Psychopathology*, *10*(4), 1-15. doi: 10.1177/2043808719876149

Towler, J., Kelly, M., & Eimer, M. (2016). The focus of spatial attention determines the number and precision of face representations in working memory. *Cerebral Cortex*, *26*(6), 2530-2540. doi: 10.1093/cercor/bhv083

van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention?. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1450-1465. doi: 10.1037/a0036229

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. (2013). Where do we store the memory representations that guide attention?. *Journal of vision*, *13*(3), 1-17. doi: 10.1167/13.3.1.

Burra, N., Coll, S. Y., Barras, C., & Kerzel, D. (2017). Electrophysiological evidence for attentional capture by irrelevant angry facial expressions: Naturalistic faces. *Neuroscience Letters*, *637*, 44-49. doi: 10.1016/j.neulet.2016.11.055

Appendix B: Hierarchical multiple linear regression analyses with CU as IV and exploratory correlations including CU traits

Table B1

Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$)

	β	*SE*	*p*	R^2 Total/Change
Step 1				0.003
Gender	-0.08	0.10	0.41	
Education	0.06	0.10	0.53	
Age	-0.14	0.10	0.15	
Step 2				0.04/0.028*
Gender	-0.03	0.09	0.75	
Education	0.06	0.09	0.52	
Age	-0.07	0.10	0.48	
Reactive aggression and anger	0.22	0.10	0.03*	
Step 3				0.03/0.75
Gender	-0.03	0.10	0.71	
Education	0.06	0.09	0.51	
Age	-0.07	0.10	0.48	
Reactive aggression and anger	024	0.11	0.04*	
Proactive Aggression	-0.01	0.04	0.76	
Step 4				0.05/ 0.08
Gender	-0.03	0.10	0.69	
Education	0.05	0.10	0.62	

Age	-0.07	0.10	0.49
Reactive aggression and anger	0.19	0.11	0.08
Proactive Aggression	0.13	0.10	0.21
Callous-unemotional traits	-0.14	0.10	0.15

Note. β, standardized regression coefficient.

*p < .05.

Table B2

Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$)

	β	*SE*	*p*	R^2 *Total/Change*
Step 1				0.03
Gender	0.15	0.09	0.10	
Education	-0.02	0.09	0.84	
Age	0.19	0.10	0.05	
Step 2				0.06/0.05*
Gender	0.11	0.10	0.25	
Education	-0.01	0.09	0.84	
Age	0.12	0.10	0.22	
Reactive aggression and anger	-0.19	0.10	0.05*	
Step 3				0.05/0.41
Gender	0.11	0.10	0.24	
Education	-0.006	0.09	0.95	
Age	0.13	0.10	0.21	
Reactive aggression and anger	-0.16	0.11	0.14	

Proactive Aggression	-0.09	0.10	0.41	
Step 4				0.05/0.43
Gender	0.12	0.10	0.23	
Education	-0.008	0.09	0.93	
Age	0.12	0.10	0.22	
Reactive aggression and anger	-0.17	0.11	0.12	
Proactive Aggression	-0.09	0.10	0.34	
Callous-unemotional traits	0.08	0.10	0.43	

Note. β, standardized regression coefficient.

Table B3

Hierarchical multiple linear regression analysis predicting attentional bias away from neutral faces whilst holding an angry face in WM ($BS3_{NEGATIVE}$)

	β	***SE***	***p***	***R^2*** ***Total/Change***
Step 1				0.02
Gender	0.05	0.09	0.54	
Education	-0.06	0.09	0.50	
Age	0.04	0.09	0.64	
Step 2				-0.009/0.18
Gender	0.03	0.10	0.78	
Education	-0.06	0.10	0.50	
Age	0.001	0.10	0.99	
Reactive aggression and anger	-0.14	0.10	0.18	

Step 3				- 0.006/0.24
Gender	0.03	0.10	0.76	
Education	-0.05	0.10	0.63	
Age	0.004	0.10	0.97	
Reactive aggression and anger	-0.09	0.11	0.44	
Proactive Aggression	-0.12	0.11	0.24	
Step 4				-0.01/0.60
Gender	0.03	0.10	0.75	
Education	-0.05	0.10	0.62	
Age	0.002	0.10	0.98	
Reactive aggression and anger	-0.09	0.11	0.41	
Proactive Aggression	-0.13	0.11	0.22	
Callous-unemotional traits	0.05	0.10	0.60	

Note. β, standardized regression coefficient.

*$p < .05$.

Table B4

Correlation matrix for Trait Anger, Aggression, Inventory of CU traits scores, and scores on the SEIP-Q

		1	2	3	4	5	6	7	8	9	10	11	12	13
1	BS1$_{POSITIVE}$	-												
2	BS1$_{NEGATIVE}$	*-.58****	-											
3	BS3$_{NEGATIVE}$	*.78****	*-.55****	-										
4	Trait Anger	*-.25***	*.25***	*-.14*	-									
5	Reactive Aggression	*-.16*	*.25***	*-.08*	*.63****	-								
6	CU scores	*-.07*	*-.04*	*.04*	*.23**	*.35****	-							
7	Hostile Attribution	*-.21**	*.20**	*-.24***	*.27***	*.16*	*.21**	-						
8	Instrumental Attribution	*-.03*	*.11*	*.03*	*-.16*	*-.21**	*-.18*	*.08*	-					
9	Benign Attribution	*.16*	*-.12*	*.06*	*-.15*	*-.17*	*-.21**	*-.43****	*-.06*	-				
10	Negative Emotional Response	*.04*	*.07*	*.13*	*.12*	*.13*	*-.03*	*.37****	*.25***	*-.19**	-			
11	Response Enactment: Appropriate	*-.04*	*.06*	*-.09*	*.04*	*.03*	*-.13*	*.13*	*.14*	*.33****	*.06*	-		
12	Response Enactment: Direct Aggression	*-.18*	*.17*	*-.21**	*.55****	*.51****	*.31****	*.54****	*-.17*	*-.27***	*.30***	*-.01*	-	

13	Response Enactment: Relational Aggression	*-.27***	*.29***	*-.12*	*.37****	*.35****	*.18*	*.56****	*.05*	*-.36****	*.49****	*-.06*	*.63****	-

Note. BS= Bias Scores.

**p<0.05, **p<0.01, ***p<0.001.*

Correlations between BS and self-report measures: exploratory analysis

Pearson correlations measured associations between self-report measures of externalizing behaviour and subscales of the SEIP-Q so as to explore the specificity of relationships between aggression, trait anger, CU traits, and SIP steps. Table B4 (supporting information) shows that both reactive aggression and trait anger were correlated to BS. Whilst significant positive correlations were found between reactive aggression, trait anger, and CU traits, the latter were not correlated to any BS (see Table B5 for means and standard deviations for self-reported measures). Of the SEIP-Q variables, only hostile attribution and response enactment (for direct and relational aggression) were correlated to BS. Hostile attribution was significantly positively correlated to trait anger and CU traits. Although no significant relationship was found between hostile attribution and reactive aggression, a significant negative correlation was found between the latter and instrumental attribution. Finally, there was a negative correlation between CU traits and instrumental attribution.

Whilst significantly correlated with trait anger and reactive aggression, CU traits did not significantly predictor of WM modulated attentional bias to emotional faces. CU traits have previously been linked to deficits in recognizing fearful expressions (Szabó et al.. 2017), although a systematic review indicated the recognition deficit was pervasive across emotions (Dawel et al., 2012) including happiness (Hastings et al., 2008) and anger (Muñoz 2009). The lack of an increased WM modulated attentional bias in relation to CU traits could be explained by an amplified emotion recognition impairment resulting in neglect of the target's emotional valence. Given the relationship between reduced reactivity to emotional information and higher CU traits (Fanti et al., 2016), it is possible that individuals displaying higher CU traits were able to disengage from the emotional valence of the stimuli and focus on task requirements. These results suggest that higher CU traits are likely to affect later processing stages in the model. Specifically, CU traits were associated with attribution of hostile intent and increased self-reported likelihood of engaging in directly aggressive behaviour.

Table B5

Means and Standard Deviations for Aggression, CU traits, Trait Anger, Attributional and Emotional response variables

Measure	*Mean*	*Standard Deviation*
Reactive Proactive Questionnaire	11.67	6
Reactive Aggression scores	8.84	3.75
CU Traits	20.91	8.07
Trait Anger	20.55	5.19
Hostile attribution	2.1	0.4
Instrumental attribution	3.3	0.3
Negative evaluation and response	2.8	0.4
Benign attribution	2.55	0.35

Dawel A, O'Kearney R, McKone E, Palermo R. Not just fear and sadness: Meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. Neuroscience & Biobehavioral Reviews. 2012 Nov 1;36(10):2288-2304. doi: 10.1016/j.neubiorev.2012.08.006.

Fanti KA, Panayiotou G, Lombardo MV, Kyranides MN. Unemotional on all counts: evidence of reduced affective responses in individuals with high callous-unemotional traits across emotion systems and valences. Social Neuroscience. 2016 Jan 2;11(1):72-87. doi: 10.1080/17470919.2015.1034378.

Hastings ME, Tangney JP, Stuewig J. Psychopathy and identification of facial expressions of emotion. Personality and Individual Differences. 2008 May 1;44(7):1474-1483. doi: 10.1016/j.paid.2008.01.004.

Muñoz LC. Callous-unemotional traits are related to combined deficits in recognizing afraid faces and body poses. Journal of the American Academy of Child & Adolescent Psychiatry. 2009 May 1;48(5):554-562. doi: 10.1097/CHI.0b013e31819c2419.

Szabó E, Kocsel N, Edes A, Pap D, Galambos A, Zsombok T, et al. Callous-unemotional traits and neural responses to emotional faces in a community sample of young adults. Personality and Individual Differences. 2017 Jun 1;111:312-317. doi: 10.1016/j.paid.2017.02.026.

Appendix C: Exploratory Analyses- Are WM visual selection biases predicted by social-emotional information processing variables?

The third hypothesis investigated the relationship between the effect of WM on visual selection bias and later stages in social-information processing. Higher hostile attribution scores were expected to predict attentional bias toward angry faces when holding an angry face in WM compared to neutral faces when holding a neutral face in WM ($BS1_{POSITIVE}$) and slower RTs to (or attentional bias away from) neutral faces when holding an angry face in WM compared to neutral faces when holding a neutral face in WM (BS3) arrays. Hierarchical regressions were planned to evaluate the relationship between attentional bias (BS1 and BS3 as DVs) and later stages in SIP (as predictors). Age, gender, and education were added at step 1, hostile attribution scores were added at step 2, instrumental attribution and negative emotional response scores were added at step 3, and response enactment (for aggressive responses) scores were added at step 4.

Results

Response enactment for aggressive responses was a positive predictor of attentional bias toward angry faces whilst holding an angry face in mind ($BS1_{POSITIVE}$, $\beta = 0.28$, $p= 0.04$, model at step 4) and a negative predictor of attentional bias away from angry faces whilst holding an angry face in mind ($BS1_{NEGATIVE}$, $\beta = -0.27$, $p= 0.04$, model at step 4). Negative emotional response was a significant positive predictor of attentional bias away from angry faces whilst holding an angry face in mind ($BS1_{NEGATIVE}$, $\beta = 0.23$, $p= 0.04$, model at step 4) and of attentional bias away from neutral faces identified whilst holding an angry face in mind ($BS2_{NEGATIVE}$, $\beta = 0.27$, $p= 0.01$, model at step 3). Hostile attribution scores were also a significant negative predictor of attentional bias away from neutral faces identified whilst holding an angry face in mind ($BS2_{NEGATIVE}$, β= -0.34, p= 0.001, model at step 3). A summary of hierarchical regression models is shown in Tables C1-C3 below.

Table C1

Hierarchical multiple linear regression analysis predicting attentional bias toward angry faces whilst holding an angry face in WM ($BS1_{POSITIVE}$) with SEIP-Q variables as predictors

	β	*SE*	*p*	$R^2_{Total/Change}$
Step 1				*0.003*
Gender	-0.08	0.10	0.41	
Education	0.06	0.10	0.53	
Age	-0.14	0.10	0.15	
Step 2				*0.02/ 0.07*
Gender	-0.06	0.09	0.51	
Education	0.05	0.10	0.63	
Age	-0.13	0.10	0.20	
Hostile Attribution	0.17	0.09	0.08	
Step 3				*0.02/0.53*
Gender	-0.07	0.10	0.48	
Education	0.03	0.10	0.77	
Age	-0.13	0.09	0.18	
Hostile Attribution	0.17	0.10	0.10	
Instrumental Attribution	0.11	0.10	0.27	
Negative Emotional Response	-0.03	0.11	0.77	

Step 4				*0.05/0.04**
Gender	-0.03	0.10	0.75	
Education	0.05	0.10	0.58	
Age	-0.08	0.10	0.40	
Hostile Attribution	0.03	0.12	0.80	
Instrumental Attribution	0.15	0.10	0.13	
Negative Emotional Response	-0.11	0.11	0.32	
Response Enactment (aggressive responses)	0.28	0.13	0.04*	

β, standardized regression coefficient.

*$p < .05$.

Table C2

Hierarchical multiple linear regression analysis predicting attentional bias away from angry faces whilst holding an angry face in WM ($BS1_{NEGATIVE}$) with SEIP-Q variables as predictors

	β	*SE*	*p*	$R^2_{Total/Change}$
Step 1				*0.03*
Gender	0.15	0.09	0.10	
Education	-0.02	0.09	0.84	
Age	0.19	0.10	0.05	
Step 2				*0.05/0.07*
Gender	0.14	0.09	0.14	
Education	-0.004	0.09	0.96	
Age	0.17	0.09	0.08	
Hostile Attribution	-0.17	0.09	0.07	

Step 3				*0.05/0.32*
Gender	0.11	0.10	0.26	
Education	0.03	0.10	0.79	
Age	0.18	0.09	0.06	
Hostile Attribution	-0.23	0.10	0.03	
Instrumental Attribution	-0.07	0.10	0.50	
Negative Emotional Response	0.16	0.11	0.15	
Step 4				*0.08/0.04**
Gender	0.07	0.10	0.46	
Education	0.001	0.10	0.99	
Age	0.14	0.10	0.15	
Hostile Attribution	-0.09	0.12	0.44	
Instrumental Attribution	-0.11	0.10	0.27	
Negative Emotional Response	0.23	0.11	0.04*	
Response Enactment (aggressive responses)	-0.27	0.13	0.04*	

β, standardized regression coefficient.

*$p < .05$.

Table C3

Hierarchical multiple linear regression analysis predicting attentional bias away from neutral faces whilst holding an angry face in WM ($BS2_{NEGATIVE}$) with SEIP-Q variables as predictors

	β	***SE***	***p***	***R^2 Total/Change***
Step 1				*-0.01*

Gender	0.06	0.10	0.55	
Education	-0.07	0.10	0.50	
Age	0.05	0.10	0.64	
Step 2				*0.03/0.01**
Gender	0.03	0.09	0.70	
Education	-0.04	0.09	0.63	
Age	0.02	0.09	0.80	
Hostile Attribution	-0.23	0.09	0.02*	
Step 3				*0.07/0.04**
Gender	-0.03	0.10	0.82	
Education	-0.01	0.10	0.94	
Age	0.04	0.09	0.64	
Hostile Attribution	-0.34	0.12	0.001**	
Instrumental Attribution	-0.01	0.10	0.95	
Negative Emotional Response	0.27	0.11	0.01*	
Step 4				*0.08/0.14*
Gender	-0.06	0.10	0.57	
Education	-0.02	0.10	0.79	
Age	0.01	0.10	0.90	
Hostile Attribution	-0.24	0.12	0.05	
Instrumental Attribution	-0.04	0.10	0.72	
Negative Emotional Response	0.33	0.11	0.005**	
Response Enactment (aggressive response)	-0.19	0.13	0.14	

Note. β, standardized regression coefficient.

*$p < .05$, **$p<0.01$

Appendix D: Anger manipulation check

Emotional state is measured by asking the following 4 questions:

How angry are you feeling?

How frustrated are you feeling?

How irritated are you feeling?

How annoyed are you feeling?

The scores for each (measured on a scale from 0-Not at all to 8-Extremely) are averaged to produce an emotional score before emotion induction and an emotional score after completion of the visual task.

I have provided histograms and descriptives for each condition/question below and results to comparisons between condition and across times (before and after).

1. Visualisation

1.1 Emotional state before induction (all participants)

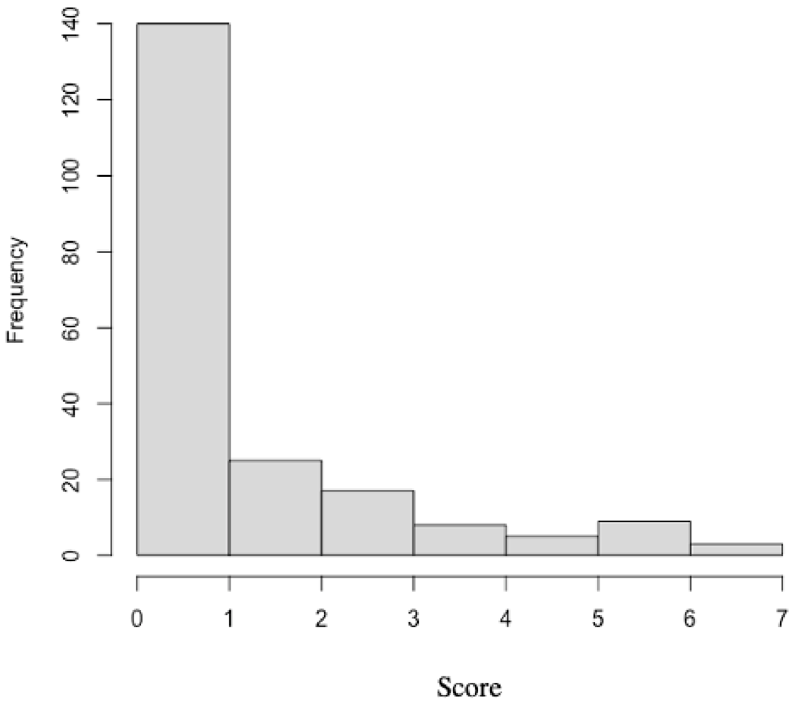
Median= 0.5

Mean= 1.17

SD= 1.64

Figure D1

Emotional state before induction (all participants)



1.2 Emotional state after induction (all participants)

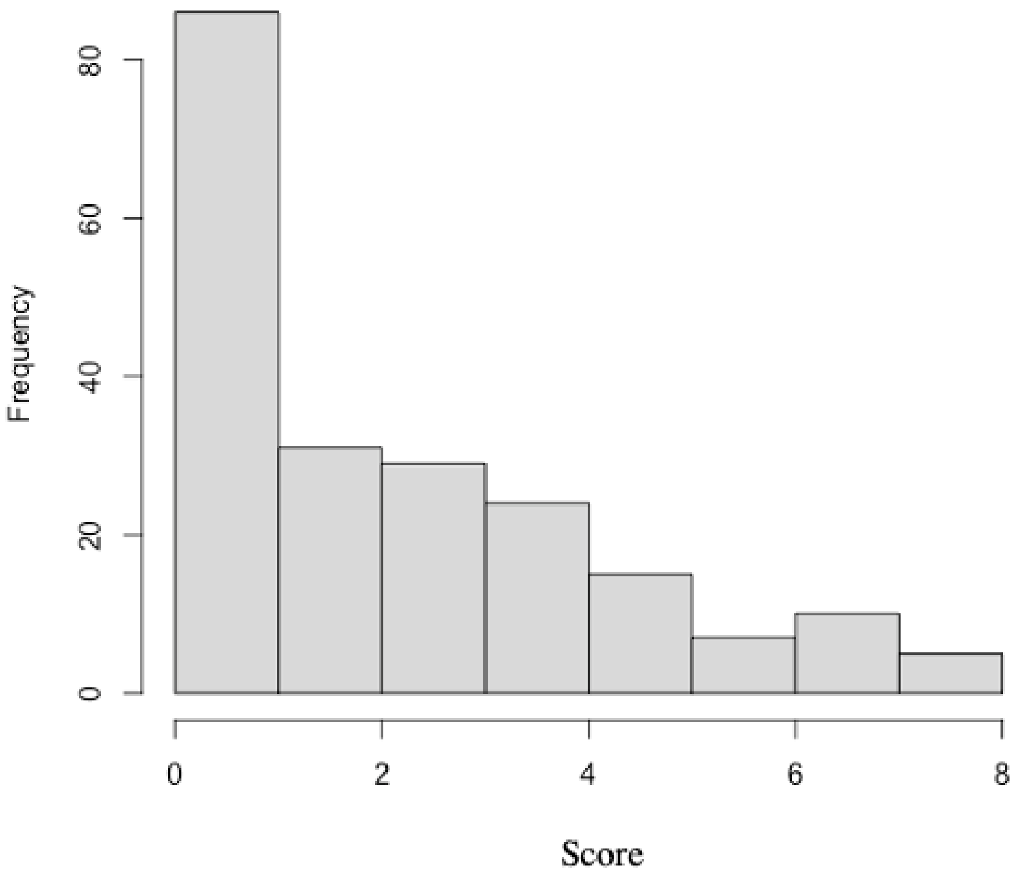
Median= 1.75

Mean= 2.19

SD= 2.07

Figure D2

Emotional state after induction (all participants)



1.3 Emotional state in NEUTRAL group- before induction

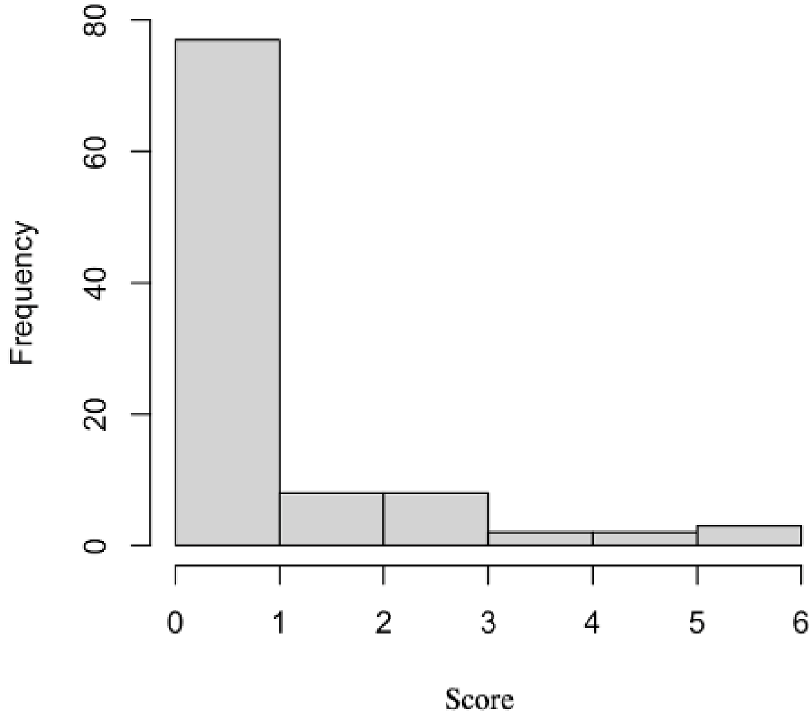
Median= 0

Mean= 0.85

SD= 1.4

Figure D3

Emotional state before induction: scores in the neutral emotion induction condition



1.4 Emotional state in ANGRY group- before induction

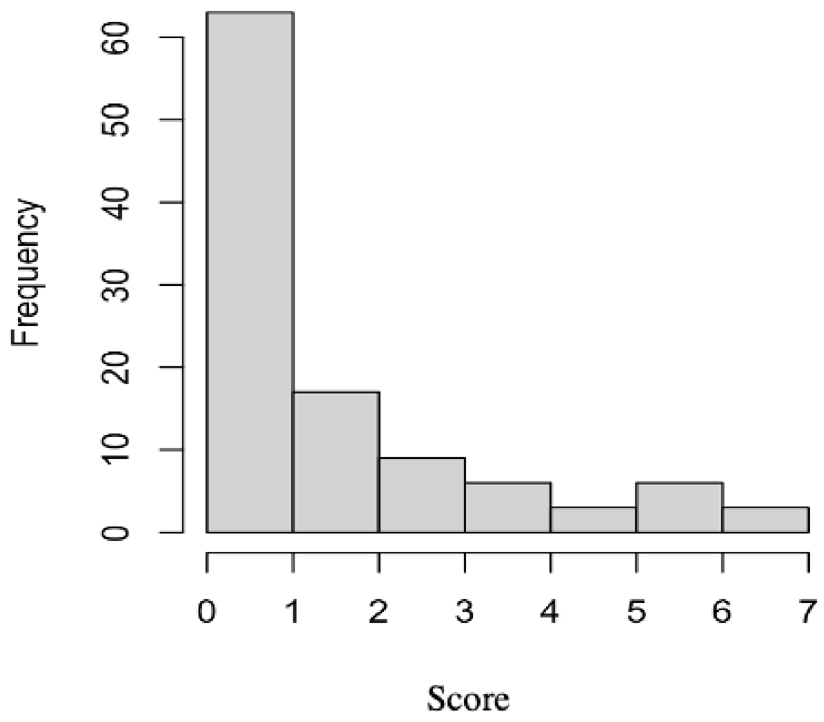
Median= 1

Mean= 1.48

SD= 1.79

Figure D4

Emotional state before induction: scores in the anger induction condition



1.5 Emotional state in NEUTRAL group – after induction

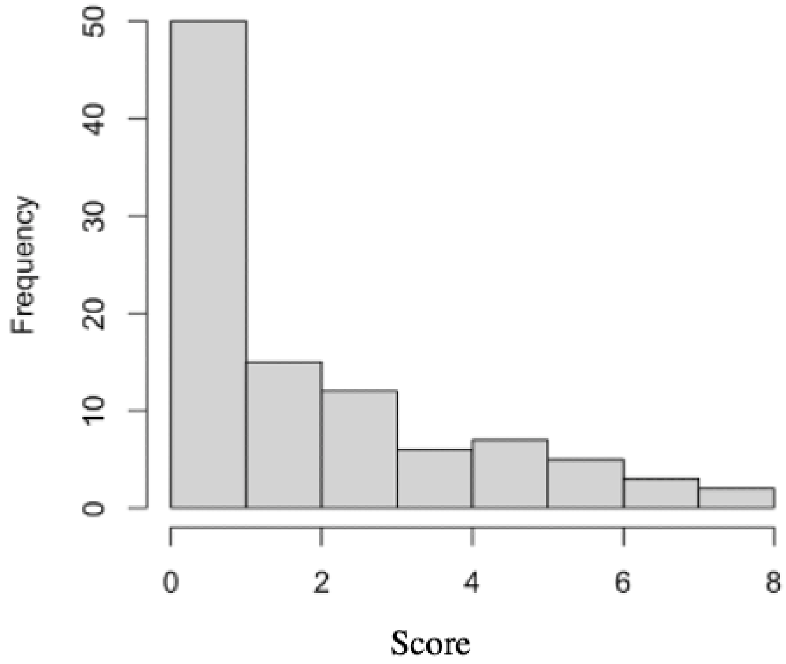
Median= 1.12

Mean= 1.9

SD= 2.03

Figure D5

Emotional state after induction: scores in the neutral emotion induction condition



1.6 Emotional state in ANGRY group – after induction

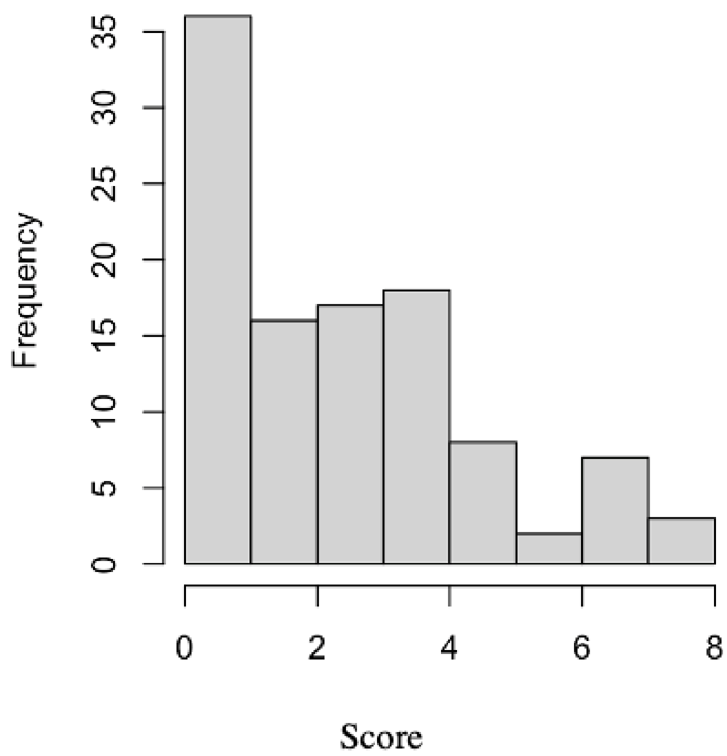
Median= 2.25

Mean= 2.48

SD= 2.08

Figure D6

Emotional state after induction: scores in the anger induction condition



2. Scores comparison between groups just for the anger subquestion (question 1 of the 4).

2.1 Anger scores before anger induction (anger induction condition)

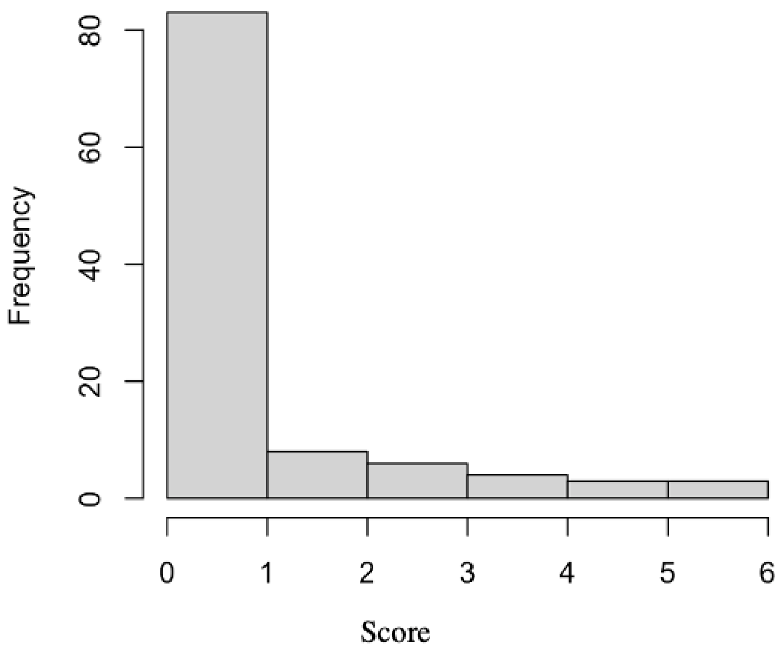
Median= 0

Mean= 0.97

SD= 1.54

Figure D7

Anger sub-scores before anger induction: anger induction condition



2.2 Anger sub-scores after anger induction (anger induction condition)

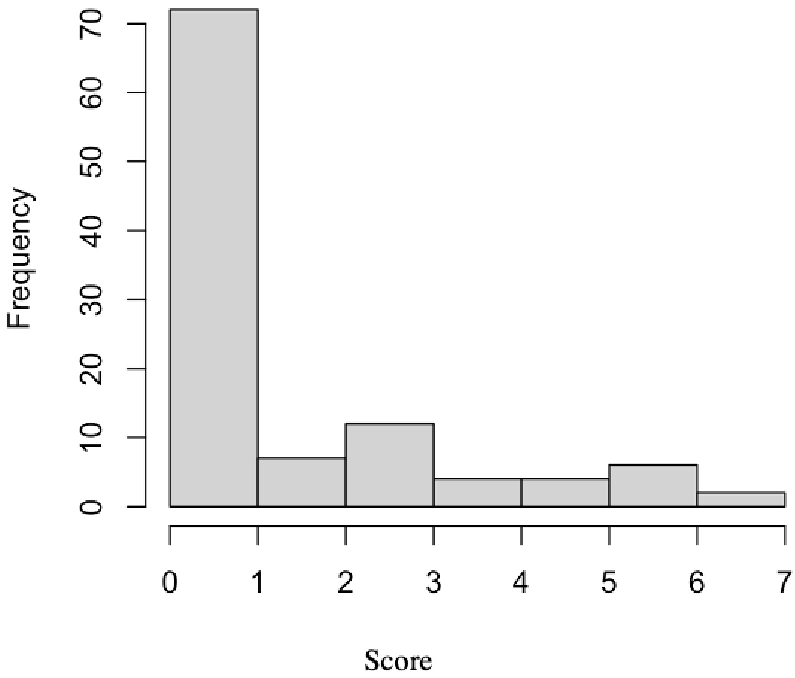
Median= 1

Mean= 1.51

SD= 1.93

Figure D8

Anger sub-scores after anger induction: anger induction condition



2.3 Anger sub-scores before emotion induction (neutral state induction condition)

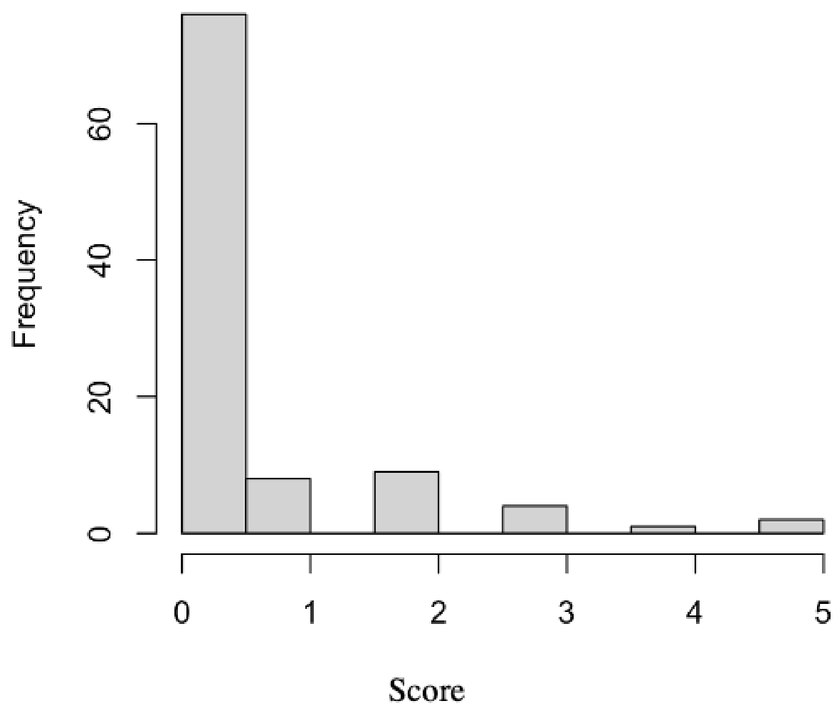
Median= 0

Mean= 0.52

SD= 1.1

Figure D9

Anger sub-scores before emotion induction: neutral state induction condition



2.4 Anger scores after emotion induction (neutral state induction condition)

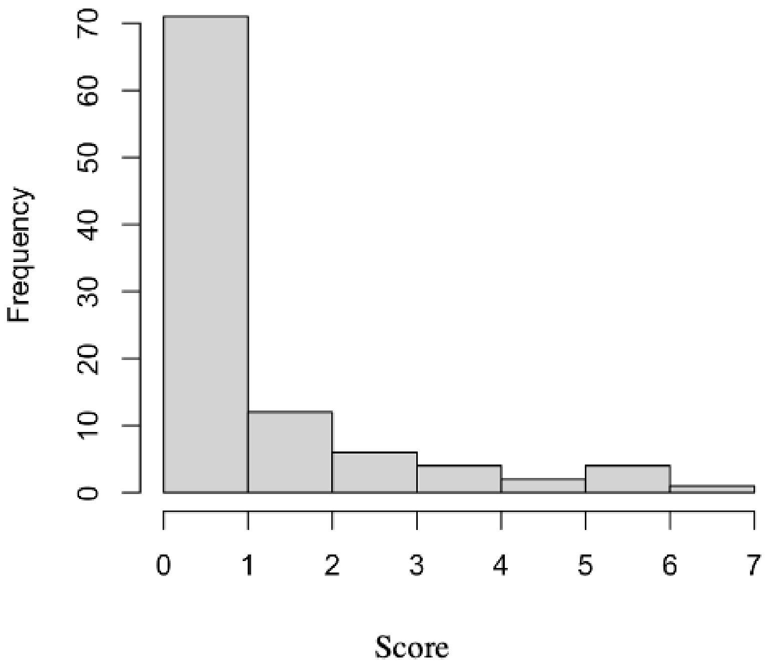
Median= 0

Mean= 1.05

SD= 1.75

Figure D10

Anger sub-scores after emotion induction: neutral state induction condition



<u>Group Comparisons</u>

Table D1

Wilcoxon signed rank tests: Comparison of emotion induction conditions

Groups	W	p	Effect size (r) 0.1= small 0.3 = moderate 0.5 = large
Emotion state before and after (paired-all participants)	14399	$p< 0.0001$	1.30
Emotional state before and after (paired-induced anger group)	3384	$p< 0.001$	0.54
Emotional state before and after (paired-induced neutral emotion group)	2891	$p<0.001$	0.65
Emotional state after induction (between induced anger and induced neutral groups)	4370	$p= 0.02$	0.16

Table D2

Mean and median scores for induced emotional conditions: composite scores, before and after emotion induction

Group	Composite anger before	Composite anger after
Induced Neutral	Mean= 0.85, Median= 0	Mean= 1.9, Median= 1.12
Induced Anger	Mean= 1.48, Median= 1	Mean= 2.48, Median= 2.25

Table D3

Mean and median scores for induced emotional conditions: anger sub-scores, before and after emotion induction

Group	Anger subscale before	Anger subscale after
Induced Neutral	Mean= 0.52, Median= 0	Mean= 1.05, Median= 0
Induced Anger	Mean= 0.97, Median= 0	Mean= 1.51, Median= 1

Table D4

Comparing trait anger scores across emotion induction conditions

Group	Trait anger (M/SD)
Induced Neutral	20.06/5.2
Induced Anger	20.64/5.28

<u>Paired plots</u>

Figure D11

Emotional state before-after: Induced anger group

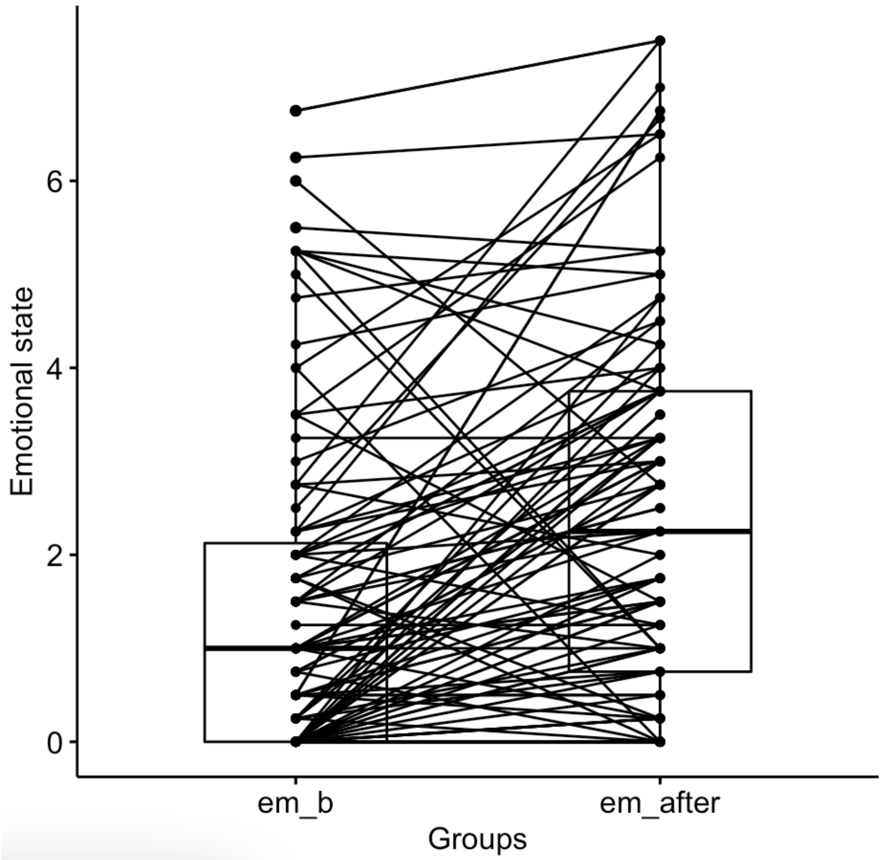
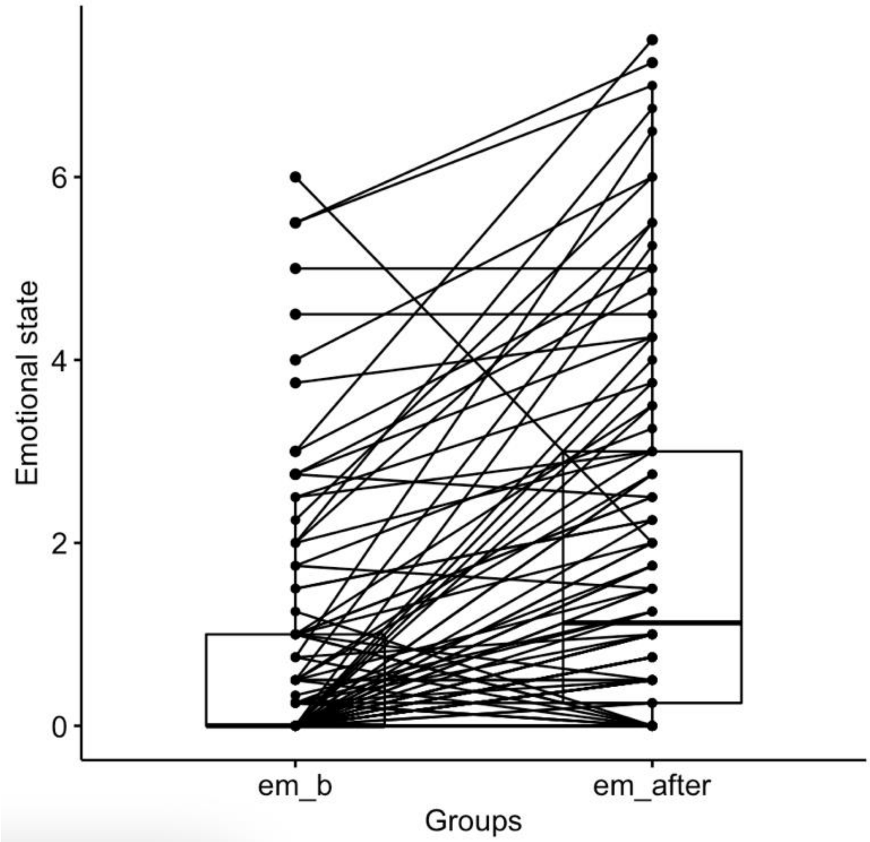


Figure D12

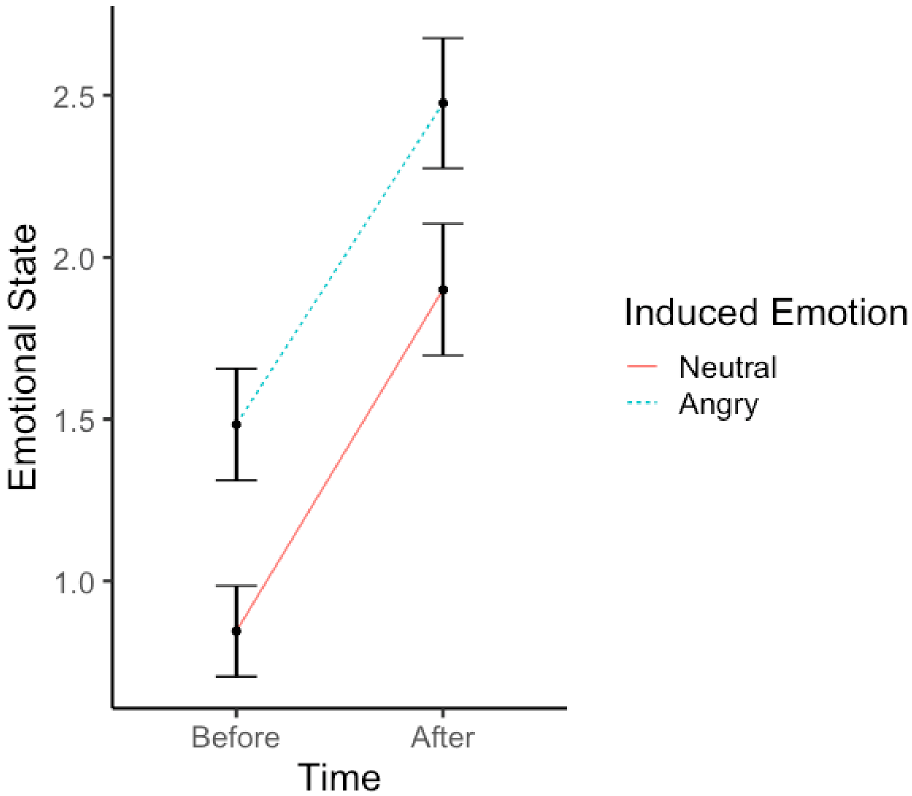
Emotional state before-after: Induced neutral emotion group



Finally, I have conducted an ANOVA with emotional state score as DV and induced emotion (anger vs neutral, between) and time (before and after, within). The outcome variable was significantly positively skewed. There was no interaction effect, only a main effect of induced emotion ($p<0.001$).

Figure D13

Interaction between induced emotion condition and time (before vs after emotion induction). Error bars show one standard error above and below mean.



<u>Exploratory Analyses:</u>

There was a significant difference in baseline state anger with participants allocated to the induced anger condition displaying higher state anger scores at baseline compared to participants allocated to the neutral condition. This difference in anger was found prior to emotion induction. Following the emotion induction procedure, participants in the induced anger group displayed significantly higher anger compared to the neutral group. However, there was an increase in anger in both groups.

Exploratory analyses were conducted to verify whether participants in the induced anger group showed increased proneness to anger. There was no evidence of difference in trait anger between emotion induction groups.

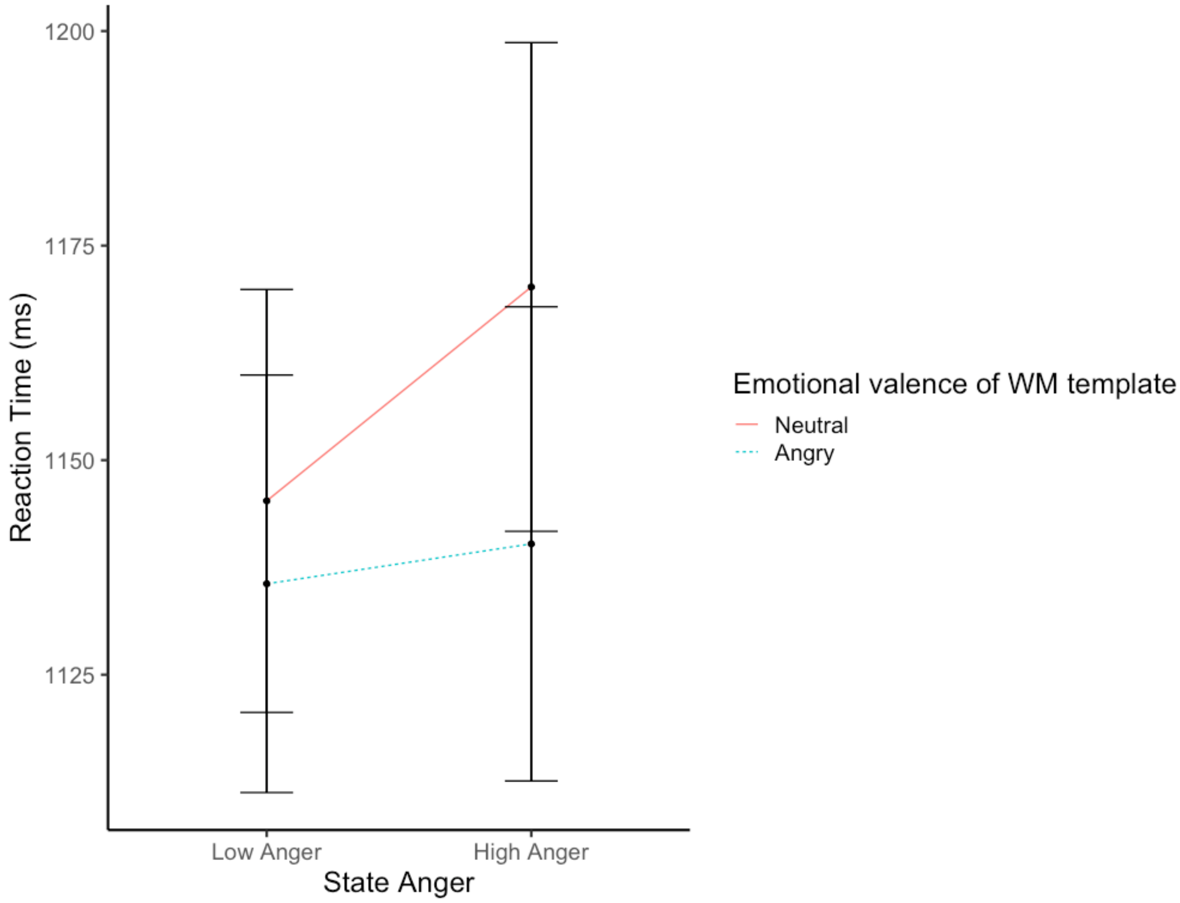
The effect of emotional state on attention: median split of emotional state scores prior to emotion induction

As can be seen in this Appendix, in the summary table for emotional state (composite scores),

both groups displayed an increase in anger following the emotion induction procedure. Participants were divided into groups using a median split of the emotional anger state measure prior to emotion induction (Median = 0.5). A 3-way mixed ANOVA with state anger group (between participants), emotional valence of target, and emotional valence of WM template (within-participants) as IVs and RTs as DV showed a main effect of emotional valence of the WM template, with faster/slower RTs to angry compared to neutral WM templates. There was also a significant interaction between emotional valence of targets and WM templates. The main effect of emotional state group was not significant and no other interactions were significant although the interaction between state anger group and emotional valence of the WM template was marginally non-significant ($F(1, 204)= 3.42$, $p= 0.07$). Figure D14 shows the interaction between high and low state anger groups and emotional valence of the WM template. Participants in the high anger group displayed considerably longer RTs to targets when holding neutral faces in WM.

Figure D14

Interaction between state anger group and emotional valence of the WM template. Error bars show one standard error above and below mean.



The effect of emotional state on attention: median split of emotional state scores following emotion induction and completion of visual task

As emotional state was measured prior to emotion induction, it is possible that the lack of an effect is due to the autobiographical recall task (i.e., the emotion induction procedure) introducing variability in emotional state post-measurement. To clarify whether the absence of an effect on visual processing was due to the autobiographical recall task, participants were divided into low and high state anger groups using a median split of state anger scores as measured after the visual task was completed (Med= 1.75). A 3-way ANOVA with state anger group (post-visual task; between-participants), emotional valence of target and emotional valence of WM template (within-participants) as IVs and RTs as DV showed a main effect of emotional valence of the WM template and an interaction effect between the emotional valence of the target in the visual search display and that of the WM template. No other interactions were significant, although the 3-way interaction between state anger group, emotional valence of the target and of the WM template was marginally non-significant ($F(1, 204)= 3.42$, $p= 0.07$). Figures below show interactions between emotional valences of WM template and target across low and high state anger group as measured following the completion of the visual task.

Figure D15

Interaction between emotional valences of WM and target in the low state anger (post visual task) group. Error bars show one standard error above and below mean.

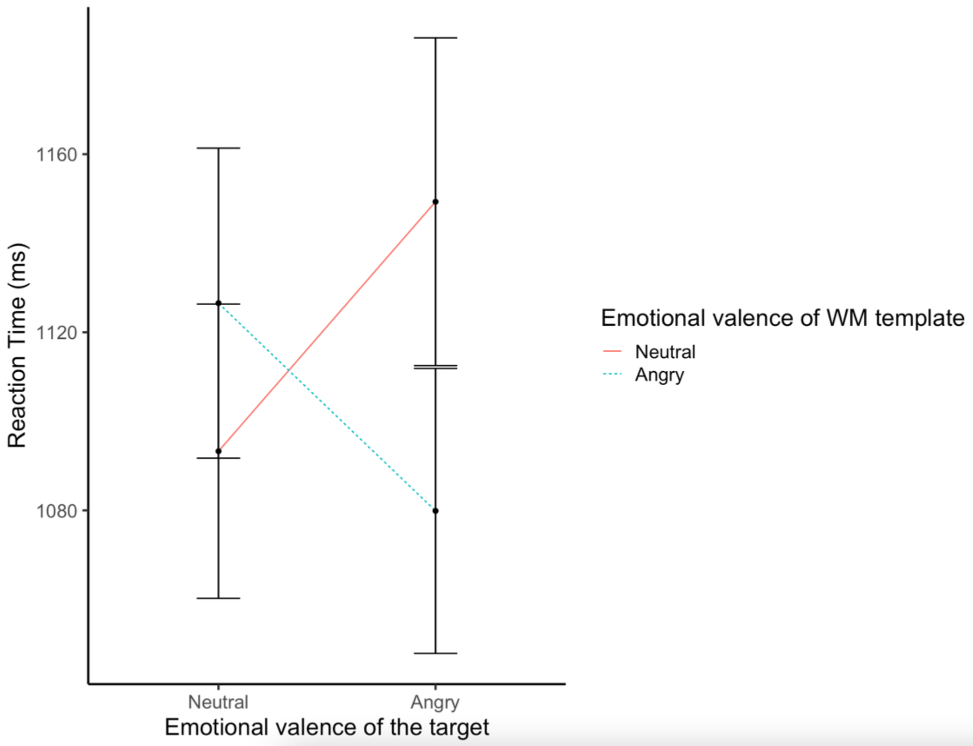


Figure D16

Interaction between emotional valences of WM and target in the high state anger (post visual task) group. Error bars show one standard error above and below mean

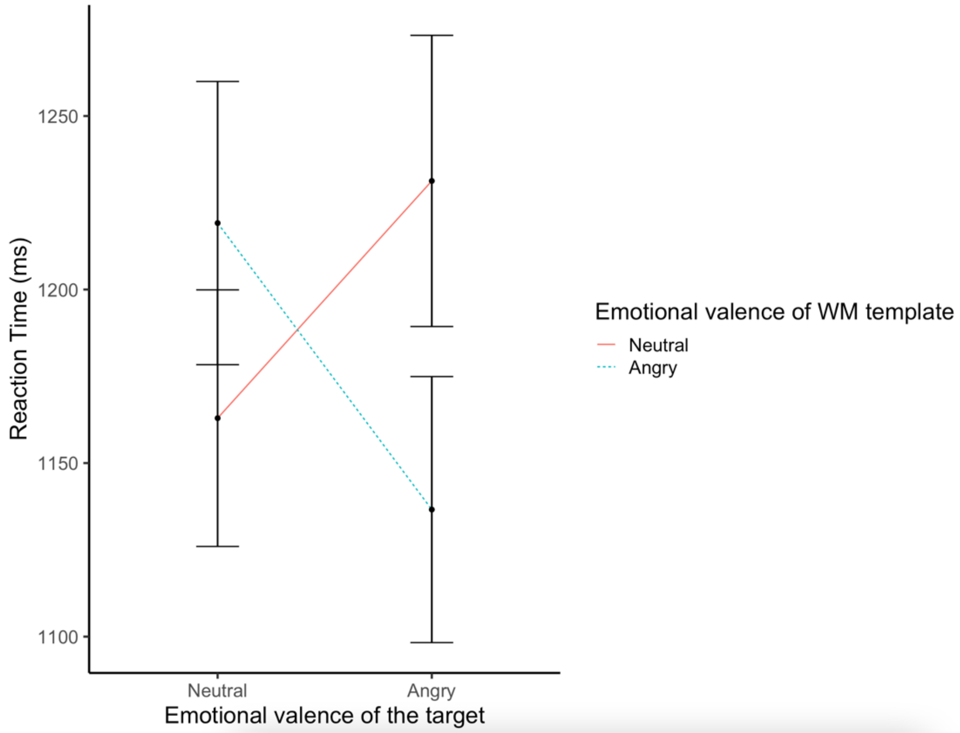


Table D5

Means and standard deviations for RTs in the high state anger (post visual task) group

	Angry WM template (Mean/SD)	Neutral WM template (Mean/SD)
Angry Target	1136.6/ 373.54	1231.29/408.85
Neutral Target	1219.14/397.69	1162.94/360.28

Table D6

Means and standard deviations for RTs in the low state anger (post visual task) group

	Angry WM template (Mean/SD)	Neutral WM template (Mean/SD)
Angry Target	1079.89/337.11	1149.32/387.75
Neutral Target	1126.55/366.78	1093.29/348.13

Whilst the emotion induction procedure resulted in an increase in self-reported state anger in both induced anger and induced neutral state groups, interactions were found between induced emotional state and emotional valences of WM template and target. Applying median splits to emotional state scores before and after emotion induction showed differences in patterns of attention allocation for each split although the interactions were not statistically significant. This suggested that the emotion induction procedure had an effect on emotional state, which was supported by the results of the main analyses. As such, analyses using a median split of emotional scores prior to emotion induction are unlikely to reflect meaningful relationships between emotional state and WM guidance of attention. The marginally non-significant 3-way interaction effect found using a median split of emotional state scores measured following completion of the visual task showed similar patterns of attention allocation to those in the main analyses. Specifically, in both low and high state anger groups participants displayed the fastest RTs to angry targets identified whilst holding angry faces in WM. There were two notable differences in results using a median split compared to induced anger condition. In the former, participants in the lower anger state displayed overall faster RTs to targets although this main effect was not significant. Comparatively, in the main analyses results showed a significant main effect of induced emotion, with participants in the induced anger condition displaying faster RTs to targets compared to participants in the induced neutral condition. Additionally, in the induced anger condition, participants showed similar mean RTs for neutral targets regardless of the emotional valence of the WM, whereas this effect was not seen in the higher state anger group.

A potential explanation for these differences in findings could be that state anger is dynamic, which means that analyses using measurement of state anger following completion of the visual task may not accurately depict relationships between emotional state and visual processing. Indeed, studies have found that completing a demanding cognitive task and appraisal of one's emotional states resulted in a decrease in the intensity of emotions experienced (Van Dillen & Koole, 2007). This suggests that high and low state anger groups as divided using a median split post visual task may have been completely different had measurements been taken during the task or immediately before. A t-test showed that participants in the high state anger group had significantly higher emotion dysregulation scores than participants in the low state anger group ($p < 0.001$), suggesting that this difference in emotional state may be partly due to emotion regulation abilities.

A secondary explanation focuses on the findings of the main analyses. Specifically, it is possible that the results observed following division of participants into groups according to their emotion induction category were partly due to emotional state as we have found significantly higher self-reported anger in the induced anger group. We did not find differences in emotion dysregulation or trait anger between the induced emotion conditions. This suggested that participants in the two conditions did not differ in their proneness toward anger or their ability to regulate their emotional states. As there was a significant increase in anger for both groups, it may be argued that the anger induction procedure had the same effect on all participants. However, had this been the case, the effect of anger of visual processing we found by splitting participants in groups based on emotion induction condition would have been enhanced in analyses using the post-visual task emotional state scores. This was not found in the data. A more likely explanation is that the emotional state score did not accurately capture the dimensions of emotion induced using the autobiographical recall task. Recall of intensely negative situations and associated emotional states may have resulted in an activation of aggressive and anger-related internal representations. That is, it is possible that by asking participants to recall an intense negative experience, this activated cognitive schemas related to anger-provoking experiences which influenced subsequent processing. There is considerable evidence linking maladaptive internal representations to aggressive behaviour and violence (Dunne et al., 2018). This study may provide initial evidence of the effect of schema activation and anger-rumination on cognition immediately preceding behaviour.

Additional Exploratory Analysis following review of responses provided in the Autobiographical Recall Task

Given that an increase in anger scores was observed for participants in both anger and neutral emotion induction conditions, I examined the responses provided to the Autobiographical Recall Task (participants were asked to recall a time when they felt little emotion (neutral condition) or a time when their (ex)partner or best friend acted unfairly toward them, resulting in an angry emotion (anger condition)). A proportion of participants in the neutral induction condition (N= 20) appeared to misunderstand the assignment and recalled feelings of anger and conflict or arguments (in which participants were directly involved not merely witnessing). These participants were included in the anger induction condition. Conversely, 3 participants in the anger induction condition stated they had never been in a situation where their (ex)partner or best friend acted unfairly toward them, resulting in an angry emotion. These participants were included in the neutral emotion induction condition.

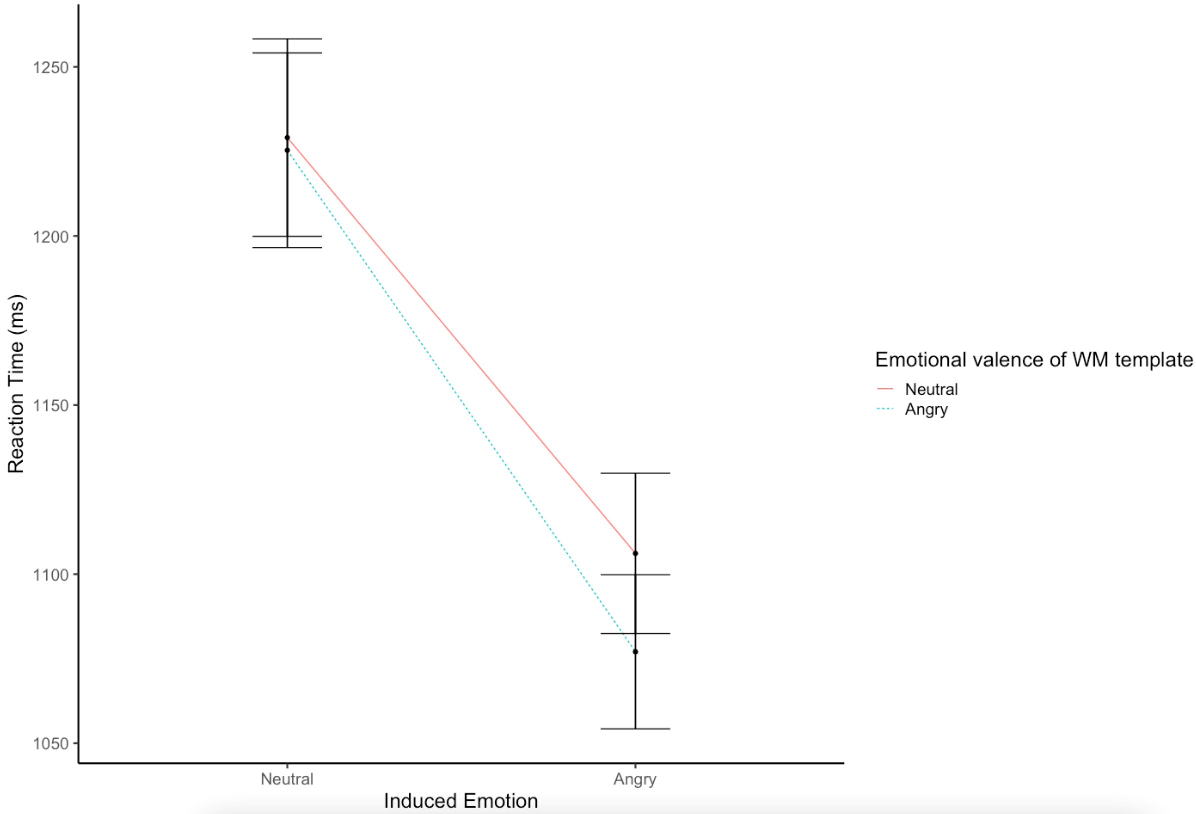
A 3-way mixed ANOVA was performed to evaluate the effects of induced emotional state (with groups adjusted as outlined above), emotional valence of the WM template, and emotional valence of the target in the visual search display on RTs to targets in the visual search array. As hypothesized, a main effect of induced emotion condition was found ($F(1, 204)= 7.02$, $p< 0.01$). Slower RTs were expected in the induced anger compared to neutral condition. However, faster RTs were found in the angry ($M= 1091.59$, $SD= 363.17$) compared to neutral ($M= 1227.22$, $SD= 375.17$) induced emotion condition.

A main effect of the emotional valence of the WM template was found ($F(1, 204)= 11.95$, $p< 0.001$) with faster RTs to targets identified whilst holding an angry ($M= 1137.52$, $SD= 369.79$) compared to neutral ($M= 1156.27$, $SD= 378.14$) template in WM. The main effect of the emotional valence of the target in the visual search display ($F(1, 204)= 0.02$) and the interaction between the former and induced emotion ($F(1, 204)= 1.03$) were non-significant.

A two-way interaction was found between emotional valence of the WM template and induced emotion condition ($F(1, 204)= 5.27$, $p< 0.05$). The Bonferroni adjusted p-value (significance being accepted at $p<0.025$), the simple main effect of emotional valence of the WM template was significant in the anger ($p< 0.001$) but not in the neutral ($p= 0.76$) emotion induction condition. In the induced anger condition, RTs were faster for targets identified whilst holding an angry ($M= 1077.06$, $SD= 355.97$) compared to a neutral ($M= 1106.12$, $SD= 370.39$) face in WM.

Figure D17

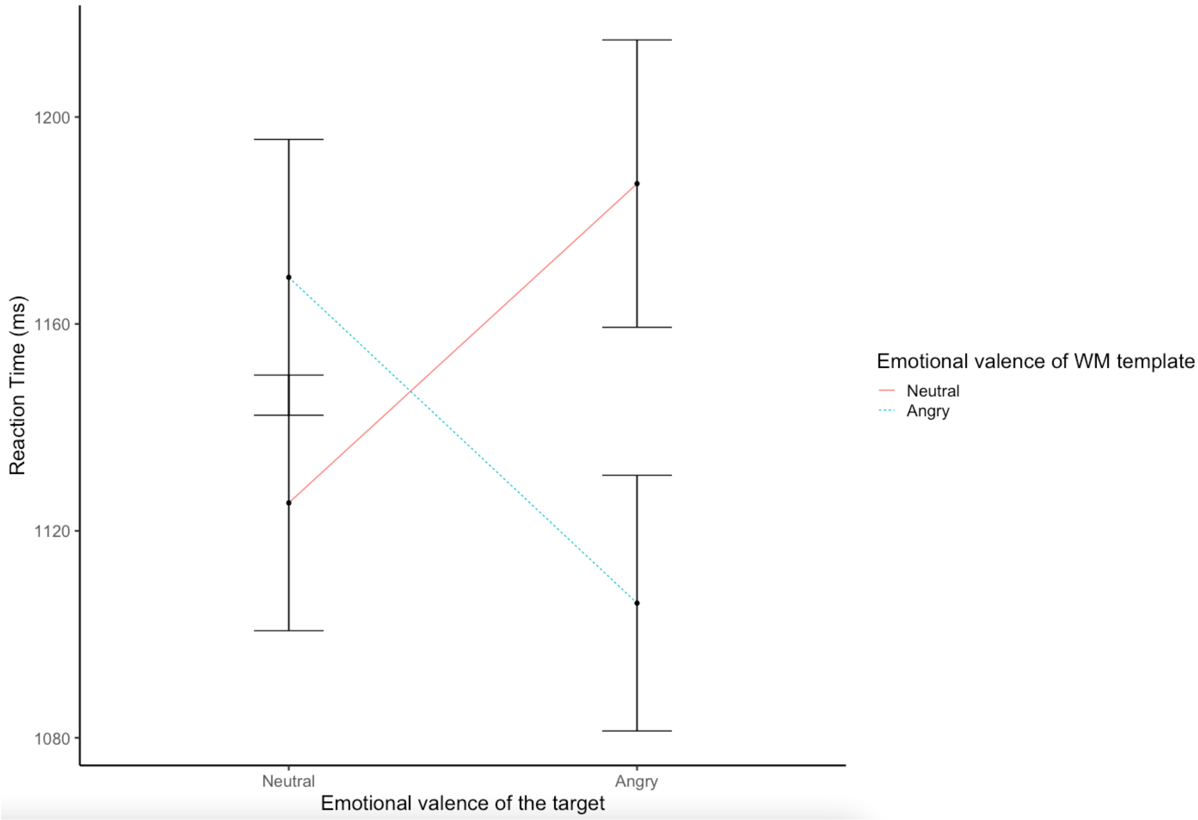
Interaction between emotional valences of WM template and emotion induction condition. Error bars show one standard error above and below mean



As expected, a main effect of WM guidance of visual attention was found; that is, a two-way interaction was found between the emotional valence of the WM template and that of the target in the visual search display ($F(1, 204) = 94.41$, $p < 0.001$). Considering the Bonferroni adjusted p-value, the simple main effect of emotional valence of the target was significant for both angry ($p < 0.001$) and neutral ($p < 0.001$) WM template conditions.

Figure D18

Interaction between emotional valences of WM template and target. Error bars show one standard error above and below mean



There was a statistically significant three-way interaction between induced emotion condition, emotional valence of the target, and emotional valence of the WM template ($F(1, 204)= 7.37$, $p< 0.01$). For the simple two-way interactions and simple main effects, a Bonferroni adjustment was applied leading to statistical significance being accepted at the $p < 0.025$ level. There were two statistically significant two-way interactions between emotional valence of the WM template and that of the target for both neutral ($F(1, 83)= 50.4$, $p< 0.001$) and angry ($F(1, 121)= 44$, $p< 0.001$) emotion induction conditions.

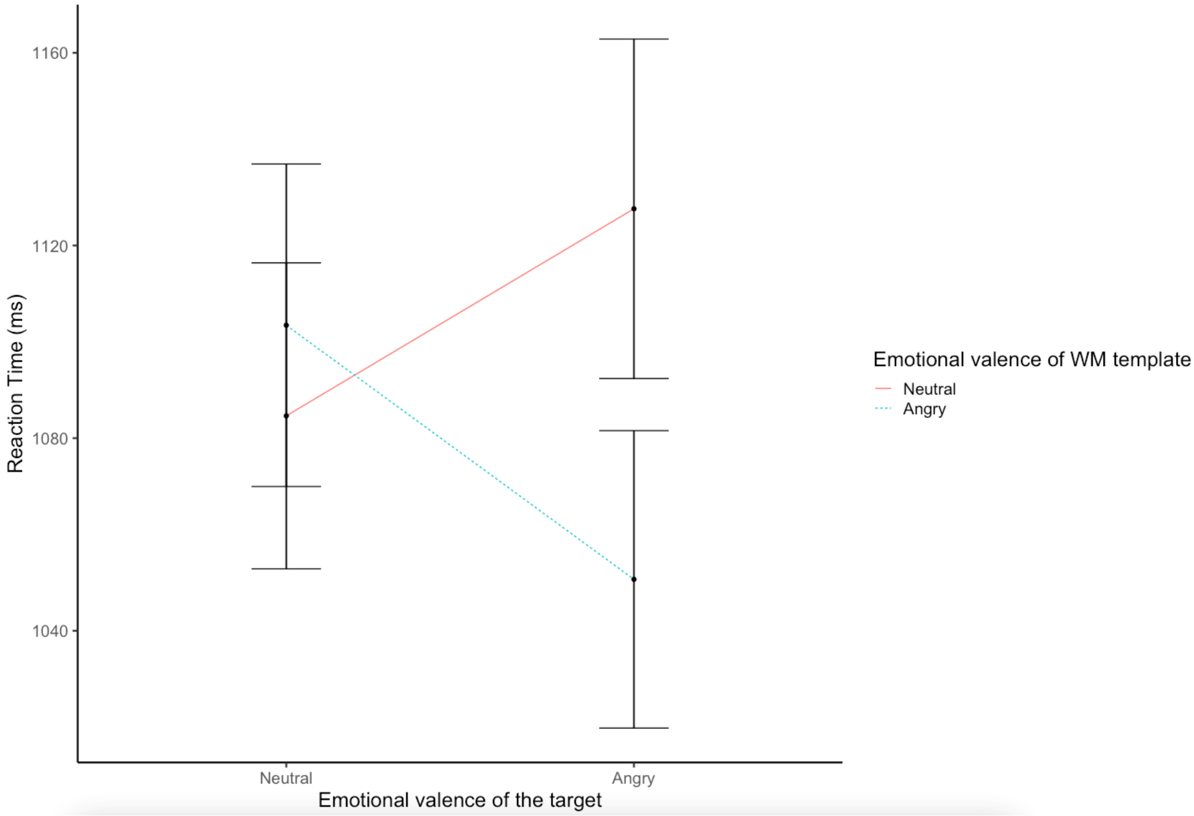
Simple main effects were conducted next. Statistical significance of a simple main effect was accepted at a Bonferroni-adjusted alpha level of 0.0125. The simple main effects of the emotional valence of the target on RTs to the visual search display were statistically significant in the **induced anger condition** for both neutral: $F(1, 121)= 17.7$, $p< 0.0001$ and angry: $F(1, 121)= 36.3$, $p< 0.0001$ WM templates. Similarly, simple main effects of the emotional valence of the target on RTs to the visual search display were statistically significant in the **induced neutral condition** for both neutral: $F(1, 83)= 38.3$, $p< 0.0001$ and angry: $F(1, 83)= 28.4$, $p< 0.0001$ WM templates.

Pairwise comparisons were conducted for each emotion induction condition. First, pairwise comparisons were run in the **induced anger condition** between RTs to the two types of emotional valence for targets in the visual search display across emotional valences for WM templates (see figure below). A Bonferroni adjustment was applied. When preceded by a neutral WM template, mean RTs were significantly higher ($p< 0.0001$) for angry targets ($M=$ 1127.62, $SD=$ 389.21) compared to neutral targets ($M=$ 1084.62, $SD=$ 350.85). When preceded by an angry WM template, mean RTs were significantly lower ($p< 0.0001$) to angry targets ($M=$ 1050.68, $SD=$ 341.01) compared to neutral targets ($M=$ 1103.44, $SD=$ 369.84).

Figure D19

Interaction between emotional valences of WM and target in the induced anger condition.

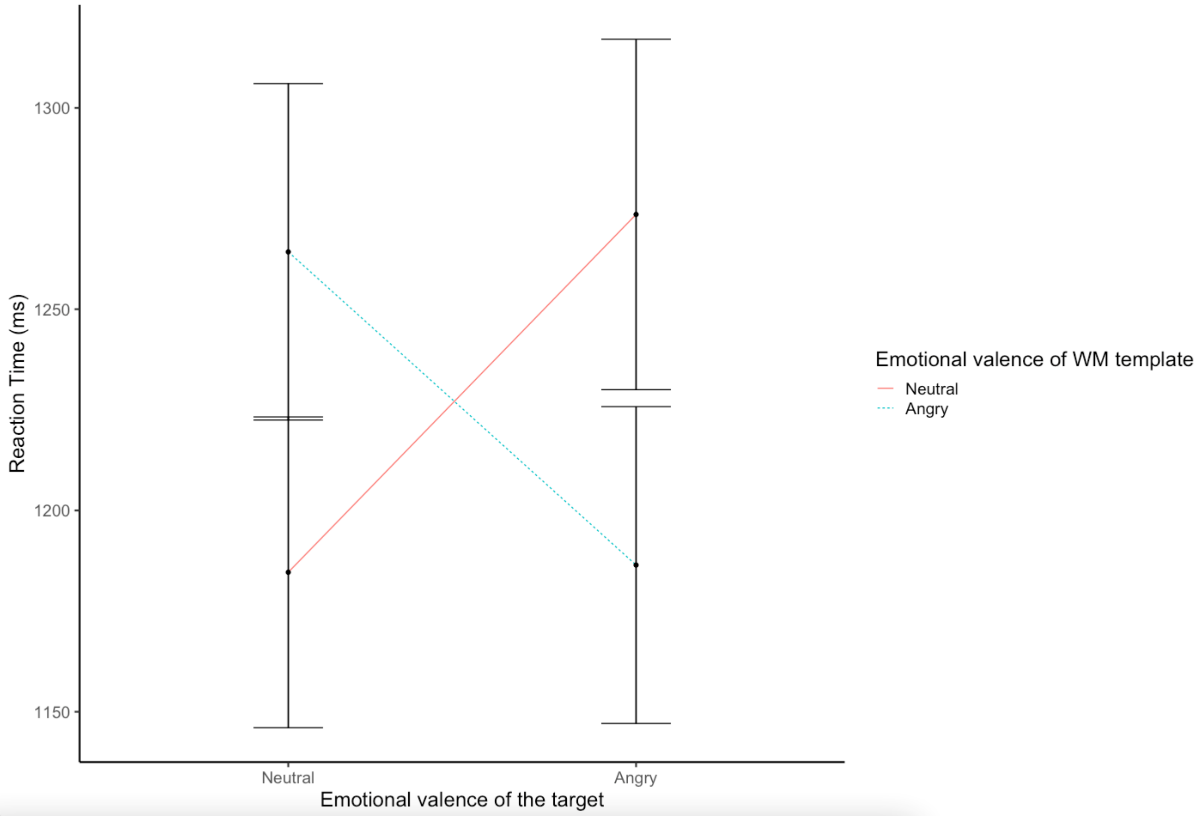
Error bars show one standard error above and below mean



Next, pairwise comparisons were run in the **induced neutral condition** between RTs to the two types of emotional valence for targets in the visual search display across emotional valences for WM templates (see Fig. 6). A Bonferroni adjustment was applied. When preceded by a neutral WM template, mean RTs were significantly higher ($p < 0.0001$) to angry targets (M= 1273.53, SD= 398.9) compared to neutral targets (M= 1184.65, SD= 353.79). When preceded by an angry WM template, mean RTs were significantly lower ($p < 0.0001$) to angry targets (M= 1186.45, SD= 360.62) compared to neutral targets (M= 1264.24, SD= 383). The main effect of the emotional valence of the target in the visual search display and the interaction between the former and induced emotion were non-significant.

Figure D20

Interaction between emotional valences of WM and target in the induced neutral condition. Error bars show one standard error above and below mean.



A paired t-test was conducted to demonstrate the expected difference in RTs to congruent compared to incongruent trials. Faster RTs were expected to congruent (emotional valence of the WM matching that of the target) compared to incongruent (emotional valence of the WM different to that of the target) trials. As expected, the t-test was significant ($t = 10.72$, $p<0.001$, one-tailed) with faster RTs to congruent ($M= 1115.73$, $SD= 354.33$) compared to incongruent ($M= 1178.06$, $SD= 390.4$) trials.

Exploratory analyses: Scores of difference

A score of difference was computed by subtracting state anger scores prior to anger induction from state anger scores following anger induction and completion of the visual task. This variable was normally distributed so participants were divided into two conditions using a median split (*Median*= 0.75, *SD*= 1.71). This resulted in 54.11% (N=112) of participants being allocated to induced anger (showed highest increase in anger following emotion induction procedure) and 45.89% (N= 95) participants being allocated to induced neutral emotional state (displayed lowest increase in anger).

A 2x2x2 mixed ANOVA was conducted with latency as DV and emotional valence of memory template and target in the visual display as IVs (within-participants) and emotional state anger as IV (between participants, median split as above). Results showed that there were no significant interactions between high/low state anger groups and emotional valence of either WM template ($F(1, 204)= 1.25$) or target ($F(1, 204)= 0.01$). There was a marginally non-significant 3-way interaction ($F(1, 204)= 2.89$, $p= 0.09$). Significant main effects were found for high/low state anger conditions ($F(1, 204)= 3.90$, $p< 0.05$) and emotional value of the WM template ($F(1, 204)= 11.57$, $p< 0.001$). A significant 2-way interaction was found between emotional valences of the WM template and that of the targets in the visual search display ($F(1, 204)= 92.33$, $p< 0.001$).

References

Abend R, Pine DS, & Bar-Haim Y (2014). *The TAU-NIMH attention bias measurement toolbox*. Laboratory for Research on Anxiety and Trauma. https://people.socsci.tau.ac.il/mu/anxietytrauma/research/

Abend, R., de Voogd, L., Salemink, E., Wiers, R. W., Pérez-Edgar, K., Fitzgerald, A., White, L. K., Salum, G. A., He, J., Silverman, W. K., Pettit, J. W., Pine, D. S., & Bar-Haim, Y. (2018). Association between attention bias to threat and anxiety symptoms in children and adolescents. *Depression and anxiety*, *35*(3), 229-238. http://dx.doi.org/10.1002/da.22706

Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The general aggression model. *Current opinion in psychology*, *19*, 75-80. http://dx.doi.org/10.1016/j.copsyc.2017.03.034

AlMoghrabi, N., Franken, I. H., Mayer, B., & Huijding, J. (2021). A Single-Session Combined Cognitive Bias Modification Training Targeting Attention and Interpretation Biases in Aggression. *Behaviour Change*, 1-20. http://dx.doi.org/10.1017/bec.2021.11

AlMoghrabi, N., Huijding, J., & Franken, I. H. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of behavior therapy and experimental psychiatry*, *58*, 36-42. http://dx.doi.org/10.1016/j.jbtep.2017.08.003

AlMoghrabi, N., Huijding, J., Mayer, B., & Franken, I. H. (2019). Gaze-contingent attention bias modification training and its effect on attention, interpretations, mood, and aggressive behavior. *Cognitive Therapy and Research*, *43*(5), 861-873.

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27e51. http://dx.doi.org/10.1146/annurev.psych.53.100901.135231

APA (2020). *APA Dictionary of Psychology*. American Psychological Association. https://dictionary.apa.org/antisocial

APA (2020). *APA Dictionary of Psychology*. American Psychological Association. https://dictionary.apa.org/aggression.

Arad, G., & Bar-Haim, Y. (2021). Cognitive bias interventions. In A. B. Adler & D. Forbes (Eds.), *Anger at work: Prevention, intervention, and treatment in high-risk occupations* (pp. 275–301). American Psychological Association. https://doi.org/10.1037/0000244-010

Armitage, J., & Eerola, T. (2020). Reaction time data in music cognition: Comparison of pilot data from lab, crowdsourced, and convenience Web samples. *Frontiers in psychology*, *10*(2883). http://dx.doi.org/10.3389/fpsyg.2019.02883

Arsenio, W. F. (2010). Social information processing, emotions, and aggression: Conceptual and methodological contributions of the special section articles. *Journal of abnormal child psychology*, *38*(5), 627-632. DOI: 10.1007/s10802-010-9408-z

Arsenio, W. F., & Lemerise, E. A. (2004). Aggression and moral development: Integrating social information processing and moral domain models. *Child development*, *75*(4), 987-1002.

Assink, M., van der Put, C. E., Hoeve, M., de Vries, S. L., Stams, G. J. J., & Oort, F. J. (2015). Risk factors for persistent delinquent behavior among juveniles: A meta-analytic review. *Clinical psychology review*, *42*, 47-61. http://dx.doi.org/10.1016/j.cpr.2015.08.002

Bacon, A. M., Corr, P. J., & Satchell, L. P. (2018). A reinforcement sensitivity theory explanation of antisocial behaviour. *Personality and Individual Differences*, *123*, 87-93. http://dx.doi.org/10.1016/j.paid.2017.11.008

Bahle, B., Matsukura, M., & Hollingworth, A. (2018). Contrasting gist-based and template-based guidance during real-world visual search. *Journal of Experimental Psychology:*

Human Perception and Performance, *44*(3), 367- 386. http://dx.doi.org/10.1037/xhp0000468

Bahn, D., Vesker, M., García Alanis, J. C., Schwarzer, G., & Kauschke, C. (2017). Age-dependent positivity-bias in children's processing of emotion terms. *Frontiers in Psychology, 8*, 1268. http://dx.doi.org/10.3389/fpsyg.2017.01268

Bailey, C. A., & Ostrov, J. M. (2008). Differentiating forms and functions of aggression in emerging adults: Associations with hostile attribution biases and normative beliefs. *Journal of Youth and Adolescence, 37*, 713–722. http://dx.doi.org/10.1007/s10964-007-9211-5

Bar-Haim, Y., Lamy, D., Pergamin, L., Bakermans-Kranenburg, M.J., & Van Ijzendoorn, M.H. (2007). Threat-related attentional bias in anxious and nonanxious individuals: a meta-analytic study. *Psychological bulletin, 133*(1), 1-24. http://dx.doi.org/10.1037/0033-2909.133.1.1

Bardeen, J. R., Daniel, T. A., Hinnant, J. B., & Orcutt, H. K. (2017). Emotion dysregulation and threat-related attention bias variability. *Motivation and Emotion*, *41*(3), 402-409. http://dx.doi.org/10.1007/s11031-017-9604-z

Baskin-Sommers, A. R., Krusemark, E. A., Curtin, J. J., Lee, C., Vujnovich, A., & Newman, J. P. (2014). The impact of cognitive control, incentives, and working memory load on the P3 responses of externalizing prisoners. *Biological Psychology*, *96*, 86-93. http://dx.doi.org/10.1016/j.biopsycho.2013.12.005

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*(1), 92-107. http://dx.doi.org/10.1111/nyas.13320

Baurain, C., & Nader-Grosbois, N. (2013). Theory of mind, socio-emotional problem-solving, socio-emotional regulation in children with intellectual disability and in typically developing children. *Journal of autism and developmental disorders*, *43*(5), 1080-1097. http://dx.doi.org/10.1007/s10803-012-1651-4

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10):7, 1–11, https://doi.org/10.1167/9.10.7.

Beaver, K. M., Vaughn, M. G., DeLisi, M., Barnes, J. C., & Boutwell, B. B. (2011). The neuropsychological underpinnings to psychopathic personality traits in a nationally representative and longitudinal sample. *Psychiatric Quarterly, 83*(2), 1-15, https://doi.org/10.1007/s11126-011-9190-2

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision research*, *49*(10), 1154-1165. https://doi.org/10.1016/j.visres.2008.07.012

Berggren, N., & Eimer, M. (2018). Object-based target templates guide attention during visual search. Journal of Experimental Psychology: *Human Perception and Performance, 44*(9), 1368-1382. https://doi.org/10.1037/xhp0000541

Bertsch, K., Böhnke, R., Kruk, M. R., & Naumann, E. (2009). Influence of aggression on information processing in the emotional Stroop task-an event-related potential study. *Frontiers in behavioral neuroscience*, *3*, 28- 38. https://doi.org/10.3389/neuro.08.028.2009

Bertsch, K., Florange, J., & Herpertz, S. C. (2020). Understanding brain mechanisms of reactive aggression. *Current psychiatry reports*, *22*(12), 1-16. https://doi.org/10.1007/s11920-020-01208-6

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

Bondü, R., & Richter, P. (2016). Interrelations of justice, rejection, provocation, and moral disgust sensitivity and their links with the hostile attribution bias, trait anger, and aggression. *Frontiers in Psychology*, *7*, 795- 810. https://doi.org/10.3389/fpsyg.2016.00795

Boxer, P., & Dubow, E. F. (2001). A social-cognitive information-processing model for school-based aggression reduction and prevention programs: Issues for research and practice. *Applied and Preventive Psychology*, *10*(3), 177-192. https://doi.org/10.1016/S0962-1849(01)80013-5

Boxer, P., Guerra, N. G., Huesmann, L. R., & Morales, J. (2005). Proximal peer-level effects of a small-group selected prevention on aggression in elementary school children: An investigation of the peer contagion hypothesis. *Journal of abnormal child psychology*, *33*(3), 325-338. https://doi.org/10.1007/s10802-005-3568-2

Brennan, G. M., & Baskin-Sommers, A. R. (2020). Aggressive realism: More efficient processing of anger in physically aggressive individuals. *Psychological science*, *31*(5), 568-581. https://doi.org/10.1177/0956797620904157

Bridges, L. J., Denham, S. A., & Ganiban, J. M. (2004). Definitional issues in emotion regulation research. *Child Development,* *75*(2), 340–345. https://doi.org/10.1016/j.avb.2011.09.006

Brislin, S. J., Yancey, J. R., Perkins, E. R., Palumbo, I. M., Drislane, L. E., Salekin, R. T., Fanti, K. A., Kimonis, E. R., Frick, P. J., Blair, R. J. R., & Patrick, C. J. (2018). Callousness and affective face processing in adults: Behavioral and brain-potential indicators. *Personality Disorders: Theory, Research, and Treatment*, *9*(2), 122- 132. doi: 10.1037/per0000235.

Brugman, S., Cornet, L. J., Smeijers, D., Smeets, K., Oostermeijer, S., Buitelaar, J. K., Verkes, R. J., Lobbestael, J., de Kogel, C. H., & Jansen, L. M. (2017). Examining the reactive proactive questionnaire in adults in forensic and non-forensic settings: A variable-and person-based approach. *Aggressive Behavior*, *43*(2), 155-162. doi: 10.1002/ab.21671

Brugman, S., Lobbestael, J., Sack, A. T., Cima, M. J., Schuhmann, T., Emmerling, F., & Arntz, A. (2018). Cognitive predictors of reactive and proactive aggression in a forensic sample: A comparison with a non-clinical sample. *Psychiatry research*, *269*, 610-620. https://doi.org/10.1016/j.psychres.2018.08.095

Brugman, S., Lobbestael, J., von Borries, A. K. L., Bulten, B. E. H., Cima, M., Schuhmann, T., Dambacher, F., Sack, A. T., & Arntz, A. (2016). Cognitive predictors of violent incidents in forensic psychiatric inpatients. *Psychiatry research*, *237*, 229-237. doi: 10.1016/j.psychres.2016.01.035

Buades-Rotger, M., & Krämer, U. M. (2018). From words to action: Implicit attention to antisocial semantic cues predicts aggression and amygdala reactivity to angry faces in healthy young women. *Aggressive behavior*, *44*(6), 624-637. https://doi.org/10.1002/ab.21787

Burra, N., & Kerzel, D. (2019). Task demands modulate effects of threatening faces on early perceptual encoding. *Frontiers in psychology*, 1-9. https://doi.org/10.3389/fpsyg.2019.02400

Burra, N., Barras, C., Coll, S. Y., & Kerzel, D. (2016). Electrophysiological evidence for attentional capture by irrelevant angry facial expressions. *Biological Psychology*, *120*, 69-80. https://doi.org/10.1016/j.biopsycho.2016.08.008

Burra, N., Coll, S. Y., Barras, C., & Kerzel, D. (2017). Electrophysiological evidence for attentional capture by irrelevant angry facial expressions: Naturalistic faces. *Neuroscience Letters*, *637*, 44-49. doi: 10.1016/j.neulet.2016.11.055

Bushman, B. J. (2016). Violent media and hostile appraisals: A meta-analytic review. *Aggressive Behavior*, *42*(6), 605-613. https://doi.org/10.1002/ab.21655

Butler, S. M., Parry, R., & Fearon, R. M. (2015). Antisocial thinking in adolescents: Further psychometric development of the Antisocial Beliefs and Attitudes Scale (ABAS). *Psychological Assessment, 27*(1), 291-301. https://doi.org/10.1037/a0038143

Byrd, A. L., Kahn, R. E., & Pardini, D. A. (2013). A validation of the Inventory of Callous-Unemotional Traits in a community sample of young adult males. *Journal of psychopathology and behavioral assessment*, *35*(1), 20-34. https://doi.org/10.1007/s10862-012-9315-4

Cabral, J. C. C., & de Almeida, R. M. M. (2019). Effects of anger on dominance-seeking and aggressive behaviors. *Evolution and Human Behavior*, *40*(1), 23-33. https://doi.org/10.1016/j.evolhumbehav.2018.07.006

Calvete, E., & Orue, I. (2012). The role of emotion regulation in the predictive association between social information processing and aggressive behavior in adolescents. *International Journal of Behavioral Development*, *36*(5), 338-347. https://doi.org/10.1177/0165025412444079

Caramaschi, D., de Boer, S. F., de Vries, H., & Koolhaas, J. M. (2008). Development of violence in mice through repeated victory along with changes in prefrontal cortex neurochemistry. *Behavioural Brain Research, 189*(2), 263-272. https://doi.org/http://dx.doi.org/10.1016/j.bbr.2008.01.003

Cardinale, E. M., & Marsh, A. A. (2020). The reliability and validity of the Inventory of Callous Unemotional Traits: A meta-analytic review. *Assessment*, *27*(1), 57-71. https://doi.org/10.1177/1073191117747392

Carlisle, N. B. (2019). Focus: Attention Science: Flexibility in Attentional Control: Multiple Sources and Suppression. *The Yale Journal of Biology and Medicine*, *92*(1), 103- 113.

Carlisle, N. B., & Woodman, G. F. (2011). When memory is not enough: Electrophysiological evidence for goal-dependent use of working memory representations in guiding visual attention. *Journal of Cognitive Neuroscience*, *23*(10), 2650-2664. https://doi.org/10.1162/jocn.2011.21602

Carlisle, N. B., & Woodman, G. F. (2013). Reconciling conflicting electrophysiological findings on the guidance of attention by working memory. *Attention, Perception, & Psychophysics*, *75*(7), 1330-1335. https://doi.org/10.3758/s13414-013-0529-7

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, *31*(25), 9315-9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

Carolan, P. L., Gaspar, J. M., Kleffner, K., & Liotti, M. (2020). Electrophysiological evidence that psychopathic personality traits are associated with atypical response to salient distractors. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(1), 195-213. https://doi.org/10.3758/s13415-019-00762-8

Carroll, G. A., Montrose, V. T., & Burke, T. (2021). Correlates of Social Cognition and Psychopathic Traits in a Community-Based Sample of Males. *Frontiers in Psychology, 12*, 656299. https://doi.org/10.3389/fpsyg.2021.656299

Cenkseven-Önder, F., Avci, R., & Çolakkadioglu, O. (2016). Validity and Reliability of the Reactive-Proactive Aggression Questionnaire in Turkish Adolescents. *Educational Research and Reviews*, *11*(20), 1931-1943. https://doi.org/10.5897/ERR2016.2937

Chabrol, H., Goutaudier, N., Melioli, T., van Leeuwen, N., & Gibbs, J. C. (2014). Impact of antisocial behavior on psychopathic traits in a community sample of adolescents. *Bulletin of the Menninger Clinic, 78*(3), 228-242.

Chambers, J. C., Eccleston L., Day, A., Ward, T., & Howells, K. (2008). Treatment readiness in violent offenders: The influence of cognitive factor on engagement in violence programs. *Aggression and Violent Behavior, 13*, 276-284. https://doi.org/10.1016/j.avb.2008.04.003

Chan, S. C., Raine, A., & Lee, T. M. (2010). Attentional bias towards negative affect stimuli and reactive aggression in male batterers. *Psychiatry Research*, *176*(2-3), 246-249. https://doi.org/10.1016/j.psychres.2008.12.013

Chapman, A., Devue, C., & Grimshaw, G. M. (2019). Fleeting reliability in the dot-probe task. *Psychological Research*, *83*(2), 308-320. https://doi.org/10.1007/s00426-017-0947-6

Chen, P., Coccaro, E.F., Jacobson, K.C., 2012. Hostile attributional bias, negative emotional responding, and aggression in adults: Moderating effects of gender and impulsivity. *Aggressive Behaviour, 38* (1), 47–63.

Chereji, S. V., Pintea, S., & David, D. (2012). The relationship of anger and cognitive distortions with violence in violent offenders population: a meta-analytic review. *European Journal of Psychology Applied to Legal Context*, *4*(1). 59-77.

Chester, D.S., & DeWall, C.N. (2017). Combating the sting of rejection with the pleasure of revenge: A new look at how emotion shapes aggression. *Journal of Personality and Social Psychology*, *112*(3), 413- 430. https://doi.org/10.1037/pspi0000080

Chiles, J. A., Von Cleve, E., Jemelka, R. P., & Trupin, E. W. (1990). Substance abuse and psychiatric disorders in prison inmates. *Hospital & Community Psychiatry, 41*, 1132–1134. https://doi.org/10.1176/ps.41.10.1132

Choe, D. E., Lane, J. D., Grabell, A. S., & Olson, S. L. (2013). Developmental precursors of young school-age children's hostile attribution bias. *Developmental Psychology, 49*(12), 2245–2256. https://doi.org/10.1037/a0032293

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia, 49*(6), 1407-1409. https://doi.org/10.1016/j.neuropsychologia.2011.01.029

Cisler, J. M., & Koster, E. H. (2010). Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clinical psychology review*, *30*(2), 203-216. https://doi.org/10.1016/j.cpr.2009.11.003

Ciucci, E., Kimonis, E., Frick, P. J., Righi, S., Baroncelli, A., Tambasco, G., & Facci, C. (2018). Attentional orienting to emotional faces moderates the association between callous-unemotional traits and peer-nominated aggression in young adolescent school children. *Journal of abnormal child psychology*, *46*(5), 1011-1019. https://doi.org/10.1007/s10802-017-0357-7

Coccaro, E. F., Fanning, J. R., Keedy, S. K. & Lee, R. J. (2016). Social cognition in Intermittent Explosive Disorder and aggression. *Journal of psychiatric research, 83*, 140-150. https://doi.org/10.1016/j.jpsychires.2016.07.010

Coccaro, E. F., Fanning, J., & Lee, R. (2017). Development of a social emotional information processing assessment for adults (SEIP-Q). *Aggressive behavior*, *43*(1), 47-59. https://doi.org/10.1002/ab.21661

Coccaro, E. F., Noblett, K. L., & McCloskey, M. S. (2009). Attributional and emotional responses to socially ambiguous cues: Validation of a new assessment of social/emotional information processing in healthy adults and impulsive aggressive patients. *Journal of Psychiatric Research*, *43*(10), 915-925. https://doi.org/10.1016/j.jpsychires.2009.01.012

Coelho, C. M., Cloete, S., & Wallis, G. (2010). The face-in-the-crowd effect: When angry faces are just cross (es). *Journal of Vision*, *10*(1), 7-7. https://doi.org/10.1167/10.1.7

Cohen, D. J., Eckhardt, C. I., & Schagat, K. D. (1998). Attention allocation and habituation to anger-related stimuli during a visual search task. *Aggressive Behavior, 24*, 399-409.

Cohen, J. (1988). A power primer. *Psychological Bulletin, 112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cohen, M. A., Piquero, A. R., & Jennings, W. G. (2010). Studying the costs of crime across offender trajectories. *Criminology & Public Policy*, *9*(2), 279-305.

Cohn, A. M., Jakupcak, M., Seibert, L. A., Hildebrandt, T. B., & Zeichner, A. (2010). The role of emotion dysregulation in the association between men's restrictive emotionality and use of physical aggression. *Psychology of Men & Masculinity*, *11*(1), 53-64. https://doi.org/10.1037/a0018090

Cooper, R. M., & Langton, S. R. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour research and therapy*, *44*(9), 1321-1329. https://doi.org/10.1016/j.brat.2005.10.004

Cougle, J. R., Summers, B. J., Allan, N. P., Dillon, K. H., Smith, H. L., Okey, S. A., & Harvey, A. M. (2017). Hostile interpretation training for individuals with alcohol use disorder and elevated trait anger: A controlled trial of a web-based intervention. *Behaviour research and therapy, 99*, 57-66. https://doi.org/10.1016/j.brat.2017.09.004 0005-7967

Coyne, S. M. (2016). Effects of viewing relational aggression on television on aggressive behavior in adolescents: A three-year longitudinal study. *Developmental Psychology, 52*(2), 284–295. https://doi.org/10.1037/dev0000068

Crago, R. V., Renoult, L., Biggart, L., Nobes, G., Satmarean, T., & Bowler, J. O. (2019). Physical aggression and attentional bias to angry faces: an event related potential study. *Brain research*, *1723*, 146387. https://doi.org/10.1016/j.brainres.2019.146387

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101. http://dx.doi.org/10.1037/0033-2909.115.1.74.

Crick, N., & Dodge, K. (1996). Social information processing mechanisms in reactive and proactive aggression. *Child Development, 67*, 993–1002. https://doi.org/10.2307/1131875

Crocker, L., Heller, W., Spielberg, J., Warren, S., Bredemeier, K., Sutton, B., Banich, M. T., & Miller, G. (2012). Neural mechanisms of attentional control differentiate trait and state negative affect. *Frontiers in Psychology*, 3, 298. https://doi.org/10.3389/fpsyg.2012.00298

Crowe, S. E., & Wilkowsky, B. M. (2013). Looking for trouble: Revenge-planning and preattentive vigilance for angry facial expressions. *Emotion, 13*(4), 774–781. https://doi.org/10.1037/a0032252

Curby, K. M., Johnson, K. J., & Tyson, A. (2012). Face to face with emotion: Holistic face processing is modulated by emotional state. *Cognition & emotion*, *26*(1), 93-102. https://doi.org/1080/02699931.2011.555752

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory. *Annual Review of Psychology, 66*(1), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Davies, P. T., Coe, J. L., Hentges, R. F., Sturge-Apple, M. L., & Ripple, M. T. (2020). Temperamental emotionality attributes as antecedents of children's social information processing. *Child development, 91*(2), 508-526. https://doi.org/10.1111/cdev.13191

Day, A., Daffern, M., Polaschek, D. L., Dunne, A., & Senn, A. (2021). The classification of people with a history of interpersonal violence for correctional treatment purposes:

Possibilities for a schema-informed approach. *Aggression and violent behavior*, *59*, 101450. https://doi.org/10.1016/j.avb.2020.101450

de Beeck, H. P. O., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends in cognitive sciences*, *14*(1), 22-30. https://doi.org/10.1016/j.tics.2009.11.002

de Castro, B., Merk, W., Koops, W., Veerman, J. W., & Bosch, J. D. (2005). Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology*, *34*(1), 105-116. https://doi.org/10.1207/s15374424jccp3401_10

de Castro, B.O., Slot, N. W., Bosch, J. D., Koops, W., & Veerman, J. W. (2003). Negative feelings exacerbate hostile attributions of intent in highly aggressive boys. *Journal of Clinical Child and Adolescent Psychology*, *32*(1), 56-65.

Denson, T. F. (2009). Angry rumination and the self-regulation of aggression. In J. P. Forgas, R. F. Baumeister, & D. M. Tice (Eds.), *The psychology of self-regulation* (pp. 233–248). New York, NY: Psychology Press.

Denson, T. F., DeWall, C. N., & Finkel, E. J. (2012). Self-control and aggression. *Current Directions in Psychological Science, 21*(1), 20–25. https://doi.org/10.1177/0963721411429451

Denson, T. F., Pedersen, W. C., Friese, M., Hahm, A., & Roberts, L. (2011). Understanding impulsive aggression: angry rumination and reduced self-control capacity are mechanisms underlying the provocation- aggression relationship. *Personality and Social Psychology Bulletin, 37*, 850–862. https://doi.org/10.1177/0146167211401420

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dickins, D. S., & Lipp, O. V. (2014). Visual search for schematic emotional faces: Angry faces are more than crosses. *Cognition and Emotion, 28*(1), 98–114. https://doi.org/10.1080/02699931.2013.809331

Dodge, K. A, & Rabiner, D. L. (2004). Returning to roots: On social information processing and moral development. *Child Development*, *75*(4), 1003-1008.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51*, 162–170.

Dodge, K. A., & McCourt, S. N. (2010). Translating models of antisocial behavioral development into efficacious intervention policy to prevent adolescent violence. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *52*(3), 277-285. https://doi.org/10.1002/dev.20440

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental psychology*, *39*(2), 349- 371. https://doi.org/10.1037/0012-1649.39.2.349

Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science*, *250*(4988), 1678-1683.

Dodge, K. A., Godwin, J., & Conduct Problems Prevention Research Group. (2013). Social-information-processing patterns mediate the impact of preventive intervention on adolescent antisocial behavior. *Psychological Science*, *24*(4), 456-465. https://doi.org/10.1177/0956797612457394

Dodge, K. A., Lansford, J. E., Burks, V. S., Bates, J. E., Pettit, G. S., Fontaine, R., & Price, J. M. (2003). Peer rejection and social information-processing factors in the development of aggressive behavior problems in children. *Child development*, *74*(2), 374-393.

Dodge, K. A., Malone, P. S., Lansford, J. E., Sorbring, E., Skinner, A. T., Tapanya, S., Uribe Tirado, L. M., Zelli, A., Peña Alampay, L., Al-Hassan, S. M., Bacchini, D., Bombi, A.

S., Bornstein, M. H., Chang, L., Deater-Deckard, K., Di Giunta, L., Oburu, P. & Pastorelli, C. (2015). Hostile attributional bias and aggressive behavior in global context. *Proceedings of the National Academy of Sciences*, *112*(30), 9310-9315.

Dodge, K. A., Pettit, G. S., Bates, J. E., & Valente, E. (1995). Social information-processing patterns partially mediate the effect of early physical abuse on later conduct problems. *Journal of Abnormal Psychology, 104*(4), 632-643. https://doi.org/10.1037/0021-843X.104.4.632

Dodge, K. A., Pettit, G. S., McClaskey, C. L., Brown, M. M., & Gottman, J. M. (1986). Social competence in children. *Monographs of the society for research in child development*, *51*(2), 1-89.

Dodge, K. A., Price, J. M., Bachorowski, J. A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology*, *99*(4), 385-392.

Dodge, K.A., & Coie, J.D. (1987). Social-information-processing factors in reactive and proactive aggression in childrens' peer groups. *Journal of Personality and Social Psychology, 53* (6), 1146–1158.

Doherty, B. R., Patai, E. Z., Duta, M., Nobre, A. C., & Scerif, G. (2017). The functional consequences of social distraction: Attention and memory for complex scenes. *Cognition*, *158*, 215-223. https://doi.org/10.1016/j.cognition.2016.10.015 0010-0277/Ó

Domes, G., Mense, J., Vohs, K., & Habermeyer, E. (2013). Offenders with antisocial personality disorder show attentional bias for violence-related stimuli. *Psychiatry Research*, *209*(1), 78-84. https://doi.org/10.1016/j.psychres.2012.11.005

Donahue, J. J., Goranson, A. C., McClure, K. S., & Van Male, L. M. (2014). Emotion dysregulation, negative affect, and aggression: A moderated, multiple mediator analysis. *Personality and Individual Differences*, *70*, 23-28. https://doi.org/10.1016/j.paid.2014.06.009

Dong, Y., De Beuckelaer, A., Yu, L., & Zhou, R. (2017). Eye-movement evidence of the time-course of attentional bias for threatening pictures in test-anxious students. *Cognition and Emotion*, *31*(4), 781-790. https://doi.org/10.1080/02699931.2016.1152953

Dowd, E. W., Kiyonaga, A., Egner, T., & Mitroff, S. R. (2015). Attentional guidance by working memory differs by paradigm: An individual-differences approach. *Attention, Perception, & Psychophysics*, *77*(3), 704-712. https://doi.org/10.3758/s13414-015-0847-z

Dunne, A. L., Gilbert, F., Lee, S., & Daffern, M. (2018). The role of aggression-related early maladaptive schemas and schema modes in aggression in a prisoner sample. *Aggressive behavior*, *44*(3), 246-256. https://doi.org/10.1002/ab.21747

Dunne, A. L., Lee, S., & Daffern, M. (2019). Extending the general aggression model: Contributions of DSM-5 maladaptive personality facets and schema modes. *Psychology, Crime & Law, 25*(9), 875–895. https://doi.org/10.1080/1068316X.2019.1597089

Edalati, H., Walsh, Z., & Kosson, D. S. (2016). Attentional bias in psychopathy: An examination of the emotional dot-probe task in male jail inmates. *International Journal of Offender Therapy and Comparative Criminology*, *60*(11), 1344-1357. https://doi.org/10.1177/0306624X15577791

Elbert, T., Shauer, M., & Moran, J. (2017). Two pedals drive the bi-cycle of violence: Reactive and appetitive aggression. *Current Opinion in Psychology, 19*, 135–138.

Eme, R. (2009). Male life-course persistent antisocial behavior: A review of neurodevelopmental factors. *Aggression and Violent Behavior*, *14*(5), 348-358.

Enzmann, D. (2013). The impact of questionnaire design on prevalence and incidence rates of self-reported delinquency: Results of an experiment modifying the ISRD-2 questionnaire. *Journal of Contemporary Criminal Justice*, *29*(1), 147-177. https://doi.org/10.1177/1043986212470890

Erdley, C. A., & Asher, S. R. (1996) Children's social goals and self-efficacy perceptions as influences on their responses to ambiguous provocation. *Child Development*, *67*, 1329–1344.

Evans, T. C., Bar-Haim, Y., Fox, N. A., Pine, D. S., & Britton, J. C. (2020). Neural mechanisms underlying heterogeneous expression of threat-related attention in social anxiety. *Behaviour research and therapy*, *132*, 103657. https://doi.org/10.1016/j.brat.2020.103657

Fan, L., Ding, C., Guo, R., Xu, M., Diao, L., & Yang, D. (2016). Visual working memory representations guide the detection of emotional faces: An ERP study. *Vision Research*, *119*, 1-8. https://doi.org/10.1016/j.visres.2015.08.021

Fanning, J. R., Berman, M. E., & Long, J. M. (2014). P3 and provoked aggressive behavior. *Social neuroscience*, *9*(2), 118-129. https://doi.org/10.1080/17470919.2013.866596

Fanti, K. A., Frick, P. J., & Georgiou, S. (2009). Linking callous-unemotional traits to instrumental and non-instrumental forms of aggression. *Journal of Psychopathology and Behavioral Assessment*, *31*(4), 285-298. https://doi.org/10.1007/s10862-008-9111-3

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience, 4,* 215-238. https://doi.org/10.3389/fnhum.2010.00215

Ferguson, C. J. (2010). Blazing Angels or Resident Evil? Can violent video games be a force for good? *Review of General Psychology, 14*(2), 68–81.

Ferguson, C. J., & Dyck, D. (2012). Paradigm change in aggression research: The time has come to retire the General Aggression Model. *Aggression and Violent Behavior*, *17*(3), 220-228.

Fiebelkorn, I. C., & Kastner, S. (2019). A rhythmic theory of attention. *Trends in cognitive sciences*, *23*(2), 87-101. https://doi.org/10.1016/j.tics.2018.11.009

Figueira, J. S., Oliveira, L., Pereira, M. G., Pacheco, L. B., Lobo, I., Motta-Ribeiro, G. C., & David, I. A. (2017). An unpleasant emotional state reduces working memory capacity: electrophysiological evidence. Social Cognitive and Affective Neuroscience, 12(6), 984-992. https://doi.org/10.1093/scan/nsx030

Fontaine, R. G., & Dodge, K. A. (2006). Real-time decision making and aggressive behavior in youth: A heuristic model of response evaluation and decision (RED). *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *32*(6), 604-624. https://doi.org/10.1002/ab.20150

Fontaine, R. G., Tanha, M., Yang, C., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2010). Does response evaluation and decision (RED) mediate the relation between hostile attributional style and antisocial behavior in adolescence?. *Journal of abnormal child psychology, 38*(5), 615-626. https://doi.org/10.1007/s10802-010-9397-y

Fontaine, R. G., Yang, C., Burks, V. S., Dodge, K. A., Price, J. M., Pettit, G. S., & Bates, J. E. (2009). Loneliness as a partial mediator of the relation between low social preference in childhood and anxious/depressed symptoms in adolescence. *Development and Psychopathology, 21* (2), 479–491. https://doi.org/10.1017/S0954579409000261

Fontaine, R. G., Yang, C., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2008). Testing an individual systems model of response evaluation and decision (RED) and antisocial behavior across adolescence. *Child Development*, **79**(2), 462-475.

Ford, B. Q., Tamir, M., Brunyé, T. T., Shirer, W. R., Mahoney, C. R., & Taylor, H. A. (2010). Keeping your eyes on the prize: Anger and visual attention to threats and rewards. *Psychological Science*, *21*(8), 1098-1105. https://doi.org/10.1177/0956797610375450

Ford, B. Q., Tamir, M., Gagnon, S. A., Taylor, H. A., & Brunyé, T. T. (2012). The angry spotlight: Trait anger and selective visual attention to rewards. *European Journal of Personality*, *26*(2), 90-98. https://doi.org/10.1002/per.1840

Fox, E., Derakshan, N., & Shoker, L. (2008). Trait anxiety modulates the electro- physiological indices of rapid spatial orienting towards angry faces. *Neuroreport, 19*, 259–263. https://doi.org/10.1097/WNR.0b013e3282f53d2a

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203), 1502-1505. https://doi.org/10.1126/science.1255484

Fraser, M. W., Galinsky, M. J., Smokowski, P. R., Day, S. H., Terzian, M. A., Rose, R. A., & Guo, S. (2005). Social information-processing skills training to promote social competence and prevent aggressive behavior in the third grade. *Journal of consulting and clinical psychology*, *73*(6), 1045- 1055. https://doi.org/10.1037/0022-006X.73.6.1045.

Frick, P. J. (2004). *The Inventory of Callous-Unemotional Traits*. University of New Orleans, Developmental Psychopathology Laboratory. http://labs.uno.edu/developmental-psychopathology/ICU.html

Frick, P. J., & White, S. F. (2008). Research review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of child* *psychology and psychiatry*, *49*(4), 359-375. https://doi.org/10.1111/j.1469-7610.2007.01862.x

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological bulletin*, *140*(1), 1-57.

Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in cognitive sciences*, *13*(7), 293-301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127-138. https://doi.org/10.1038/nrn2787

Friston, K. & Kiebel, S. (2009) Cortical circuits for perceptual inference. *Neural Networks, 22,* 1093–104.

Fu, X., & Pérez-Edgar, K. (2019). Threat-related attention bias in socioemotional development: A critical review and methodological considerations. *Developmental Review, 51*, 31–57. https ://doi.org/10.1016/j.dr.2018.11.002

Fundamental Rights Report (2021, June 10). *Violence and harassment across Europe much higher than official records*. European Union Agency for Fundamental Rights. https://fra.europa.eu/en/news/2021/violence-and-harassment-across-europe-much-higher-official-records

Gable, P. A., Poole, B. D., & Harmon-Jones, E. (2015). Anger perceptually and conceptually narrows cognitive scope. *Journal of Personality and Social Psychology*, *109*(1), 163-174. https://doi.org/10.1037/a0039226

Gagnon, J., & Rochat, L. (2017). Relationships between hostile attribution bias, negative urgency, and reactive aggression. *Journal of Individual Differences*, 38(4), 211-219. https://doi.org/10.1027/1614-0001/a000238

Gagnon, J., McDuff, P., Daelman, S., & Fournier, S. (2015). Is hostile attributional bias associated with negative urgency and impulsive behaviors? A social-cognitive conceptualization of impulsivity. *Personality and Individual Differences, 72*, 18-23. https://doi.org/10.1016/j.paid.2014.08.011

Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Nonnormality of data in structural equation models. *Transportation Research Record, 2082*(1), 116-124. https://doi.org/10.3141/2082-14

Garofalo, C., Neumann, C. S., & Velotti, P. (2018). Difficulties in emotion regulation and psychopathic traits in violent offenders. *Journal of criminal justice*, *57*, 116-125. https://doi.org/10.1016/j.jcrimjus.2018.05.013

Garofalo, C., Neumann, C. S., & Velotti, P. (2020). Psychopathy and aggression: The role of emotion dysregulation. *Journal of Interpersonal Violence*, *00*(0), 1-25. https://doi.org/10.1177/0886260519900946

Gasse, A., Kim, W. S., & Gagnon, J. (2020). Association between depression and hostile attribution bias in hostile and non-hostile individuals: An ERP study. *Journal of affective disorders*, *276*, 1077-1083. https://doi.org/10.1016/j.jad.2020.06.025

Gauthier, I. and Logothetis, N.K. (2000) Is face recognition not so unique after all? *Cognitive Neuropsychology, 17*, 125–142.

Gerin, W., Davidson, K. W., Christenfeld, N. J. S., Goyal, T., & Schwartz, J. E. (2006). The role of angry rumination and distraction in blood pressure recovery from emotional arousal. *Psychosomatic Medicine, 68,* 64–72.

Geurts, H. M., Verté, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *Journal of Child Psychology and Psychiatry, 45*(4), 836–854

Gilbert, F. and Daffern, M., 2017. Aggressive scripts, violent fantasy and violent behavior: a conceptual clarification and review. *Aggression and violent behavior, 36,* 98-107. https://doi.org/10.1016/j.avb.2017.05.001

Gilbert, F., Daffern, M., Talevski, D., & Ogloff, J. R. (2013). The role of aggression-related cognition in the aggressive behavior of offenders: A general aggression model perspective. *Criminal justice and behavior*, *40*(2), 119-138. https://doi.org/10.1177/0093854812467943.

Gomes, H. S., Farrington, D. P., Maia, Â., & Krohn, M. D. (2019). Measurement bias in self-reports of offending: a systematic review of experiments. *Journal of experimental criminology*, *15*(3), 313-339. https://doi.org/10.1007/s11292-019-09379-w

Gong, M., & Smart, L. J. (2021). The anger superiority effect revisited: a visual crowding task. *Cognition and Emotion*, *35*(2), 214-224. https://doi.org/10.1080/02699931.2020.1818552

Gooding, D. C., & Tallent, K. A. (2003). Spatial, object, and affective working memory in social anhedonia: An exploratory study. Schizophrenia Research, 63, 247–260. https://doi.org/10.1016/s0920-9964(02)00326-2

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: development, factor structure, and initial validation of the difficulties in emotion regulation scale. Journal of Psychopathology and Behavioral Assessment. https://doi.org/10. 1023/B:JOBA.0000007455.08539.94.

Grecucci, A., Soto, D., Rumiati, R. I., Humphreys, G. W., & Rotshtein, P. (2010). The interrelations between verbal working memory and visual selection of emotional faces. *Journal of Cognitive Neuroscience*, *22*(6), 1189-1200.

Grubert, A., Schmid, P., & Krummenacher, J. (2013). Happy with a difference, unhappy with an identity: Observers' mood determines processing depth in visual search. *Attention,* *Perception, & Psychophysics*, *75*(1), 41-52. https://doi.org/10.3758/s13414-012-0385-x

Guerra, N. G. and Slaby, R. G. (1990). Cognitive mediators of aggression in adolescent offenders: II. Intervention. *Developmental psychology*, **26**(2), 269-277.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917–924.

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2013). Does negative affect always narrow and positive affect always broaden the mind? Considering the influence of motivational intensity on cognitive scope. *Current Directions in Psychological Science, 22*, 301-307. https://doi.org/10.1177%2F0963721413481353

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychology review*, *20*(3), 290-322.

Harper, B. D., Lemerise, E. A. and Caverly, S. L. (2010). The effect of induced mood on children's social information processing: Goal clarification and response decision. *Journal of Abnormal Child Psychology, 38*(5), 575-586. https://doi.org/10.1007/s10802-009-9356-7

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 968-987). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470939376.ch25

Hastings, M. E., Tangney, J. P., & Stuewig, J. (2008). Psychopathy and identification of facial expressions of emotion. *Personality and individual differences*, *44*(7), 1474-1483. https://doi.org/10.1016/j.paid.2008.01.004

Hatfield, J., & Dula, C. S. (2014). Impulsivity and physical aggression: examining the moderating role of anxiety. *The American journal of psychology, 127*(2), 233-243. https://doi.org/10.5406/amerjpsyc.127.2.0233

Hawkins, K. A., & Cougle, J. R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy*, *44*(3), 479-488. https://doi.org/10.1016/j.beth.2013.04.005

Hazebroek, J. F., Howells, K., & Day, A. (2001). Cognitive appraisals associated with high trait anger. *Personality and Individual Differences, 30*(1), 31–45. https://doi.org/ 10.1016/S0191-8869(00)00007-6

Helfritz-Sinville, L. E., & Stanford, M. S. (2015). Looking for trouble? Processing of physical and social threat words in impulsive and premeditated aggression. *The Psychological Record*, *65*(2), 301-314.

Helmsen, J., Koglin, U., & Petermann, F. (2012). Emotion regulation and aggressive behavior in preschoolers: The mediating role of social information processing. *Child Psychiatry & Human Development*, *43*(1), 87-101. https://doi.org/10.1007/s10578-011-0252-3

Helseth, S. A., Waschbusch, D. A., King, S., & Willoughby, M. T. (2015). Aggression in children with conduct problems and callous-unemotional traits: Social information processing and response to peer provocation. *Journal of abnormal child psychology*, *43*(8), 1503-1514. https://doi.org/10.1007/s10802-015-0027-6

Hesser, H., Axelsson, S., Bäcke, V., Engstrand, J., Gustafsson, T., Holmgren, E., Jeppsson, U., Pollack, M., Nordén, K., Rosenqvist, D., & Andersson, G. (2017). Preventing intimate partner violence via the Internet: A randomized controlled trial of emotion-regulation and conflict-management training for individuals with aggression problems. *Clinical psychology & psychotherapy*, *24*(5), 1163-1177. https://doi.org/10.1002/cpp.2082

Hiemstra, W., De Castro, B. O., & Thomaes, S. (2019). Reducing aggressive children's hostile attributions: A cognitive bias modification procedure. *Cognitive therapy and research*, *43*(2), 387-398. https://doi.org/10.1007/s10608-018-9958-x

Hilbig, B. E. (2016). Reaction time effects in lab-versus web-based research: experimental evidence. *Behavioural Research Methods, 48,* 1718–1724. https://doi.org/10.3758/s13428-015-0678-9

Home Office (2018). *The economic and social costs of crime. Second edition*. HM Government. Retrieved December 26, 2021, from The economic and social costs of crime (publishing.service.gov.uk)

Horsley, T. A., de Castro, B. O., & Van der Schoot, M. (2010). In the eye of the beholder: Eye-tracking assessment of social information processing in aggressive behavior. *Journal of Abnormal Child Psychology*, *38*(5), 587-599. https://doi.org/10.1007/s10802-009-9361-x

Hosie, J., Dunne, A. L., Meyer, D., & Daffern, M. (2021). Aggressive script rehearsal in adult offenders: Relationships with emotion regulation difficulties and aggressive behavior. *Aggressive behavior*, *48*(1), 5-16. https://doi.org/10.1002/ab.22000

Houssa, M., Jacobs, E., & Nader-Grosbois, N. (2017). Impact of Short-Term Training in Social Cognition in Preschoolers with Externalizing Behavior. *Journal of Education and Training Studies*, *5*(1), 110-123. https://doi.org/10.11114/jets.v5i1.2081

Huang, S.-L., Chang, Y.-C., & Chen, Y.-J. (2011). Task-irrelevant angry faces capture attention in visual search while modulated by resources. *Emotion, 11*(3), 544–552. https://doi.org/10.1037/a0022763

Huesmann, L. R. (1988). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In R. G. Geen & E.

Donnerstein, *Human aggression: Theories, Research, and Implications for Social Policy* (pp. 73-109). Academic Press. Santa Barbara.

Huesmann, L., & Taylor, L. D. (2006). The role of media violence in violent behavior. *Annual Review of Public Health,* 27393– 415. https://doi.org/10.1146/annurev.publhealth.26.021304.144640.

Hur, J., Miller, G. A., McDavitt, J. R. B., Spielberg, J. M., Crocker, L. D., Infantolino, Z. P., Towers, D. N., Warren, S. L., & Heller, W. (2015). Interactive effects of trait and state affect on top-down control of attention. *Social Cognitive and Affective Neuroscience, 10*(8), 1128–1136. https://doi.org/10.1093/scan/nsu163

Isaacowitz, D. M., Wadlinger, H. A., Goren, D., & Wilson, H. R. (2006). Is there an age-related positivity effect in visual attention? A comparison of two methodologies. *Emotion, 6*(3), 511–516. https://doi.org/10.1037/1528-3542.6.3.511

Jackson, M. C., Linden, D. E., & Raymond, J. E. (2014). Angry expressions strengthen the encoding and maintenance of face identity representations in visual working memory. *Cognition & Emotion*, *28*(2), 278-297. https://doi.org/10.1080/02699931.2013.816655

Jackson, M. C., Wu, C. Y., Linden, D. E., & Raymond, J. E. (2009). Enhanced visual short-term memory for angry faces. *Journal of Experimental Psychology Human Perception and Performance, 35*(2), 363–374. https://doi.org/10.1037/a0013895.

James, M., & Seager, J. A. (2006). Impulsivity and schemas for a hostile world: Postdictors of violent behaviour. *International Journal of Offender Therapy and Comparative Criminology, 50*, 47-56. https://doi.org/10.1177/0306624X05279804

Jenness, J. L., Lambert, H. K., Bitrán, D., Blossom, J. B., Nook, E. C., Sasse, S. F., Sommerville, L. H., & McLaughlin, K. A. (2021). Developmental variation in the associations of attention bias to emotion with internalizing and externalizing psychopathology. *Research on child and adolescent psychopathology, 49*(6), 711-726. https://doi.org/10.1007/s10802-020-00751-3

Jolliffe, D., Farrington, D. P., Piquero, A. R., Loeber, R., & Hill, K. G. (2017). Systematic review of early risk factors for life-course-persistent, adolescence-limited, and late-onset offenders in prospective longitudinal studies. *Aggression and violent behavior*, *33*, 15-23. https://doi.org/10.1016%2Fj.avb.2017.01.009

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, *146*(4), 355- 375. https://doi.org/10.1037/bul0000224

Jusyte, A., & Schönenberg, M. (2017). Impaired social cognition in violent offenders: perceptual deficit or cognitive bias?. *European archives of psychiatry and clinical neuroscience*, *267*(3), 257-266. https://doi.org/10.1007/s00406-016-0727-0

Kahn, R. E., Byrd, A. L., & Pardini, D. A. (2013). Callous-unemotional traits robustly predict future criminal offending in young men. *Law and human behavior*, *37*(2), 87- 97.

Kappenman, E. S., Farrens, J. L., Luck, S. J., & Proudfit, G. H. (2014). Behavioral and ERP measures of attentional bias to threat in the dot-probe task: Poor reliability and lack of correlation with anxiety. *Frontiers in psychology*, *5*, 1-9. https://doi.org/10.3389/fpsyg.2014.01368

Kauschke, C., Bahn, D., Vesker, M., & Schwarzer, G. (2019). The role of emotional valence for the processing of facial and verbal stimuli—positivity or negativity bias?. *Frontiers in psychology*, 1654- 1669. https://doi.org/10.3389/fpsyg.2019.01654

Keltner, D., Ellsworth, P. C., & Edwards, K. (1993). Beyond simple pessimism: effects of sadness and anger on social perception. *Journal of personality and social psychology*, *64*(5), 740- 752. https://doi.org/0022-3514/93/S3.00

Kimonis, E. R., Fanti, K. A., Isoma, Z., & Donoghue, K. (2013). Maltreatment profiles among incarcerated boys with callous-unemotional traits. *Child maltreatment*, *18*(2), 108-121. https://doi.org/10.1177/1077559513483002.

Kimonis, E. R., Frick, P. J., Cauffman, E., Goldweber, A., & Skeem, J. (2012). Primary and secondary variants of juvenile psychopathy differ in emotional processing. *Development and psychopathology,* *24*(3), 1091-1103. https://doi.org/10.1017/S0954579412000557

Kimonis, E. R., Frick, P. J., Fazekas, H., & Loney, B. R. (2006). Psychopathy, aggression, and the processing of emotional stimuli in non-referred girls and boys. *Behavioral sciences & the law*, *24*(1), 21-37. https://doi.org/10.1002/bsl.668

Kimonis, E. R., Frick, P. J., Munoz, L. C., & Aucoin, K. J. (2008). Callous-unemotional traits and the emotional processing of distress cues in detained boys: Testing the moderating role of aggression, exposure to community violence, and histories of abuse. *Development and psychopathology*, *20*(2), 569-589. https://doi.org/10.1017/S095457940800028X

Kimonis, E. R., Graham, N., & Cauffman, E. (2018). Aggressive male juvenile offenders with callous-unemotional traits show aberrant attentional orienting to distress cues. *Journal of abnormal child psychology*, *46*(3), 519-527. https://doi.org/10.1007/s10802-017-0295-4

Kimonis, E. R., Kidd, J., Most, S. B., Krynen, A., & Liu, C. (2020). An elusive deficit: Psychopathic personality traits and aberrant attention to emotional stimuli. *Emotion*, *20*(6), 951- 954. https://doi.org/10.1037/emo0000601

King, A. P., Muzik, M., Hamilton, L., Taylor, A. B., Rosenblum, K. L., & Liberzon, I. (2016). Dopamine receptor gene DRD4 7-repeat allele X maternal sensitivity interaction on child externalizing behavior problems: independent replication of effects at 18 months. *PloS one*, *11*(8), e0160473. https://doi.org/10.1371/journal.pone.0160473

Kiyonaga, A., & Egner, T. (2013). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic bulletin & review*, *20*(2), 228-242. https://doi.org/10.3758/s13423-012-0359-y

Klein Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and violent behavior*, *46*, 66-81. https://doi.org/10.1016/j.avb.2019.01.009

Kofler, M. J., Larsen, R., Sarver, D. E., & Tolan, P. H. (2015). Developmental trajectories of aggression, prosocial behavior, and social–cognitive problem solving in emerging adolescents with clinically elevated attention-deficit/hyperactivity disorder symptoms. *Journal of abnormal psychology*, *124*(4), 1027-1042.

Kokko, K., Simonton, S., Dubow, E., Lansford, J. E., Olson, S. L., Huesmann, L. R., Boxer, P., Pulkkinen, L., Bates, J. E., Dodge, K. A., & Pettit, G. S. (2014). Country, sex, and parent occupational status: Moderators of the continuity of aggression from childhood to adulthood. *Aggressive behavior*, *40*(6), 552-567. https://doi.org/10.1002/ab.21546

Kosson, D. S., Lorenz, A. R., & Newman, J. P. (2006). Effects of comorbid psychopathy on criminal offending and emotion processing in male offenders with antisocial personality disorder. *Journal of Abnormal Psychology*, *115*(4), 798- 806. https://doi.org/10.1037/0021-843X.115.4.798

Kosson, D. S., McBride, C. K., Miller, S. A., Riser, N. R., & Whitman, L. A. (2018). Attentional bias following frustration in youth with psychopathic traits: Emotional deficit versus negative preception. *Journal of Experimental Psychopathology*, *9*(2), 1-21. https://doi.org/10.5127/jep.060116

Koster, E. H., Crombez, G., Verschuere, B., & De Houwer, J. (2004). Selective attention to threat in the dot probe paradigm: Differenti- ating vigilance and difficulty to disengage. Behavioral Research Therapy, 42(10), 1183–1192. https://doi.org/10.1016/j.brat.2003.08.001

Kreibig, S. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84,* 394–421. https://doi.org/10.1016/j.biopsy- cho.2010.03.010

Krusemark, E. A., Kiehl, K. A., & Newman, J. P. (2016). Endogenous attention modulates early selective attention in psychopathy: An ERP investigation. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(5), 779-788. https://doi.org/10.3758/s13415-016-0430-7

Kuin, N. C., Masthoff, E. D., Munafò, M. R., & Penton-Voak, I. S. (2017). Perceiving the evil eye: Investigating hostile interpretation of ambiguous facial emotional expression in violent and non-violent offenders. *PloS one*, *12*(11), e0187080. https://doi.org/10.1371/journal.pone.0187080

Kumar, S., Soto, D., & Humphreys, G. W. (2009). Electrophysiological evidence for attentional guidance by the contents of working memory. *European Journal of Neuroscience*, *30*(2), 307-317. https://doi.org/10.1111/j.1460-9568.2009.06805.x

Lam, C. L., Leung, C. J., Yiend, J., & Lee, T. M. (2020). The implication of cognitive processes in emotional bias. *Neuroscience & Biobehavioral Reviews*, *114*, 156-157. https://doi.org/10.1016/j.neubiorev.2020.04.022

Lansford, J. E. (2018). Development of aggression. *Current opinion in psychology*, *19*, 17-21. https://doi.org/10.1016/j.copsyc.2017.03.015

Law, H., & Falkenbach, D. M. (2018). Hostile attribution bias as a mediator of the relationships of psychopathy and narcissism with aggression. *International journal of offender therapy and comparative criminology*, *62*(11), 3355-3371. https://doi.org/10.1177/0306624X17742614

Le Pelley, M. E., Vadillo Nistal, M., & Luque, D. (2013). Learned predictiveness influences rapid attentional capture: Evidence From the Dot Probe Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1888-1900. https://doi.org/10.1037/a0033700

Lee, H. J., and Cho, Y. S. (2019). Memory facilitation for emotional faces: Visual working memory trade-offs resulting from attentional preference for emotional facial expressions. Memory & Cognition, 47, 1231–1243. https://doi.org/10.3758/s13421-019-00930- 8.

Lench, H. C., Bench, S. W., Darbor, K. E., & Moore, M. (2015). A functionalist manifesto: Goal-related emotions from an evolutionary perspective. *Emotion Review*, *7*(1), 90-98. https://doi.org/10.1177/1754073914553001.

Lepsien, J., Griffin, I. C., Devlin, J. T., & Nobre, A. C. (2005). Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *Neuroimage, 26*(3), 733-743 https://doi.org/10.1016/j.neuroimage.2005.02.026

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology, 81*, 146–159. https://doi.org/10.1037/0022-3514.81.1.146

Li, J., Fraser, M. W., & Wike, T. L. (2013). Promoting social competence and preventing childhood aggression: A framework for applying social information processing theory in intervention research. *Aggression and Violent Behavior*, *18*(3), 357-364. https://doi.org/10.1016/j.avb.2013.01.001

Lievaart, M., Franken, I. H., & Hovens, J. E. (2016). Anger assessment in clinical and nonclinical populations: Further validation of the State–Trait Anger Expression Inventory-2. *Journal of clinical psychology*, *72*(3), 263-278. https://doi.org/10.1002/jclp.22253

Lijffijt, M., Cox, B., Acas, M. D., Lane, S. D., Moeller, F. G., & Swann, A. C. (2012). Differential relationships of impulsivity or antisocial symptoms on P50, N100, or P200 auditory sensory gating in controls and antisocial personality disorder. *Journal of psychiatric research*, *46*(6), 743-750. https://doi.org/10.1016/j.jpsychires.2012.03.001

Lobbestael, J., Arntz, A., & Wiers, R. W. (2008). How to push someone's buttons: A comparison of four anger-induction methods. *Cognition & Emotion*, *22*(2), 353-373. https://doi.org/10.1080/02699930701438285

Lobbestael, J., Cima, M., & Arntz, A. (2013). The relationship between adult reactive and proactive aggression, hostile interpretation bias, and antisocial personality disorder. *Journal of personality disorders*, *27*(1), 53-66. https://doi.org/10.1521/pedi.2013.27.1.53

Lobbestael, J., Cousijn, J., Brugman, S., & Wiers, R. W. (2016). Approach and avoidance towards aggressive stimuli and its relation to reactive and proactive aggression. *Psychiatry Research*, *240*, 196-201. https://doi.org/10.1016/j.psychres.2016.04.038

Lobbestael, J., Van Vreeswijk, M., & Arntz, A. (2007). Shedding light on schema modes: A clarification of the mode concept and its current research status. *Netherlands Journal of Psychology, 63*(3), 76–85.

Long, F., Ye, C., Li, Z., Tian, Y., & Liu, Q. (2020). Negative emotional state modulates visual working memory in the late consolidation phase. *Cognition and Emotion*, *34*(8), 1646-1663. https://doi.org/10.1080/02699931.2020.1795626

Loper, A. B. (2003). The relationship of maladaptive beliefs to personality and behavioral adjustment among incarcerated women. *Journal of Cognitive Psychotherapy*, *17*(3), 253-266.

Lotfali, S., Moradi, A., & Ekhtiari, H. (2016). On the Effectiveness of Emotion Regulation Training in Anger Management and Emotional Regulation Difficulties in Adolescents. *Modern Applied Science*, *11*(1), 114- 124. https://doi.org/10.5539/mas.v11n1p114

Lutz, J. (2015). The validity of crowdsourcing data in studying anger and aggressive behavior. *Social Psychology, 47*(1), 38-51. https://doi.org/10.1027/1864-9335/a000256

MacLeod, C., & Holmes, E. A. (2012). Cognitive bias modification: An intervention approach worth attending to. *American Journal of Psychiatry*, *169*(2), 118-120. https://doi.org/10.1176/appi.ajp.2011.11111682

Mado Proverbio, A., Lozano Nasi, V., Alessandra Arcari, L., De Benedetto, F., Guardamagna, M., Gazzola, M., & Zani, A. (2015). The effect of background music on episodic memory and autonomic responses: listening to emotionally touching music enhances facial memory capacity. *Scientific reports*, *5*(1), 1-13. https://doi.org/10.1038/srep15219

Mallett, R., Mummaneni, A., & Lewis-Peacock, J. A. (2020). Distraction biases working memory for faces. *Psychonomic bulletin & review*, *27*(2), 350-356.

Mancke, F., Herpertz, S. C., & Bertsch, K. (2018). Correlates of aggression in personality disorders: an update. *Current Psychiatry Reports, 20*(8), 1-14. https://doi.org/10.1007/s11920-018-0929-4

Mancke, F., Herpertz, S. C., Kleindienst, N., & Bertsch, K. (2017). Emotion dysregulation and trait anger sequentially mediate the association between borderline personality disorder and aggression. *Journal of Personality Disorders*, *31*(2), 256-272. doi:

Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., & Robertson, I. H. (2001). The differential assessment of children's attention: The test of everyday attention for children (TEA-Ch), normative sample and ADHD performance. *Journal of Child Psychology and Psychiatry, 42*(8), 1065–1081.

Manning, K. E. (2020). Seeing red? A systematic review of the evidence for attentional biases to threat-relevant stimuli in propensity to reactive aggression. *Aggression and violent behavior*, *50*, 101359. https://doi.org/10.1016/j.avb.2019.101359.

Maoz, K., Adler, A. B., Bliese, P. D., Sipos, M. L., Quartana, P. J., & Bar-Haim, Y. (2017). Attention and interpretation processes and trait anger experience, expression, and control. *Cognition and Emotion*, *31*(7), 1453-1464. https://doi.org/10.1080/02699931.2016.1231663

Maran, T., Sachse, P., & Furtner, M. (2015). From specificity to sensitivity: affective states modulate visual working memory for emotional expressive faces. *Frontiers in Psychology*, *6*, 1297- 1313. https://doi.org/10.3389/fpsyg.2015.01297

Martinelli, A., Ackermann, K., Bernhard, A., Freitag, C. M., & Schwenck, C. (2018). Hostile attribution bias and aggression in children and adolescents: A systematic literature review on the influence of aggression subtype and gender. *Aggression and violent behavior*, *39*, 25-32. https://doi.org/10.1016/j.avb.2018.01.005

Massah, O., Sohrabi, F., A'azami, Y., Doostian, Y., Farhoudian, A., & Daneshmand, R. (2016). Effectiveness of gross model-based emotion regulation strategies training on anger reduction in drug-dependent individuals and its sustainability in follow-up. *International journal of high risk behaviors & addiction*, *5*(1), 1-6. https://doi.org/10.5812/ijhrba.24327

Mathys, C. D., Lomakina, E. I., Daunizeau, J., Iglesias, S., Brodersen, K. H., Friston, K. J., & Stephan, K. E. (2014). Uncertainty in perception and the Hierarchical Gaussian Filter. *Frontiers in Human Neuroscience, 8*. http://dx.doi.org/10.3389/fnhum.2014.00825

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience*, *5*, 39. https://doi.org/10.3389/fnhum.2011.00039

Matjasko, J. L., Vivolo-Kantor, A. M., Massetti, G. M., Holland, K. M., Holt, M. K., & Cruz, J. D. (2012). A systematic meta-review of evaluations of youth violence prevention programs: Common and divergent findings from 25 years of meta-analyses and systematic reviews. *Aggression and Violent Behavior*, *17*(6), 540-552. https://doi.org/10.1016/j.avb.2012.06.006.

Matsumoto, D., & Ekman, P. (1989). *The Japanese and Causacian facial expressions of emotion (JACFEE) and neutrals (JACNeuF).* San Francisco State University, Department of Psychology, Intercultural and Emotion Research Laboratory.

Mathews, A. & MacLeod, C. (2005). Cognitive Vulnerability to Emotional Disorders. *Annual Review of Clinical Psychology*, 1, 167-195. http://dx.doi.org/10.1146/annurev.clinpsy.1.102803.143916

Mazza, V., Dallabona, M., Chelazzi, L., & Turatto, M. (2011). Cooperative and opposing effects of strategic and involuntary attention. *Journal of cognitive neuroscience*, *23*(10), 2838-2851.

McDonald, K. L. and Lochman, J. E. (2012). Predictors and outcomes associated with trajectories of revenge goals from fourth grade through seventh grade. *Journal of abnormal child psychology*, *40*(2), 225-236. https://doi.org/10.1007/s10802-011-9560-0

Mellentin, A. I., Dervisevic, A., Stenager, E., Pilegaard, M., & Kirk, U. (2015). Seeing enemies? A systematic review of anger bias in the perception of facial expressions among anger-prone and aggressive populations. *Aggression and violent behavior*, *25*, 373-383. https://doi.org/10.1016/j.avb.2015.09.001

Meloy, J. R., Hempel, A. G., Mohandie, K., Shiva, A. A., & Gray, B. (2001). Offender and offense characteristics of a nonrandom sample of adolescent mass murderers. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(6), 719–728.

Merrill, K. L., Smith, S. W., Cumming, M. M., & Daunic, A. P. (2017). A review of social problem-solving interventions: Past findings, current status, and future directions. *Review of Educational Research, 87*(1), 71-102. https://doi.org/10.3102/0034654316652943

Metropolitan Police (2022). *What is antisocial behaviour?.* https://www.met.police.uk/advice/advice-and-information/asb/asb/antisocial-behaviour/what-is-antisocial-behaviour/

Mikels, J. A., & Reuter-Lorenz, P. A. (2019). Affective working memory: An integrative psychological construct. *Perspectives on Psychological Science*, *14*(4), 543-559. https://doi.org/10.1177/1745691619837597

Miller, N. V., & Johnston, C. (2019). Social threat attentional bias in childhood: Relations to aggression and hostile intent attributions. *Aggressive behavior*, *45*(3), 245-254. https://doi.org/10.1002/ab.21813

Millisecond Software, LLC. (2019). Inquisit 5 [Computer software]. (2016). https://www.millisecond.com.

Milner, R. J., & Webster, S. D. (2005). Identifying schemas in child molesters, rapist, and violent offenders. Sexual Abuse: *A Journal of Research and Treatment, 17*, 425-439. https://doi.org/10.1177/107906320501700406

Miloff, A., Savva, A., & Carlbring, P. (2015). Cognitive bias measurement and social anxiety disorder: Correlating self-report data and attentional bias. *Internet Interventions*, *2*(3), 227-234. https://doi.org/10.1016/j.invent.2015.03.006

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674-701.

Moons, W. G., & Mackie, D. M. (2007). Thinking straight while seeing red: The influence of anger on information processing. *Personality and Social Psychology Bulletin*, *33*(5), 706-720. https://doi.org/10.1177/0146167206298566.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119-124. https://doi.org/10.1177/1754073912468165

Morales, S., Pérez-Edgar, K., & Buss, K. (2016). Longitudinal relations among exuberance, externalizing behaviors, and attentional bias to reward: The mediating role of effortful control. *Developmental Science, 19*, 853–862. https://doi.org/10.1111/desc.12320.

Moriya, J., Koster, E. H., & De Raedt, R. (2014). The influence of working memory on the anger superiority effect. *Cognition and Emotion*, *28*(8), 1449-1464. https://doi.org/10.1080/02699931.2014.890094

Muftić, L. R., Grubb, J. A., Bouffard, L. A., & Maljević, A. (2014). The impact of life domains on juvenile offending in Bosnia and Herzegovina: Direct, indirect, and moderating effects in Agnew's integrated general theory. *Journal of Research in Crime and Delinquency*, *51*(6), 816-845. https://doi.org/10.1177/0022427814546198

Muñoz Centifanti, L. C. M., Kimonis, E. R., Frick, P. J., & Aucoin, K. J. (2013). Emotional reactivity and the association between psychopathy-linked narcissism and aggression in detained adolescent boys. *Development and psychopathology*, *25*(2), 473-485. https://doi.org/10.1017/S0954579412001186

Muñoz, L. C., Frick, P. J., Kimonis, E. R., & Aucoin, K. J. (2008). Types of aggression, responsiveness to provocation, and callous-unemotional traits in detained adolescents. *Journal of abnormal child psychology*, *36*(1), 15-28. https://doi.org/10.1007/s10802-007-9137-0

Murray-Close, D., Ostrov, J. M., Nelson, D. A., Crick, N. R., & Coccaro, E. F. (2010). Proactive, reactive, and romantic relational aggression in adulthood: Measurement, predictive validity, gender differences, and association with intermittent explosive disorder. *Journal of Psychiatric Research*, *44*(6), 393-404. https://doi.org/10.1016/j.jpsychires.2009.09.005

Nesset, M. B., Bjørngaard, J. H., Whittington, R., & Palmstierna, T. (2021). Does cognitive behavioural therapy or mindfulness-based therapy improve mental health and emotion regulation among men who perpetrate intimate partner violence? A randomised controlled trial. *International journal of nursing studies*, *113*, 103795. https://doi.org/10.1016/j.ijnurstu.2020.103795

Neumann, A., van Lier, P. A., Gratz, K. L., & Koot, H. M. (2010). Multidimensional assessment of emotion regulation difficulties in adolescents using the difficulties in emotion regulation scale. *Assessment*, *17*(1), 138-149. https://doi.org/10.1177/1073191109349579

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. https://doi.org/10.1016/j.biopsych.2009.07.035.

Norris, C. J. (2021). The negativity bias, revisited: Evidence from neuroscience measures and an individual differences approach. *Social neuroscience*, *16*(1), 68-82. https://doi.org/10.1080/17470919.2019.1696225

Novaco, R. W., & Taylor, J. L. (2015). Reduction of assaultive behavior following anger treatment of forensic hospital patients with intellectual disabilities. *Behaviour Research and Therapy*, *65*, 52-59. https://doi.org/10.1016/j.brat.2014.12.001

O'Reilly, K., Donohoe, G., Coyle, C., O'Sullivan, D., Rowe, A., Losty, M., McDonagh, T., McGuiness, L., Ennis, Y., Brennan, L., Owens, E., Davoren, M., Mullaney, R., Abidin, Z., & Kennedy, H. G. (2015). Prospective cohort study of the relationship between neuro-cognition, social cognition and violence in forensic patients with schizophrenia and schizoaffective disorder. *BMC psychiatry*, *15*(1), 1-17. https://doi.org/10.1186/s12888-015-0548-0

Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (45–100). San Diego: Academic Press. https://doi.org/10.1016/S0079-7421(09)51002-X

Ogilvie, J. M., Stewart, A. L., Chan, R. C., & Shum, D. H. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology*, *49*(4), 1063-1107. https://doi.org/10.1111/j.1745-9125.2011.00252.x

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological review*, *108*(3), 483- 522. https://doi.org/10.1037//0033-295X.108.3.483

Olivers, C. N. (2008). Interactions between visual working memory and visual attention. *Frontiers in Bioscience*, *13*(3), 1182-1191.

Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in cognitive sciences*, *15*(7), 327-334. https://doi.org/10.1016/j.tics.2011.05.004

Oostermeijer, S., Nieuwenhuijzen, M., Van de Ven, P. M., Popma, A., & Jansen, L. M. C. (2016). Social information processing problems related to reactive and proactive aggression of adolescents in residential treatment. *Personality and Individual Differences, 90*, 54-60. https://doi.org/10.1016/j.paid.2015.10.035

Opitz, P. C., Gross, J. J., & Urry, H. L. (2012). Selection, optimization, and compensation in the domain of emotion regulation: Applications to adolescence, older age, and major depressive disorder. *Social and Personality Psychology Compass, 6*(2), 142–155. https://doi.org/10.1111/j.1751-9004.2011.00413.x

Orue, I., Calvete, E., & Fernández-González, L. (2021). Early maladaptive schemas and social information processing in child-to-parent aggression. *Journal of interpersonal violence*, *36*(15-16), 6931-6955. https://doi.org/10.1177/0886260519831395

Osgood, J. M. (2017). Is revenge about retributive justice, deterring harm, or both? *Social and Personality Psychology Compass, 11*(1), e12296.

Osgood, J. M., Kase, S. E., Zaroukian, E. G., & Quartana, P. J. (2021). Online intervention reduces hostile attribution bias, anger, aggressive driving, and cyber-aggression, results of two randomized trials. *Cognitive Therapy and Research*, *45*(2), 310-321. https://doi.org/10.1007/s10608-020-10147-8

Owen, J. M. (2011). Transdiagnostic cognitive processes in high trait anger. *Clinical Psychology Review, 31*(2), 193–202. https://doi.org/10. 1016/j.cpr.2010.10.003

Pan, Y., & Soto, D. (2010). The modulation of perceptual selection by working memory is dependent on the focus of spatial attention. *Vision Research*, *50*(15), 1437-1444. https://doi.org/10.1016/j.visres.2009.10.016

Pan, Y., Lin, B., Zhao, Y., & Soto, D. (2014). Working memory biasing of visual perception without awareness. *Attention, Perception, & Psychophysics, 76*(7), 2051-2062. https://doi.org/10.3758/s13414-013-0566-2

Pasion, R., Fernandes, C., Pereira, M. R., & Barbosa, F. (2018). Antisocial behaviour and psychopathy: Uncovering the externalizing link in the P3 modulation. *Neuroscience & Biobehavioral Reviews*, *91*, 170-186. https://doi.org/10.1016/j.neubiorev.2017.03.012

Peacock, A., Leung, J., Larney, S., Colledge, S., Hickman, M., Rehm, J., Giovino, G. A., West, R., Hall, W., Griffiths, P., Ali, R., Gowing, L., Marsden, J., Ferrari, A. J., Grebely, J., Farrell, M., & Degenhardt, L. (2018). Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. *Addiction, 113*(10), 1905-1926. https://doi.org/10.1111/add.14234

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in cognitive sciences*, *18*(5), 242-250. https://doi.org/10.1016/j.tics.2014.02.004

Peng, X., Zhang, M., Zhang, D., Xie, D., & Guan, Y. (2015). From anger to attribution: Could attention be a bridge?. *Social Behavior and Personality: an international journal*, *43*(3), 505-518. https://doi.org/10.2224/sbp.2015.43.3.505

Pfabigan, D. M., Alexopoulos, J., & Sailer, U. (2012). Exploring the effects of antisocial personality traits on brain potentials during face processing. *PLoS One*, *7*(11), e50283. https://doi.org/10.1371/journal.pone.0050283

Pinkham, A. E., Griffin, M., Baron, R., Sasson, N. J., & Gur, R. C. (2010). The face in the crowd effect: Anger superiority when using real faces and multiple identities. *Emotion, 10*(1), 141–146. https://doi.org/10.1037/a0017387.

Poldrack, R. A., Monahan, J., Imrey, P. B., Reyna, V., Raichle, M. E., Faigman, D., & Buckholtz, J. W. (2017). Predicting violent behavior: what can neuroscience add? *Trends in cognitive sciences, 22*(2), 111-123.

Postle, B.R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*, 23–38.

Preston, O. C., & Anestis, J. C. (2020). The indirect relationships between psychopathic traits and proactive and reactive aggression through empathy and emotion dysregulation. *Journal of Psychopathology and Behavioral Assessment*, *42*(3), 409-423. https://doi.org/10.1007/s10862-019-09760-z

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., Dahl, R. E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological assessment*, *27*(2), 365- 376. https://doi.org/10.1037/pas0000036

Puls, S., & Rothermund, K. (2018). Attending to emotional expressions: no evidence for automatic capture in the dot-probe task. *Cognition and Emotion*, *32*(3), 450-463. https://doi.org/10.1080/02699931.2017.1314932

Putman, P., Hermans, E., & van Honk, J. (2004). Emotional stroop performance for masked angry faces: It'sBAS, not BIS. *Emotion, 4*(3), 305–311.

Qualtrics (2019). Provo: Utah. https://www.qualtrics.com/

Quan, F., Yang, R., Zhu, W., Wang, Y., Gong, X., Chen, Y., Dong, Y., & Xia, L. X. (2019a). The relationship between hostile attribution bias and aggression and the mediating effect of anger rumination. *Personality and Individual Differences*, *139*, 228-234. https://doi.org/10.1016/j.paid.2018.11.029

Quan, F., Zhu, W., Dong, Y., Qiu, J., Gong, X., Xiao, M., Zheng, Y., Zhao, Y., Chen, X., & Xia, L. X. (2019b). Brain structure links trait hostile attribution bias and attitudes toward violence. *Neuropsychologia, 125,* 42-50. https://doi.org/10.1016/j.neuropsychologia.2019.01.015

Quarto, T., Blasi, G., Pallesen, K. J., Bertolino, A., & Brattico, E. (2014). Implicit processing of visual emotions is affected by sound-induced affective states and individual affective traits. *PLoS One*, *9*(7), e103278. https://doi.org/10.1371/journal.pone.0103278

Quoidbach, J., Mikolajczak, M., & Gross, J. J. (2015). Positive interventions: An emotion regulation perspective. *Psychological bulletin*, *141*(3), 655- 693. https://doi.org/10.1037/a0038648

Rabiner, D. L., Lenhart, L., & Lochman, J. E. (1990). Automatic versus reflective social problem solving in relation to children's sociometric status. *Developmental Psychology*, *26*(6), 1010-1016, https://doi.org/0012-1649/90/$00.75

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *32*(2), 159-171. https://doi.org/10.1002/ab.20115

Reidy, D.E., Foster, J.D. and Zeichner, A. (2010). Narcissism and unprovoked aggression. *Aggressive Behavior*, *36*(6), 414-422. https://doi.org/10.1002/ab.20356

Repple, J., Pawliczek, C. M., Voss, B., Siegel, S., Schneider, F., Kohn, N., & Habel, U. (2017). From provocation to aggression: the neural network. *BMC neuroscience*, *18*(1), 1-9. https://doi.org/10.1186/s12868-017-0390-z

Ribeiro, F. S., Santos, F. H., & Albuquerque, P. B. (2019). How does allocation of emotional stimuli impact working memory tasks? An overview. *Advances in Cognitive Psychology*, *15*(2), 155- 168. https://doi.org/10.5709/acp-0265-y

Richey, A., Brown, S., Fite, P. J., & Bortolato, M. (2016). The role of hostile attributions in the associations between child maltreatment and reactive and proactive aggression. *Journal of Aggression, Maltreatment & Trauma*, *25*(10), 1043-1057. https://doi.org/10.1080/10926771.2016.1231148.

Rissanen, E., Kuvaja-Köllner, V., Elonheimo, H., Sillanmäki, L., Sourander, A., & Kankaanpää, E. (2021). The long-term cost of childhood conduct problems: Finnish Nationwide 1981 Birth Cohort Study. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13506

Roberton, T., Daffern, M., & Bucks, R. S. (2012). Emotion regulation and aggression. *Aggression and violent behavior*, *17*(1), 72-82. https://doi.org/10.1016/j.avb.2011.09.006

Rodebaugh, T. L., Scullin, R. B., Langer, J. K., Dixon, D. J., Huppert, J. D., Bernstein, A., Zvielli, A., & Lenze, E. J. (2016). Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *Journal of abnormal psychology*, *125*(6), 840- 851. https://doi.org/10.1037/abn0000184.

Romeo, R., Knapp, M., & Scott, S. (2006). Economic cost of severe antisocial behaviour in children-and who pays it. *The British Journal of Psychiatry*, *188*(6), 547-553. https://doi.org/10.1007/s00787-008-0686-8

Romero, E., Richards, M. H., Harrison, P. R., Garbarino, J., & Mozley, M. (2015). The role of neighborhood in the development of aggression in urban African American youth: A multilevel analysis. *American journal of community psychology*, *56*(1), 156-169.

RStudio Team (2018). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA http://www.rstudio.com/

RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA http://www.rstudio.com/

Rudo-Hutt, A. S. (2015). Electroencephalography and externalizing behavior: a meta-analysis. *Biological psychology*, *105*, 1-19. https://doi.org/10.1016/j.biopsycho.2014.12.005

Sadek, S. A., Daniel, M. R., & Langdon, P. E. (2021). Attentional bias toward negative and positive pictorial stimuli and its relationship with distorted cognitions, empathy, and moral reasoning among men with intellectual disabilities who have committed crimes. *Aggressive Behavior*, *47*(1), 120-130. https://doi.org/10.1002/ab.21908

Salahub, C., & Emrich, S. M. (2020). Fear not! Anxiety biases attentional enhancement of threat without impairing working memory filtering. *Cognitive, Affective, & Behavioral Neuroscience*, *20*(6), 1248-1260. https://doi.org/10.3758/s13415-020-00831-3

Santone, E., Crothers, L. M., Kolbert, J. B., & Miravalle, J. (2020). Using Social Information Processing Theory to Counsel Aggressive Youth. *Journal of School Counseling*, *18*(18), n18.

Satmarean, T. S., Milne, E., & Rowe, R. (2022). Working memory guidance of visual attention to threat in offenders. *PloS One*, *17*(1), e0261882. https://doi.org/10.1371/journal.pone.0261882

Schippell, P. L., Vasey, M.W., Cravens-Brown, L. M., & Bretveld, R. A. (2003). Suppressed attention to rejection, ridicule, and failure cues: a unique correlate of reactive but not proactive aggression in youth. *Journal of Clinical Child and Adolescent Psychology, 32*, 40–55. https://doi.org/10.1207/S15374424JCCP3201_05.

Schlaghecken, F., Blagrove, E., Mantantzis, K., Maylor, E. A., and Watson, D. G. (2017). Look on the bright side: positivity bias modulates interference effects in the simon task. Journal of Experimental Psychology, 146, 763–770. https://doi.org/10.1037/xge0000316

Schmukle, S. C. (2005). Unreliability of the dot probe task. *European Journal of Personality*, *19*(7), 595-605. https://doi.org/10.1002/per.554

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European archives of psychiatry and clinical neuroscience*, *264*(1), 61-69.

Schönenberg, M., Louis, K., Mayer, S., & Jusyte, A. (2013). Impaired identification of threat-related social information in male delinquents with antisocial personality disorder. *Journal of personality disorders*, *27*(4), 496-505. https://doi.org/10.1521/pedi_2013_27_100

Schwenck, C., Gensthaler, A., Romanos, M., Freitag, C. M., Schneider, W., & Taurines (2014). Emotion recognition in girls with conduct problems. *European Child and Adolescent Psychiatry, 23*,13–22.

Searight, H. R., Rottnek, F., & Abby, S. (2001). Conduct disorder: Diagnosis and treatment in primary care. *American Family Physician*, *15;63(8):1579-1589.*

Seidl-Rathkopf, K. N., Turk-Browne, N. B., & Kastner, S. (2015). Automatic guidance of attention during real-world visual search. *Attention, Perception, & Psychophysics*, *77*(6), 1881-1895. https://doi.org/10.3758/s13414-015-0903-8

Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. *Journal of Neuroscience, 32*(34), 11812-11819. https://doi.org/10.1523/JNEUROSCI.1693-12.2012

Sell, A., Cosmides, L., & Tooby, J. (2014). The human anger face evolved to enhance cues of strength. *Evolution and Human Behavior, 35*(5), 425–429. https://doi.org/10.1016/ j.evolhumbehav.2014.05.008.

Semmler, C., & Hurst, J. (2017). The impact of state and trait anger on processing of evidential inconsistencies. *Psychiatry, Psychology and Law*, *24*(4), 594-604. https://doi.org/10.1080/13218719.2016.1258686

Shasteen, J. R., Sasson, N. J., & Pinkham, A. E. (2015). A detection advantage for facial threat in the absence of anger. *Emotion*, *15*(6), 837–845. https://doi.org/10.1037/emo0000090

Shields, G. S., Moons, W. G., Tewell, C. A., & Yonelinas, A. P. (2016). The effect of negative affect on cognition: Anxiety, not anger, impairs executive function. *Emotion*, *16*(6), 792- 797. https://doi.org/10.1037/emo0000151

Simione, L., Calabrese, L., Marucci, F. S., Belardinelli, M. O., Raffone, A., and Maratos, F. A. (2014). Emotion based attentional priority for storage in visual short-term memory. *PLoS One, 9*. https://doi.org/10.1371/journal.pone.0095261.

Simons, J. S., Wills, T. A., Emery, N. N., & Spelman, P. J. (2016). Keep calm and carry on: Maintaining self-control when intoxicated, upset, or depleted. *Cognition and emotion*, *30*(8), 1415-1429. https://doi.org/10.1080/02699931.2015.1069733

Skeem, J. L., Poythress, N. G., Edens, J. F., Lilienfeld, S. O., & Cale, E. M. (2003). Psychopathic personality or personalities? Exploring potential variants of psychopathy and their implications for risk assessment. *Aggression and Violent Behavior, 8,* 513–546.

Smeijers D, Bulten EB, Brazil IA. The Computations of hostile biases (CHB) model: Grounding hostility biases in a unified cognitive framework. Clinical psychology review. 2019 Nov 1;73:101775. doi: 10.1016/j.cpr.2019.101775.

Smeijers, D., Benbouriche, M., & Garofalo, C. (2020). The association between emotion, social information processing, and aggressive behavior: A systematic review. *European Psychologist*. 25(2), 81-91. https://doi.org/10.1027/1016-9040/a000395

Smeijers, D., Rinck, M., Bulten, E., van den Heuvel, T., & Verkes, R. J. (2017). Generalized hostile interpretation bias regarding facial expressions: Characteristic of pathological aggressive behavior. Aggressive behavior, 43(4), 386-397.

Smith, P., & Waterman, M. (2003). Processing bias for aggression words in forensic and nonforensic samples. *Cognition & Emotion*, *17*(5), 681-701. https://doi.org/10.1080/02699930302281

Smith, P., & Waterman, M. (2004). Role of experience in processing bias for aggressive words in forensic and non-forensic populations. Aggressive Behavior: *Official Journal of the International Society for Research on Aggression, 30*(2), 105-122. https://doi.org/10.1002/ab.20001.

Soto, D., & Humphreys, G. W. (2007). Automatic guidance of visual attention from verbal working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 730- 757. https://doi.org/10.1037/0096-1523.33.3.730

Soto, D., & Humphreys, G. W. (2008). Stressing the mind: The effect of cognitive load and articulatory suppression on attentional guidance from working memory. *Perception & Psychophysics*, *70*(5), 924-934. doi: 10.3758/PP.70.5.924

Soto, D., Greene, C. M., Kiyonaga, A., Rosenthal, C. R., & Egner, T. (2012). A parieto-medial temporal pathway for the strategic control over working memory biases in human visual attention. *Journal of Neuroscience, 32*(49), 17563-17571. https://doi.org/10.1523/JNEUROSCI.2647-12.2012

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248- 261. https://doi.org/10.1037/0096-1523.31.2.248

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12*(9), 342–348. https://doi.org/10.1016/j.tics.2008.05.007

Soto, D., Humphreys, G. W., & Rotshtein, P. (2007). Dissociating the neural mechanisms of memory-based guidance of visual selection. *Proceedings of the National Academy of Sciences, 104*(43), 17186-17191. doi: 10.1073/pnas.0703706104

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, & Psychophysics*, *78*(7), 1839-1860. https://doi.org/10.3758/s13414-016-1108-5

Souza, A. S., Thaler, T., Liesefeld, H. R., Santos, F. H., Peixoto, D. S., & Albuquerque, P. B. (2021). No evidence that self-rated negative emotion boosts visual working memory precision. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(2), 282- 307. https://doi.org/10.1037/xhp0000891

Spachtholz, P., Kuhbandner, C., & Pekrun, R. (2014). Negative affect improves the quality of memories: trading capacity for precision in sensory and working memory. *Journal of Experimental Psychology: General*, *143*(4), 1450- 1465. https://doi.org/10.1037/xge0000012

Spielberger, C. D. (1999). *State-Trait Anger Expression Inventory-2*. Lutz, Florida: Psychological Assessment Resources

Staugaard, S. R. (2009). Reliability of two versions of the dot-probe task using photographic faces. *Psychology Science Quarterly*, *51*(3), 339-350.

Stavraki, M., Lamprinakos, G., Brinol, P., Petty, R. E., Karantinou, K., & Díaz, D. (2021). The influence of emotions on information processing and persuasion: A differential appraisals perspective. *Journal of experimental social psychology*, *93*, 104085. https://doi.org/10.1016/j.jesp.2020.104085

Strickland, C. M., Drislane, L. E., Lucy, M., Krueger, R. F., & Patrick, C. J. (2013). Characterizing psychopathy using DSM–5 personality traits. *Assessment, 20*, 327–338. http://dx.doi.org/10.1177/10731 91113486691

Szasz, P. L., Szentagotai, A., & Hofmann, S. G. (2011). The effect of emotion regulation strategies on anger. *Behaviour research and therapy, 49*(2), 114-119. https://doi.org/10.1016/j.brat.2010.11.011

Szymaniak, K., & Zajenkowski, M. (2021). How do high trait anger people feel about rewards high and low in arousal? Disentangling the association between trait anger and subjective pleasantness of rewards. *Personality and Individual Differences*, *168*, 110278. https://doi.org/10.1016/j.paid.2020.110278

Tager, D., Good, G. E., & Brammer, S. (2010). “Walking over'em”: An exploration of relations between emotion dysregulation, masculine norms, and intimate partner abuse in a clinical sample of men. *Psychology of Men & Masculinity, 11*(3), 233–603. https://doi.org/10. 1037/a0017636

Taylor, A. J. G., & Jose, M. (2014). Physical aggression and facial expression identification. *Europe's Journal of Psychology*, *10*(4), 650-659. https://doi.org/10.5964/ejop.v10i4.816

Teglasi, H., Caputo, M. H., & Scott, A. L. (2022). Explicit and implicit theory of mind and social competence: A social information processing framework. *New Ideas in Psychology*, *64*, 100915.

Thigpen, N. N., Gruss, L. F., Garcia, S., Herring, D. R., & Keil, A. (2018). What does the dot-probe task measure? A reverse correlation analysis of electrocortical activity. *Psychophysiology*, *55*(6), e13058. https://doi.org/10.1111/psyp.13058

Thompson, R. J., & Berenbaum, H. (2011). Adaptive and aggressive assertiveness scales (AAA-S). *Journal of Psychopathology and Behavioral Assessment*, *33*(3), 323-334. https://doi.org/10.1007/s10862-011-9226-9

Thomson, N. D., & Centifanti, L. C. M. (2018). Proactive and reactive aggression subgroups in typically developing children: The role of executive functioning, psychophysiology, and psychopathy. *Child Psychiatry & Human Development, 49*(2), 197–208. https://doi.org/10.1007/s10578-017-0741-0

Torrence, R. D., & Troup, L. J. (2018). Event-related potentials of attentional bias toward faces in the dot-probe task: A systematic review. *Psychophysiology*, *55*(6), e13051. https://doi.org/10.1111/psyp.13051

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168*(3), 242–249. http://dx.doi.org/10.1016/j. psychres.2008.05.006.

Tremblay, P. F., & Dozois, D. J. A. (2009). Another perspective on trait aggressiveness: Overlap with early maladaptive schemas. *Personality and Individual Differences, 46*(5-6), 569–574. https://doi.org/10.1016/j.paid.2008.12.009

Trevors, G. J., Kendeou, P., & Butterfuss, R. (2017). Emotion processes in knowledge revision. *Discourse Processes*, *54*(5-6), 406-426. https://doi.org/10.1080/0163853X.2017.1312201

Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and violent behavior*, *46*, 66-81.

Van Bockstaele, B., van der Molen, M. J., van Nieuwenhuijzen, M., & Salemink, E. (2020). Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents. *Journal of experimental child psychology*, *194*, 104811. https://doi.org/10.1016/j.jecp.2020.104811

Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological bulletin*, *140*(3), 682- 721. https://doi.org/10.1037/a0034834

Van Goozen, S. H. M., Cohen-Kettenis, P. T., Matthys, W., & van Engeland, H. (2002). Preference for aggressive and sexual stimuli in children with disruptive behavior disorder and normal controls. *Archives of Sexual Behavior, 31*, 247–253.

Van Honk, J., & Schutter, D. (2007). Vigilant and avoidant responses to angry facial expressions. *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior,* 197–223.

Van Leeuwen, N., Rodgers, R. F., Gibbs, J. C., & Chabrol, H. (2014). Callous-unemotional traits and antisocial behavior among adolescents: The role of self-serving cognitions. *Journal of abnormal child psychology*, *42*(2), 229-237.

Van Moorselaar, D., Theeuwes, J., & Olivers, C. N. (2014). In competition for the attentional template: Can multiple items within visual working memory guide attention?. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1450-1464. https://doi.org/10.1037/a0036229

Van Nieuwenhuijzen, M., Van Rest, M. M., Embregts, P. J. C. M., Vriens, A., Oostermeijer, S., Van Bokhoven, I., & Matthys, W. (2017). Executive functions and social information processing in adolescents with severe behavior problems. *Child neuropsychology*, *23*(2), 228-241. https://doi.org/10.1080/09297049.2015.1108396

van Reemst, L., Fischer, T. F., & Zwirs, B. W. (2016). Social information processing mechanisms and victimization: A literature review. *Trauma, Violence, & Abuse*, *17*(1), 3-25. https://doi.org/10.1177/1524838014557286

Van Rooijen, R., Ploeger, A., & Kret, M. E. (2017). The dot-probe task to measure emotional attention: A suitable measure in comparative studies?. *Psychonomic bulletin & review, 24*(6), 1686-1717. https://doi.org/10.3758/s13423-016-1224-1

Van Steenbergen, H., Band, G. P., & Hommel, B. (2011). Threat but not arousal narrows attention: evidence from pupil dilation and saccade control. *Frontiers in psychology*, *2*, 281- 286. https://doi.org/10.3389/fpsyg.2011.00281

Van Wijk-Herbrink, M. F., Lobbestael, J., Bernstein, D. P., Broers, N. J., Roelofs, J., & Arntz, A. (2021). The influence of early maladaptive schemas on the causal links between perceived injustice, negative affect, and aggression. *International Journal of Forensic Mental Health*, *20*(2), 133-149. https://doi.org/10.1080/14999013.2020.1842562

Vaske, J., Galyean, K., & Cullen, F. T. (2011). Toward a biosocial theory of offender rehabilitation: Why does cognitive-behavioral therapy work?. *Journal of Criminal Justice*, *39*(1), 90-102. https://doi.org/10.1016/j.jcrimjus.2010.12.006

Veenstra, L., Bushman, B. J., & Koole, S. L. (2018). The facts on the furious: a brief review of the psychology of trait anger. *Current Opinion in Psychology*, *19*, 98-103. https://doi.org/10.1016/j.copsyc.2017.03.014

Veenstra, L., Schneider, I. K., & Koole, S. L. (2017). The effects of motivational training on state anger and aggressive impulses among people varying in trait anger. *Motivation Science*, *3*(4), 354- 368. https://doi.org/10.1037/mot0000060

Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy, 6,* 473–482.

Verhoef, R. E., van Dijk, A., & de Castro, B. O. (2021). A Dual-Mode Social-Information-Processing Model to Explain Individual Differences in Children's Aggressive Behavior. *Clinical Psychological Science*, 1-17. https://doi.org/10.1177/21677026211016396

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron*, *30*(3), 829-841. https://doi.org/10.1016/S0896-6273(01)00328-2

Wadlinger, H. A., & Isaacowitz, D. M. (2006). Positive mood broadens visual attention to positive stimuli. *Motivation and Emotion, 30*(1), 87–99. https://doi.org/10.1007/s11031-006-9021-1

Wang, X., Yang, L., Yang, J., Gao, L., Zhao, F., Xie, X., & Lei, L. (2018). Trait anger and aggression: A moderated mediation model of anger rumination and moral disengagement. *Personality and Individual Differences*, *125*, 44-49. https://doi.org/10.1016/j.paid.2017.12.029

Wang, Y., Cao, S., Dong, Y., & Xia, L. X. (2019). Hostile attribution bias and angry rumination: A longitudinal study of undergraduate students. *PloS one*, *14*(5), e0217759. https://doi.org/10.1371/journal.pone.0217759

Wang, Y., Zhu, W., Xiao, M., Zhang, Q., Zhao, Y., Zhang, H., Chen, X., Zheng, Y., & Xia, L. X. (2018). Hostile attribution bias mediates the relationship between structural variations in the left middle frontal gyrus and trait angry rumination. *Frontiers in psychology*, *9*, 526. https://doi.org/10.3389/fpsyg.2018.00526

Ward, T. (2000). Sexual offenders' cognitive distortions as implicit theories. *Aggression and Violent Behavior, 5*, 491–507. https://doi.org/10.1016/S1359-1789(98)00036-6

Waschbusch, D., Carrey, N., Willoughby, M., King, S., & Andrade, B. (2007). Effects of methylphenidate and behavior modification on the social and academic behavior of children with disruptive behavior disorders: The moderating role of callous/ unemotional traits. *Journal of Clinical Child and Adolescent Psychology, 36*, 629–644.

Wenze, S. J., Gunthert, K. C., & German, R. E. (2012). Biases in affective forecasting and recall in individuals with depression and anxiety symptoms. *Personality and Social Psychology Bulletin, 38,* 895–906. http://dx.doi.org/10.1177/0146167212447242

Wenzel, A., & Lystad, C. (2005). Interpretation biases in angry and anxious individuals. *Behaviour Research and Therapy, 43*(8), 1045–1054. https://doi.org/10.1016/ j.brat.2004.02.009

White, B. & Turner, K. A. (2014). Anger rumination and effortful control: Mediation effects of reactive but not proactive aggression. *Personality and Individual Differences, 56*, 186- 189. https://doi.org/10.1016/j.paid.2013.08.012

White, B. A., Gordon, H., & Guerra, R. C. (2015). Callous–unemotional traits and empathy in proactive and reactive relational aggression in young women. *Personality and Individual Differences*, *75*, 185-189. https://doi.org/10.1016/j.paid.2014.11.031

Wieser, M. J., Hambach, A., & Weymar, M. (2018). Neurophysiological correlates of attentional bias for emotional faces in socially anxious individuals–Evidence from a visual search task and N2pc. *Biological psychology*, *132*, 192-201. https://doi.org/10.1016/j.biopsycho.2018.01.004

Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and social psychology review*, *12*(1), 3-21. https://doi.org/10.1177/1088868307309874

Wilkowski, B. M., & Robinson, M. D. (2010). The anatomy of anger: An integrative cognitive model of trait anger and reactive aggression. *Journal of personality*, *78*(1), 9-38. https://doi.org/10.1111/j.1467-6494.2009.00607.x.

Wilkowski, B. M., Robinson, M. D., & Meier, B. P. (2006). Agreeableness and the prolonged spatial processing of antisocial and prosocial information. *Journal of Research in Personality, 40*, 1152–1168.

Wilkowski, B.M., Robinson, M.D., Gordon, R.D., & Troop-Gordon, W. (2007). Tracking the evil eye: trait anger and selective attention within ambiguously hostile scenes. *Journal of Research in Personality, 41,* 650-666.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior research methods*, *42*(3), 671-684. https://doi.org/10.3758/BRM.42.3.671.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1-8. https://doi.org/10.1038/s41562-017-0058

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. Psychonomic Bulletin & Review, 11, 269–274.

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. (2013). Where do we store the memory representations that guide attention?. *Journal of vision*, *13*(3), 1-17. https://doi.org/10.1167/13.3.1.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychological Science, 12*, 219–224.

Wrangham, R. W. (2018). Two types of aggression in human evolution. *Proceedings of the National Academy of Sciences*, *115*(2), 245-253. https://doi.org/10.1073/pnas.1713611115

Wu, L.Z., Zhang, H., Chiu, R.K., Kwan, H.K. and He, X. (2014). Hostile attribution bias and negative reciprocity beliefs exacerbate incivility's effects on interpersonal deviance. *Journal of Business Ethics*, *120*(2), 189-199. https://doi.org/10.1007/s10551-013-1658-6

Xie, W., & Zhang, W. (2017). Negative emotion enhances mnemonic precision and subjective feelings of remembering in visual long-term memory. *Cognition*, *166*, 73-83. https://doi.org/10.1016/j.cognition.2017.05.025

Xie, W., Li, H., Ying, X., Zhu, S., Fu, R., Zou, Y., & Cui, Y. (2017). Affective bias in visual working memory is associated with capacity. *Cognition and Emotion*, *31*(7), 1345-1360. https://doi.org/10.1080/02699931.2016.1223020

Xu, Q., Ye, C., Gu, S., Hu, Z., Lei, Y., Li, X., Huang, L., & Liu, Q. (2021). Negative and positive bias for emotional faces: Evidence from the attention and working memory paradigms. *Neural Plasticity, 2021,* Article 8851066. https://doi.org/10.1155/2021/8851066

Yao, N., Rodriguez, M. A., He, M., & Qian, M. (2019). The influence of visual working memory representations on attention bias to threat in individuals with high trait anxiety.

Journal of Experimental Psychopathology, *10*(4), 1-15. https://doi.org/10.1177/2043808719876149

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and emotion*, *24*(1), 3-47. https://doi.org/10.1080/02699930903205698

Yin, S., Bi, T., Chen, A., & Egner, T. (2021). Ventromedial prefrontal cortex drives the prioritization of self-associated stimuli in working memory. *Journal of Neuroscience*, *41*(9), 2012-2023. https://doi.org/10.1523/JNEUROSCI.1783-20.2020

Young, M. J., Tiedens, L. Z., Jung, H., & Tsai, M. H. (2011). Mad enough to see the other side: Anger and the search for disconfirming information. *Cognition and Emotion*, *25*(1), 10-21. https://doi.org/10.1080/02699930903534105

Yuan, K. H., Bentler, P. M., & Zhang, W. (2005). The effect of skewness and kurtosis on mean and covariance structure analysis: The univariate case and its multivariate implication. *Sociological Methods & Research*, *34*(2), 240-258. https://doi.org/10.1177/0049124105280200

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology*, *32*(2), 180-198. https://doi.org/10.1080/20445911.2020.1717498

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological science*, *22*(11), 1434-1441. https://doi.org/10.1177/0956797611417006

Zhu, W., Chen, Y., & Xia, L. X. (2020). Childhood maltreatment and aggression: The mediating roles of hostile attribution bias and anger rumination. *Personality and individual differences*, *162*, 110007. https://doi.org/10.1016/j.paid.2020.110007

Zvielli, A., Bernstein, A., & Koster, E. H. (2015). Temporal dynamics of attentional bias. *Clinical Psychological Science*, *3*(5), 772-788. https://doi.org/10.1177/2167702614551572

Zvielli, A., Bernstein, A., and Koster, E. H. W. (2014). Dynamics of attentional bias to threat in anxious adults: bias towards and/or away? PLoS ONE 9:e104025. https://doi.org/10.1371/journal.pone.0104025