RUNNING HEAD: CHB-MODEL

The Computations of Hostile Biases (CHB) model:

Grounding hostility biases in a unified cognitive framework

Danique Smeijers[1], Berend H. Bulten[1], and Inti A. Brazil[1,2]

[1]Forensic Psychiatric Centre Pompestichting, Nijmegen, The Netherlands

[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Cite as:

Smeijers, D., Bulten, E. B., & Brazil, I. A. (2019). The Computations of hostile biases (CHB) model: Grounding hostility biases in a unified cognitive framework. *Clinical psychology review*, *73*, 101775. https://doi.org/10.1016/j.cpr.2019.101775

Abstract

Our behavior is partly a product of our perception of the world, and aggressive individuals have been found to have 'hostility biases' in their perception and interpretation of social information. Four types of hostility biases can be distinguished: the hostile attribution, interpretation, expectation, and perception bias. Such low-level biases are believed to have a profound influence on decision-making, and possibly also increase the likelihood of engaging in aggressive acts. The current review systematically examined extant research on the four types of hostility bias, with a particular focus on the associations between each type of hostility bias and aggressive behavior. The results confirmed the robust association between hostility biases and aggressive behavior. However, it is still unknown how exactly hostility biases are acquired. This is also caused by a tendency to study hostility biases separately, as if they are non-interacting phenomena. Another issue is that current approaches cannot directly quantify the latent cognitive processes pertaining to the hostility biases, thus creating an explanatory gap. To fill this gap, we embedded the results of the systematic review in a state-of-the-art computational framework, which provides a novel mechanistic account with testable predictions.

Keywords:

Hostility biases, aggressive behavior, computational modelling

1. Introduction

Accurate processing of social information is crucial for normal socialization and interaction. Due to deficits in social information processing, aggressive individuals are thought to interpret, perceive, and make decisions about social stimuli in such a manner that the likelihood of engaging in aggressive acts increases (Dodge & Crick, 1990). Substantial research has focused on the associations between aggressive behavior and deficits in the interpretation and representation of social information. One of the most influential notions is that aggression occurs after making a hostile attribution that "the self" has been threatened (for a review see Orobio de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002; Dodge, 2006). This tendency of aggressive individuals to attribute hostile intent to others' actions is often referred to as "the hostile *attribution* bias" (HAB; Nasby, Hayden, & DePaulo, 1980). For instance, if someone bumps into you, a hostile attribution could be that this person did it on purpose to harm you.

The ability to match an act that causes a negative outcome for the self on the one hand, with the cognitive attribution that the intent of the actor must be consistent with the outcome on the other, seems to be innate. But, the ability to attribute benign intent to situations with bad outcomes is thought to develop around the third year of life, along with the development of theory of mind. This benign attributional style is acquired successfully by most children. However, some children fail to learn to identify cues that signal that the actor had good intentions and was not being hostile. A hostile, instead of benign, attributional style then becomes a stable personality-like characteristic that guides behavior (Dodge, 2006). The association between aggressive behavior and the HAB is thought to be robust and has previously been found in children and adults in the general population, as well as in clinical samples (for a review see Orobio de Castro et al., 2002; Dodge, 2006; Tuente, Bogaerts, & Veling, 2019).

There are also other hostility biases that may occur during social information processing, in addition to the hostility bias in the attribution of intent. Three other hostility biases have been identified; the hostile interpretation, perception, and expectation bias, respectively. It is important to note that these terms are sometimes used interchangeably. We operationalized the hostile *interpretation* bias (HIB) as the a-priori tendency to interpret social stimuli as hostile. For example, when someone is looking at you, a hostile interpretation could be that the person's facial expression signals that the person is angry even though, in reality, the expression is non-threatening. The hostile *perception* bias (HPB) alludes to the tendency to perceive ambiguous social interactions as hostile (Bushman, 2016). For example, when you see two people talking loudly during a conversation, a hostile perception could be that they are arguing or getting ready to fight. The difference between the HIB and the HPB is that the interpretation bias concerns the interpretation of social stimuli solely, whereas the perception bias is broader and concerns a social interaction as a whole. Finally, the hostile *expectation* bias (HEB) refers to the tendency to assume that someone will react to potential conflicts with hostility (Bushman, 2016). For instance, if you bump into another person, a hostile expectation could be that the person will assume that you did it on purpose and will attack you.

All four hostility biases have repeatedly been found to be associated with higher levels of aggressive behavior (for a review see Bushman, 2016; Dodge, 2006; Mellentin, Dervisevic, Stenager, Pilegaard, & Kirk, 2015; Orobio de Castro et al., 2002; Tuente et al., 2019), and contribute significantly to the development and the persistency of aggression: When attributing, interpreting, perceiving or expecting hostility in others, one is more likely to act aggressively, which in turn causes others to respond more aggressively, thus further strengthening the person's hostile view on others (e.g. Crick & Dodge, 1996). Hostility biases, therefore, are important constructs for the understanding and treatment of aggressive

behavior in clinical settings. The HAB is currently considered a target for interventions that aim to reduce behavioral problems (Orobio de Castro et al., 2002; Tuente et al., 2019). Also, altering the HIB is assumed to be an important addition to traditional interventions for antisocial pathology (Mellentin et al., 2015). However, how hostility biases exactly are acquired is still unexplained. For instance, the mutual associations, the similarities and differences between the different types of hostility biases as well as their underlying processes are yet unknown. An enhanced understanding of these single and/or shared components is needed before hostility biases can be systematically targeted in clinical settings.

The main goal of the current review was to systematically examine extant research on the four types of hostility bias, with a particular focus on the associations between each type of hostility bias and aggressive behavior. We also examined the body of evidence concerning different techniques/interventions that have been used to try to alter the biases. Finally, given the current lack of understanding of the cognitive mechanisms sub-serving hostility biases, we embedded the results of the systematic review in a state-of-the-art computational framework. Such an approach has the potential to significantly support the current state of affairs, as computational frameworks offer clearly defined theoretical models for the mechanistic underpinnings of a cognitive domain (e.g. learning, attention, perception), and provide methodological tools to directly quantify target cognitive processes that are part of such mechanisms and study them systematically (Brazil, van Dongen, Maes, Mars, & Baskin-Sommers, 2018; Wiecki, Poland, & Frank, 2015). Thus, we will propose a novel approach to understanding and studying hostility biases, in which the biases are anchored in a single, well-defined framework that integrates separate cognitive processes and their interactions and is supported by the most recent neuroscientific insights.

2. Method

2.1 Search strategy and included studies

A systematic search was conducted for articles written in English, published before March 2019 and describing the association between hostility biases and aggressive behavior. However, a meta-analysis and systematic review about the HAB (41 studies included) and about the HIB (15 studies included) were published in 2002 and 2015, respectively (Orobio de Castro et al., 2002; Mellentin et al., 2015). The review by Tuente et al. (2019) about the HAB (25 studies included) was restricted to adult populations. Part of the current review supplements these previous reviews. Therefore, the systematic searches for articles about these two biases were restricted to additional articles published between 2002 and March 2019 for the HAB and articles published between 2015 and March 2019 for the HIB.

PsycINFO and PubMed were used to search articles with the following keywords: hostile (attribution, interpretation, perception, expectation) AND bias AND anger; hostile (attribution, interpretation, perception, expectation) AND bias AND aggress*; hostile (attribution, interpretation, perception, expectation) AND bias AND violen*; hostile (attribution, interpretation, perception, expectation) AND bias AND offend*; hostile (attribution, interpretation, perception, expectation) AND bias AND forensic. This initial search yielded 375 references of which 135 were duplicates. The remaining 240 articles were screened for eligibility, resulting in 86 relevant references (see Figure 1 for a flowchart of the study selection process). All titles and abstracts were screened by two researchers independently (DS & MW; inter-rater agreement Cohen's Kappa = .93). Disagreements about the inclusion of an article were discussed by DS and MW. Relevant articles were selected using the following inclusion criteria based on the title and abstract:

1. Articles should be original research published in peer reviewed journals
2. Articles should be written in English

3. Anger or aggression and the hostile attribution, interpretation, perception and/or expectation bias should be the main focus of the article

4. The focus of the articles should not be on emotion recognition

5. The focus of the articles should not be on the instruments used to assess one of the hostility biases and their psychometric properties

6. All samples were allowed, such as community samples, undergraduates, offender samples, children, adolescents, and adults

7. Gender was not an exclusion criterion

8. Articles about the HAB were published between 2002 and March 2019

9. Articles about the HIB were published between 2015 and March 2019

10. No restrictions for period of publication were used for articles on the HPB and HEB.

Subsequently, the searches were supplemented by cross-referencing which yielded 11 extra relevant articles. An overview of included studies is presented in Table 1 (appendix).

2.2 Study characteristics

The total number of participants of the selected studies was 21582. Sixty-three studies were focused on HAB, 17 on the HIB, 2 on the HPB, and 0 on the HEB. One meta-analytic review focused on the HAB, HIB, and HEB and one study focused on both the HAB and HIB. Twenty-seven studies investigated only children, 7 studies adolescents, 14 studies undergraduate students, 24 studies adults, 3 studies children and their parents, 1 study adolescents and their parents, 1 study children and adolescents, and 3 studies examined undergraduates and adults. Of all selected studies, 25 included clinical samples, 56 were focused on healthy sample, and 5 were meta-analyses and/or systematic reviews.

3. Results

Examination of the studies included highlighted that the core topics of empirical studies on the biases could be categorized into the following 10 domains: aggression, personality, gender, peers and parents, maltreatment, media/cyber violence, neural and biological underpinnings, perceptual sensitivity, emotion, and intervention. Therefore, we organized the presentation of the results following these domains. Seven studies did not fit within these domains, and were grouped into a category called '*emerging research domains*'. Note also that no articles were found that focused on the HEB, except for one review, which is why a presentation of the results for the HEB is lacking.

3.1. Associations with aggression

3.1.1. HAB

Previous meta-analytic reviews already found a robust positive association between aggressive behavior and the HAB among children, adults, clinical and non-clinical samples (Dodge, 2006; Orobio de Castro et al., 2002; Tuente et al., 2019). This association is further supported by the additional literature identified in the current review. After the publication of the meta-analysis by Orobio de Castro et al. (2002), eight new studies examined aggressive samples as compared to non-aggressive samples (Chen, Coccaro, & Jacobson, 2012; Gagnon et al., 2017; Godleski & Ostrov, 2010; Jahoda, Pert, & Trower, 2006; Jin, Eagle, & Keat, 2008; MacBrayer, Milich, & Hundley, 2003; Orobio de Castro, Merk, Koops, Veerman, & Bosch, 2005; Orobio de Castro, Slot, Bosch, Koops, & Veerman, 2003). These studies showed that aggressive children as well as aggressive adults displayed higher levels of the HAB compared to non-aggressive individuals. An association between anger, aggression and the HAB was also found among patients with psychotic disorders (Combs et al., 2009; Darrell-Berry et al., 2017). Greater levels of the HAB were displayed by aggressive patients

with persecutory delusions compared to undergraduate students (Combs et al., 2009). The HAB was also associated with aggressive responses on a stories task in which mentally ill offenders had to create an additional sentence for an ambiguous story (Edwards & Bond, 2012). Also, children referred to mental health institutions for ADHD or oppositional defiance disorder displayed higher levels of the HAB and aggression as compared to children without psychiatric problems (r = .39; MacBrayer et al., 2003). The HAB was also found to be predictive of paranoia in a sample of psychiatric patients with persecutory delusions (Combs et al., 2009), and of aggressive behavior six months post-testing among undergraduate students (r = .37; Quan et al., 2019).

Only two studies failed to confirm the association between the HAB and aggressive behavior; one concerned a sample of urban youth, and the other incarcerated males. Leff et al. (2014) suggested that the HAB may serve as a protective factor in some situations that urban, ethnic minority youth may encounter. In a sample of incarcerated males, an association was found between high trait anger and an aggressive response (e.g. shove or punch him) towards ambiguous vignettes, but not with the HAB (Bowen, Roberts, & Kocian, 2016). Another study did not find any differences between children with conduct problems relative to a group without conduct problems and a third group with conduct problems and co-morbid callous unemotional traits (Helseth, Waschbusch, King, & Willoughby, 2015). The latter study, however, did find that children with both callous unemotional traits and conduct problems generated more aggressive responses to provocation in hypothetical scenarios relative to the other groups.

When focusing on specific types of aggression, it was found that the HAB was associated with reactive (Choe, Shaw, & Forbes, 2015; Dodge, 2006; Dodge et al., 2015; Gagnon & Rochat, 2017; Kokkinos, Karagianni, & Voulgaridou, 2017; Oriobio de Castro et al., 2002; Orobio de Castro et al., 2003), chronic aggression (Dodge et al., 2015), relational and physical

aggressive behavior (Bailey & Ostrov, 2007; Bondü, 2018; Cillessen, Lansu, & Van Den Berg, 2014; Crick, Grotpeter, & Bigbee, 2002; Gentile, Coyne, & Walsh, 2011; Godleski & Ostrov, 2010; Godleski, Ostrov, Houston, & Schlienz, 2010; Kokkinos et al., 2017; Martins, 2013; Mathieson et al., 2011; Werner, 2012; Yeung & Leadbeater, 2007), aggressive driving behavior (Matthews & Norris, 2002), and intimate partner violence (Thomas & Weston, 2019). Reactive aggression refers to impulsive responses to threat, frustration or provocation. Relational aggression inflicts harm through damage or control of friendships or other relationships. The strength of these associations varied across studies from small to moderate (*r* ranging from .08 to .34). The study by De La Osa et al. (2018) found that the HAB moderated the association between relational aggression and symptoms of the oppositional defiant disorder related to overt aggression (e.g. arguing, losing temper). Only the study by Helfritz-Sinville and Stanford (2014) failed to confirm the association between reactive aggression and the HAB. But, their results indicated that the aggressive groups were more likely to report to being threatening, rude or to use physical force in ambiguous situations (*r* ranging from .37 to .40).

The association with relational and physical aggression seems to depend on the kind of provocation situations, described in vignettes, that were used to measure the HAB. Individuals prone to show relational aggression were found to display higher levels of the HAB in response to relational provocations (Bailey & Ostrov, 2007; Crick et al., 2002; Godleski et al., 2010; Kokkinos et al., 2017; Martins, 2013; Werner, 2012), and higher levels of physical aggression were associated with higher levels of the HAB for instrumentally provocative situations (Bailey & Ostrov, 2007; Crick et al., 2002; Godleski & Ostrov, 2010; Martins, 2013). Only one study found that the HAB in response to physical provocations was associated with more relational aggressive behavior (Werner, 2012).

3.1.2. HIB

The association between the HIB and aggressive behavior has mainly been investigated in relation to emotional facial expressions. A previous systematic review including 15 studies found that individuals prone to show anger and aggression were characterized by a bias in processing facial expressions as angry or hostile in neurocognitive paradigms (Mellentin et al., 2015). The authors suggested that this HIB might be a cognitive dysfunction that possibly could mediate aggressive behavior in susceptible populations. In line with Mellentin et al. (2015), the novel studies identified in our systematic search further confirmed that aggressive individuals displayed a stronger HIB towards ambiguous emotional facial expressions as compared to healthy controls, psychiatric patients with lower levels of aggression and non-violent offenders (*r* ranging from .34 to .44; Smeijers, Rinck, Bulten, van den Heuvel, & Verkes, 2017; Wegrzyn, Westphal, & Kissler, 2017). In addition, one study found that forensic psychiatric outpatients exhibiting severe aggressive behavior also displayed higher levels of the HIB to unambiguous emotional facial expressions (Smeijers et al., 2017). There has been only one study that did not find evidence for an increased HIB in violent offenders compared to non-violent offenders and controls (Kuin, Masthoff, Munafò, & Penton-Voak, 2017). The authors suggested that the heterogeneity within their samples might provide an explanation for their results, as the samples consisted of mild to severely violent (sexual) offenders in the experimental group and of prison workers in the control group.

Studies focusing on subtypes of aggression found that the HIB was associated with reactive aggressive behavior in psychiatric and forensic psychiatric in- and outpatients (*r* ranging from .16 to .21; Lobbestael, Cima, & Arntz, 2013; Smeijers et al., 2017). In line with these studies, psychopathy, a personality disorder characterized by a tendency to use more proactive than reactive aggression, was found not to be associated with the HIB, but rather with a better recognition of hostile-looking eyes and more correct attributions of mental states

than non-psychopathic offenders (Nentjes, Bernstein, Arntz, van Breukelen, & Slaats, 2015). The latter finding suggested that individuals with high levels of psychopathy may not display a bias, but instead show an increased sensitivity to hostile mental states.

Higher levels of the HIB were also found to be associated with the disposition to act aggressively, severity of aggressive behavior, and cognitive distortions (*r* ranging from .18 to .29; Smeijers et al., 2017). Furthermore, an association has been found between the HIB, pain, and depression (*r* =.38 and *r* = .48, respectively) among individuals seeking anger management treatment (McDermott, Smith, Matheny, & Cougle, 2017; Smith, Summers, Dillon, Macatee, & Cougle, 2016). The authors provided several possible explanations for these associations; 1) individuals high in anger may tend to react to pain with more aggression due to the presence of depression; 2) such individuals show elevated levels of depression in response to their feelings of anger, which might lead to heightened pain perceptions; and 3) it may be that irritability plays a role, as depression, aggression, and the experience of pain all can be characterized by irritability.

3.1.3. HPB

The HPB was investigated in two studies, but in five experiments in total. Bartholow and Heinz (2006) studied the association between aggression, alcohol and the HPB. In order to examine this association, they primed half of the participants with alcohol-related advertisements. The other half were primed with neutral advertisements. The HPB was found to be positively associated with alcohol use and aggression-related alcohol problems (*r* = .21). Participants in the alcohol prime condition showed more hostile perceptions. Furthermore, participants with high and moderate aggression-related expectancies displayed higher levels of the HPB in the alcohol prime condition. The authors suggested that their results provide

support for the notion that alcohol can influence aggression-related outcomes even in the absence of alcohol consumption.

The association between rejection, social exclusion, aggression and the HPB in social environment was investigated in four studies by DeWall, Twenge, Gitter, and Baumeister (2009). Across these studies, participants were allocated to the social exclusion or control condition. Several manipulations of social exclusion were used. In study 1 participants were told that their confederate did not want to meet them. In study 2 participants were told that they possessed a personality type in which it was likely that they would end up alone in life. In study 3 and 4 participants were scored on the personality traits extraversion and surgency. These studies had two social exclusion conditions: interpersonal and individual failure. Interpersonal-failure participants were then told that they scored high on surgency and therefore could anticipate a future filled with professional accomplishment but may experience a dearth of long and lasting interpersonal relationships. Participants in the individual failure conditions were told that the participant's level of extraversion is a good thing for relationships and is linked to having an easy time keeping relationships together. They were then informed that their high surgency score meant they could anticipate successful interpersonal relationships but might experience a lack of professional accomplishment.

The results showed that the participants in the social exclusion conditions displayed a bias to perceive neutral or ambiguous stimuli (words and word-pairs) as more hostile as compared to the control condition. Participants in the social exclusion and interpersonal failure condition had a more hostile impression of other's traits relative to the participants in the other two conditions. Participants in social exclusion conditions displayed more aggression. Taken together, the results of these studies suggest that social exclusion leads to hostile cognitions/biases that seem to be associated with aggressive behavior.

3.1.4. Interim summary

In line with previous meta-analyses and reviews (Bushman, 2016; Dodge, 2006; Orobio de Castro et al., 2002; Mellentin et al., 2015), we found that the association between aggressive behavior and hostility biases is robust. The HAB and HIB, in particular, were both repeatedly found to be associated with reactive aggression. Next, the following domains will be discussed in relation to hostility biases: personality, gender, peers and parents, maltreatment, media/cyber violence, neural underpinnings, perceptual sensitivity, emotion, and intervention.

3.2. Personality characteristics

Four studies investigated the role of personality in relation to the HAB. One study focused on the malleability of personality. A meta-analysis including 11 studies revealed that an entity theory of personality was predictive of the HAB (r = .20) and aggressive desires following ambiguous provocations (r = .23; Yeager, Miu, Powers, & Dweck, 2013). An entity theory of personality refers to the idea that people's traits cannot change. Two follow-up experimental studies investigated whether learning to view your personality as malleable was associated with the HAB. These results showed that individuals that learned that personality traits have the potential to change, exhibited less hostile attributions (r = .33). At 8-months follow up, these individuals reported less aggressive feelings (r = .29). Furthermore, they found that an entity theory of personality was predictive of greater levels of the HAB and aggressive feelings at 8-months follow-up (r = .25 and r = .35, respectively; Yeager et al., 2013). Taken together, these results suggest that someone's beliefs of whether his/her personality is malleable or fixed could shape the individual's attribution style.

Another study found that the big five personality traits were associated with the HAB (Kokkinos et al., 2017). The Big Five personality traits consist of neuroticism (general

tendency to experience negative affects), agreeableness (e.g. altruistic, sympathetic to others), conscientiousness (e.g. active in planning organizing and carrying out task, self-control), extraversion (e.g. traits such as assertiveness, sociability, talkative) and openness to experience (e.g. attentiveness to inner feelings, active imagination, intellectual curiosity; Rothmann & Coetzer, 2003). Specifically, it was found that agreeableness, conscientiousness, extraversion and openness correlated negatively with the level of the HAB and aggressive behavior (*r* ranging from -.12 to -.45) while neuroticism correlated positively (*r* = .25; Kokkinos et al., 2017). There was also an association between relational aggression and the following combination of Big Five traits: high neuroticism (*r* = .16), and low agreeableness, conscientiousness, extraversion and openness, was mediated by the HAB (*r* ranging from .16 - .18). The authors suggested that individuals with this combination of personality traits might be more likely to make hostile attributions, which in turn is associated with an increased likelihood of displaying aggression. In addition, Li et al. (2016) showed that the HAB was associated with aggression, anger, and state narcissism (*r* ranging from .18 to .84). State narcissism was also related to anger and aggression (*r* = .27 and *r* = .21, respectively). An increase in state narcissism was found to be associated with higher levels of aggression (*r* = .02) and anger (*r* = .38), but not with higher levels of the HAB. An increase in state narcissism was induced by letting participants listen to a narcissism story about a handsome young god Narcissus, who fell in love with his own image reflected in a pool. The control condition listened to a neutral story about how Prometheus created humans. Edwards and Bond (2012), however, found that the HAB was predicted by high narcissism and low self-concept clarity in mentally disordered offenders. Self-concept clarity refers to the extent to which self-beliefs are consistent, stable, and clearly defined. Furthermore, only small correlations were found between level of the HAB and the personality dispositions victim sensitivity (*r* = .17), provocation sensitivity (*r* = .24), and moral disgust (*r* = .20; Bondü & Richter, 2016).

In summary, the results of the aforementioned studies suggest that believing that someone's personality is fixed is associated with greater levels of the HAB. Also, neuroticism, victim and provocation sensitivity and moral disgust seem to be related to greater levels of the HAB. However, the found associations were small and have to be replicated by future research.

3.3. Gender differences

The association between the HAB and aggressive behavior was often investigated across whole study samples. However, several studies investigated gender differences. One study found that physical and relational aggression displayed by boys was associated with the HAB (Cillessen et al., 2014). The study by Godleski and Ostrov (2010) reported that girls displayed higher levels of the HAB as compared to boys (r = .10). This study specifically found that girls with more physical aggression reported higher levels of the HAB for instrumentally provocative situations (r = .14; Godleski & Ostrov, 2010). The study by Möller and Krahé (2009) found that girls showed higher levels of the HAB in response to scenarios with relational content, whereas boys displayed higher levels of the HAB for scenarios with physical content. Bondü (2018) found that girls displayed lower levels of the HAB (r = .47), but also that the association between the HAB and physical aggression was stronger for boys. However, Yeung and Leadbeater (2007) did not find differences in the level of HAB displayed by boys and girls. Taken together, the results of the few studies examining the role of gender revealed mixed results leaving the role of gender inconclusive.

3.4. Peers and parents

Several factors are thought to be associated with the development or maintenance of the HAB. Especially the role of peers and parents in the development of the HAB has received attention. Ten studies were dedicated to reveal this association.

One study pointed out that higher levels of the HAB were shown towards enemies as compared to friends (Peets, Hodges, Kikas, & Salmivalli, 2007). This is in line with previous reviews which found that greater levels of the HAB were displayed in response to aggressive peers and after peer rejection (Dodge, 2006; Orobio de Castro et al., 2002). Another study revealed that the peer group HAB was predictive of an individual's own HAB (r = .35; Halligan & Philips, 2010). It also emerged that this association was even stronger when only reciprocal friendships were considered. Additionally, the effect of the HAB on aggression was stronger at lower levels of popularity (Cillessen et al., 2014). Peer contagion was also found to appear in online environments. Adolescents who chatted with a hostile confederate displayed higher levels of the HAB whereas the HAB of adolescents who chatted with a benign confederate even decreased (Freeman, Hadwin, & Halligan, 2011). To summarize, these results suggest a transmission of social information processing which occurs in natural as well as in online environments.

Besides peers, parents also seemed to play a role in the development of the HAB. This link was described in previous systematic meta-analyses (Dodge, 2006; Orobio de Castro et al., 2002). The additional studies identified in the present review yielded mixed results. First of all, the study by Mammen, Kolko, and Pilkonis (2003) showed that parental satisfaction was associated with aggressive parental behavior but not with the HAB. Parental psychological control, however, was found to be predictive of the HAB but only in boys, not in girls (Nelson & Coyne, 2009). Moreover, high levels of maternal sensitivity (e.g. appropriate responses to children's requests) were associated with low levels of the HAB in

children (r = .55; Wong, Chen, & McElwain, 2019). Children with a disorganized attachment style showed higher levels of the HAB and more aggressive goals relative to children with an organized attachment style (Zajac, Bookhout, Hubbard, Carlson, & Dozier, 2018).

Additionally, three studies examined the role of the HAB displayed by parents themselves (Halligan, Cooper, Healy, & Murray, 2007; MacBrayer et al., 2003; Werner, 2012). Attributions that were made by fathers and mothers regarding their own child were partially determined by their own HAB (Halligan et al., 2007). The mothers of children referred to psychiatric institutions, who displayed more HAB than children without psychiatric problems, showed higher levels of the HAB themselves as compared to mothers from the comparison group (r = .54; MacBrayer et al., 2003). The HAB displayed by mothers in parent-child interactions was also associated with their children's aggressive behavior (Werner, 2012). More specifically, it emerged that mothers who exhibited higher levels of the HAB regarding adult peers as well as regarding their own children and their classmates, had children who described themselves as more relationally aggressive (Werner, 2012). An association between the HAB of parents and the HAB of children could not be confirmed by Halligan et al. (2007), while MacBrayer et al. (2003) and Werner (2012) suggested that the mother's HAB was more closely related to the daughter's HAB and aggressive behavior as compared to sons.

All in all, the above studies suggest that peer contagion contributes to the occurrence of the HAB. The results of the few studies examining the role of parents in the development of the HAB revealed mixed results leaving the role of parents inconclusive.

3.5. Maltreatment

Three studies assessed the association between the HAB and maltreatment and all confirmed an association between adverse events and elevated levels of the HAB. Relational

aggression was predicted by the HAB when combined with relatively high levels of relational victimization and emotional sensitivity following provocation (Mathieson et al., 2011). Among violent batterers, it was found that childhood exposure to emotional abuse was positively associated with the HAB (Jin et al., 2008). Greater levels of the HAB were found in maltreated adolescents living in out of home care (Kay & Green, 2016). To summarize, even though only three studies were found, they all provided support for the existence of a positive association between maltreatment and the HAB.

3.6. Media/cyber violence

A recent meta-analysis including 37 studies found that exposure to violent media was positively associated with hostility biases (average r = .20; Bushman, 2016). The majority of the studies included in this meta-analysis focused on violent video games. Only a few were dedicated to the role of violent cartoons and violent television. When focusing on the different types of hostility biases, this meta-analysis revealed that exposure to violent media was associated with higher levels of the HAB and HEB (r = .18 and r = .26, respectively), but no correlation was found for the HIB. No studies about the HPB were included in the meta-analysis. Also, no firm conclusions could be drawn regarding the HIB, because only three studies had been conducted on this topic. Furthermore, the association between media violence and hostility biases was small to moderate and appeared to become larger with age. Note that the studies about the HEB included in the meta-analysis by Bushman (2016) were the only ones found in the literature. This highlights that the HEB has only been investigated in relation to violent media exposure.

Three studies that were not part of the meta-analysis by Bushman (2016) further confirmed an association between watching/exposure to violence and the HAB. Martins (2013) found that children displayed more severe levels of the HAB in response to

instrumental provocations immediately after being exposed to a clip containing physical aggression (r = .27), whereas exposure to relational aggression was associated with higher levels of the HAB in response to relational provocations (r = .24). Children who watched playful fighting episodes, on the other hand, displayed lower levels of the HAB (r = .50; Boulton, 2012). Also, Gentile et al. (2011) found a positive association between exposure to violent television/games/movies and the HAB (r = .25), and between the amount a child spend "before the screen" and the HAB (r = .18). Taken together, there seems to be a positive association between watching/exposure to violence and the HAB. However, more research is needed to examine this relation in the HIB, HEB, and HPB.

3.7. Neural and biological underpinnings

The neural underpinnings of the HAB were examined by five studies. It was found that the HAB for relational provocations was predictive of increased electrophysiological activity in frontal brain regions, indexed with an event-related potential known as the P300 (Godleski et al., 2010). The authors suggested that individuals with high levels of the HAB were more sensitive to potential salient stimuli, as larger P300 amplitude is hypothesized to reflect greater allocation of cognitive resources or enhanced attending. In a longitudinal design, it was revealed that the HAB at the age of 10 and 11 was predictive of increased left amygdala reactivity to fearful faces at age 20 (Choe et al., 2015). The N400 response to a critical word that mismatched with the character's expected hostile intention was found to be larger for aggressive psychiatric patients compared to non-aggressive individuals (Gagnon et al., 2017). The authors suggested that this might indicate that aggressive individuals developed stronger expectations about hostile intent and that this expectation made it more difficult to integrate a non-hostile resolution in a hostile context. Furthermore, an enhanced late positive potential-like component in aggressive psychiatric patients was found when hostile intentions took

place in a non-hostile context, which was associated with reactive aggressive behavior (Gagnon et al., 2017). The study by Wang et al. (2018) found a positive association between the HAB and a greater density in the left medial frontal gyrus (r = .27). As the left medial frontal gyrus is thought to be associated with memory retrieval, the authors suggested that individuals exhibiting higher levels of the HAB retrieved hostile knowledge structures and schemas to understand social information in ambiguous situations. Research among individuals with moderate to severe brain injury found that this population exhibited larger levels of the HAB for hostile, ambiguous and benign scenario's as compared to individuals without brain injury (Neumann, Malec, & Hammond, 2017).

Only one longitudinal study was focused on biological factor, namely on the role of the monoamine oxidase A (MAOA) genotype (Galán, Choe, Forbes, & Shaw, 2017). They found that Caucasian men with low activity MAOA (L-MAOA) displayed higher levels of the HAB and aggressive behavior as compared to Caucasian men with high activity (H-MAOA). Regardless of the MAOA genotype, it was found that maternal punitiveness in toddlerhood positively predicted the HAB in middle childhood, violent attitudes in adolescence, and self-reported antisocial behavior in adulthood. This result was only found among African American's. For L-MAOA Caucasian men, it was found that higher levels of the HAB also predicted greater levels of adult antisocial behavior. It is important to note that no firm conclusions can be drawn based on one monogenetic study. Also, aggression is complex behavior and is, therefore, more likely to be explained by a complex interaction between multiple genes (for a meta-analysis see Vassos, Collier, & Fazel, 2014).

In summary, the neural and biological underpinnings of the HAB were only examined in a few studies, which differed in the mechanisms and processes that were targeted. The scarcity of data on this topic highlight that there is a desperate need for future research on the neural and biological substrates of hostility biases.

3.8. Perceptual sensitivity/attention allocation

Previously, it has been suggested that the HAB might occur partly due to problems with the encoding of intent cues (Orobio de Castro et al., 2002). It often was thought that this tendency is associated with attention allocation. It especially was thought that aggressive individuals pay relatively more attention to hostile than to non-hostile cues (Crick & Dodge, 1994). The schema inconsistency hypothesis, however, proposes that hostile schemas direct attention towards schema-inconsistent information (Horsley, Orobio de Castro, & Van der Schoot, 2010), and that whether this information is interpreted accurately depends on the ambiguity and the strength of the schema. Horsley et al. (2010) examined the schema inconsistency hypothesis. They showed three cartoons displaying real-life ambiguous provocation situations. Participants were asked to pretend to be one of the two characters while viewing the first cartoon. In the second cartoon, the other character behaved in a hostile, non-hostile, or ambiguous way. In the third picture, the first character experienced a harmful (negative) event and the other character expressed a neutral or a behavior-congruent emotion expression. While performing this task, eye-movement was assessed using an eye-tracker. All pictures remained visible throughout the task, enabling participants to look back at previous pictures if desired. Their results showed that the high and low aggression group looked back longer at the second picture when more schema-inconsistent information was present. They specifically found that the high aggression group looked back longer at more non-hostile (aggressive schema-inconsistent) information and the low aggression group looked back longer at more hostile (non-aggressive schema inconsistent) information. Furthermore, the high aggression group recalled marginally less non-hostile cues even though they had looked at these cues longer. This suggested that schema-based perception made processing of schema-inconsistent information more difficult. The authors concluded that there was no

indication of hypersensitivity to hostile cues, but that the findings provided some first evidence in support of the schema inconsistency hypothesis.

Another study found that the HAB was not associated with gaze perception in violent adolescent offenders (Karadenizova & Dahle, 2018). This study also showed that happy facial expressions elicited a stronger feeling of being looked at (r = .36). Compared to other emotional facial expressions, faces displaying happiness with averted gaze gave the participant the feeling of being looked at more often (r ranging from .38 to .42). Miller and Johnston (2019) investigated the association between the HAB and attentional biases to social threat. Their results showed that being quicker in engaging attention towards social threat (i.e. attentional orienting) and perceiving the presence of social threat earlier than neutral faces (i.e. attentional prioritization) were positively associated with the HAB (r = 14 and r = .23, respectively). Attentional prioritization of social threat was also related to higher levels of aggression (r = .18).

Additionally, the role of attention allocation has also been a topic in studies on the HIB. It often has been wondered whether the HIB regarding emotional facial expressions was associated with impaired emotion recognition and whether aggressive individuals did not exhibit a bias but rather displayed a greater perceptual sensitivity. In trying to understand this phenomenon, other studies also used eye-tracking in order to examine the allocation of attention towards non-hostile and hostile cues. The results showed that non-hostile cues were processed for a longer period of time in high trait anger individuals than in low trait anger individuals (Wilkowski, Robinson, Gordon, & Troop-Gordon, 2007). Furthermore, it was found that individuals high in trait anger showed longer gaze durations for non-hostile cues than for hostile cues. This pattern was found for different trait anger measures (r ranging from .03 to .50), but not among individuals low in trait anger. The latter group did not display longer fixation times to either hostile or non-hostile cues. It was also found that aggressive

tendencies (as measured with a questionnaire) were associated with eye gaze fixation time on non-facial body parts of a character who showed more hostile emotions in the context of non-confrontational social interactions (Lin et al., 2016). This association was not found for confrontational social interactions. Also, an association occurred between aggressive tendencies and eye gaze fixation time on the mouth area. Taken together, these results support the notion that individuals high in trait anger quickly made the inference that ambiguous actions are hostile in nature, even before encoding specific hostile or non-hostile cues in the situation. This indicated that there is a bias even before social stimuli were analyzed. It was also suggested that aggression-prone individuals selectively focus their visual attention as they may not focus their attention towards crucial facial parts, but mainly on the mouth area and on non-facial body parts (e.g. pointing fingers; Lin et al., 2016)

Teige-Mocigemba, Hölzenbein, and Klauer (2016) examined the hypothesis that aggressive individuals were more accurate in perceiving subtle hostile cues, rather than displaying a bias. They made use of pictures of actual people displaying emotional facial expressions, instead of animated pictures. Their results showed that aggressive individuals were more accurate in identifying the moment when a neutral face changed into an angry face, as compared to non-aggressive controls (*r* ranging from .27 to .30). These results converged with the previous finding that physically aggressive individuals were more sensitive to subtle differences in facial expressions of anger (Wilkowski & Robinson, 2012). It also was in line with a study that showed that violent offenders did not show a more general impairment in recognizing emotional facial expressions as compared to sexual offenders (Wegrzyn et al., 2017). On the other hand, it was also found that the more aggressive individuals were, the more likely they were to give an aggressive response to neutral targets irrespective of whether the prime picture displayed an angry, a neutral, or a happy facial expression (Teige-Mocigemba et al., 2016). The authors suggested that it might be plausible

that the processes underlying the sensitivity as well as the bias hypothesis can operate independently or interdependently in different tasks.

An investigation by Jusyte and Schönenberg (2017) in violent offenders revealed that perceptual sensitivity was not altered among violent offenders relative to healthy controls and that violent offenders did not judge facial expressions more often as emotional. This conclusion was in line with the review by Mellentin et al. (2015), who concluded that the HIB was not restricted to a deficit in selective attention. A second study, however, revealed that violent offenders exhibited deficient categorization performance for ambiguous fearful expressions (*r* ranging from .32 to .39). No differences were found regarding ambiguous angry expressions. This categorization deficit was found to be associated with self-reported psychopathy and aggression. Taken together, their findings suggest that violent offenders exhibit a bias in the interpretation of emotional facial expressions, but not a heightened sensitivity to anger.

All in all, the results of the aforementioned studies seem to provide evidence for that the HAB and HIB are not restricted to a deficit in selective attention. However, also some mixed results were reported which suggests that more research is needed to elucidate the role of attention allocation in hostility biases.

3.9. Emotion

The role of emotion in attributing hostile intent to others actions was topic of investigation of ten studies (Chen et al., 2012; Choe, Lane, Grabell, & Olson, 2013; Davies, Coe, Hentges, Sturge-Apple, & Ripple, 2018; Kay & Green, 2016; Matheny et al., 2017; Mathieson et al., 2011; Orobio de Castro et al., 2005; Orobio de Castro et al., 2003; Quan et al., 2019; Wong et al., 2019). The study by Orobio de Castro et al. (2005) revealed that referred aggressive boys not only attributed more hostile intent to peers but also reported

more anger and aggressive responses, less adaptive emotion regulation strategies, and evaluated aggressive responses less negatively as compared to non-aggressive boys. They also found that aggressive behavior could best be explained as a function of the HAB minus adaptive emotion regulation. Additionally, the level of the HAB was found to increase after a negative mood induction in highly aggressive boys relative to moderate and non-aggressive boys (Orobio de Castro et al., 2003).

Other studies found that an adequately functioning theory of mind and emotion understanding were associated with reduced levels of the HAB (Choe et al., 2013; Kay & Green, 2016). Among forensic psychiatric inpatients with psychotic disorders, the HAB was not associated with theory of mind (Bratton, O'Rourke, Tansey, & Hutton, 2017). Wong et al. (2019) found that an interaction between a child's anger proneness and emotion understanding was predictive of the HAB when there were low levels of emotion understanding (r = .56).

Higher levels of the HAB were also found to be associated with emotional sensitivity (Mathieson et al., 2011). Among adults seeking treatment for problematic anger, higher levels of the HAB were associated with low distress tolerance (r = .46; Matheny et al., 2017). Also, fearful distress was associated with higher levels of the HAB at a two-year follow-up (r = .20; Davies et al., 2018). This study could not find a link between temperamental anger, positive affect, and the HAB at two-year follow-up, but it did find that higher levels of anger predicted longer attention allocation (measured using eye-tracking) towards happy, angry, and neutral faces (r = .26; r = .17; r = .20 respectively). Furthermore, anger rumination was found to be associated with the HAB and aggressive behavior (r = .34; r = .57, respectively; Quan et al., 2019). This study also found that anger rumination mediated the association between the HAB and aggression. Also, the study by Wang et al. (2018) found a positive association between anger rumination and the HAB (r = .41).

Taken together, the results of above studies suggest a robust relation between the HAB and emotion. More specifically, the HAB was found to be positively associated with emotionality and negatively with emotion regulation abilities and emotion understanding.

3.10. Emerging research domains

In addition to the aforementioned core topics, seven studies were focused on emerging research domains. These studies showed that the HAB was also associated with mindfulness, IQ, planning ability, justice sensitivity, impulsivity and belief in a just world (Bègue & Muller, 2006; Bondü, 2018; Chen et al., 2012; Choe et al., 2013; Ellis, Weiss, & Lochman, 2009; Gagnon & Rochat, 2017; Heppner et al., 2008). Higher levels of mindfulness were associated with lower levels of aggression as well as with lower levels of the HAB (Heppner et al., 2008). Also, a high IQ was found to be associated with lower levels of the HAB (Choe et al., 2013). Bondü (2018) investigated whether the HAB was associated with justice sensitivity. Three subtypes of justice sensitivity were distinguished: observer, perpetrator, and victim. Being high in observer justice sensitivity refers to an aversive perception of injustice to the disadvantages of others. Individuals high in perpetrator justice sensitivity tend to respond negatively to causing injustice. Finally, being high in victim justice sensitivity refers to frequently feeling unfairly treated and respond with anger and retaliation. The results of this study showed a positive association between the HAB and victim justice sensitivity only (r = .22).

The study by Ellis et al. (2009) revealed that the association between planning ability and reactive and proactive aggression was moderated by the HAB. Specifically, it was found that the association between reactive aggression and planning deficits became extremely strong in the positive direction as the level of the HAB increased, whereas it became strongly negative for proactive aggression.

Another study focused on impulsivity, and found that the association between aggressive behavior and the HAB was weaker at lower levels of impulsivity (Chen et al., 2012). No association was found between aggression and the HAB in individuals with below average levels of impulsivity. Also, the study by Gagnon and Rochat (2017) found an association between HAB and impulsivity. They subdivided trait impulsivity in negative urgency (i.e. the tendency to experience strong impulses, often under conditions of negative affect), lack of premeditation, lack of perseverance, and sensation seeking. A positive associated was found between the HAB and negative urgency (r = .39). They also found that negative urgency mediated the association between reactive aggression and the HAB (this model explained 28% of the variance).

Finally, Bègue and Muller (2006) investigated whether a belief in a just world moderated the HAB. The concept of belief in a just world refers to the idea that good things tend to happen to good people, whereas bad things tend to happen to bad people. Belief in a just world can be directed to yourself and to others (i.e. I am treated fairly by the world, or others are treated fairly by the world). Their results showed that high troublemakers, in an ambiguous situation, and with a high belief in a just world for their selves, showed less aggressive reactions than adolescents with a low believe in a just world. It was concluded that a belief in a just world directed at oneself functioned as a protective factor against the HAB.

In summary, these studies suggest that mindfulness, intelligence, low levels of impulsivity and high believes in a just world could be protective for the development of the HAB. It is noteworthy that the associations found in these studies were only examined once and only regarding the HAB. So research on these topics in association with the HIB, HPB, and HEB is extremely scarce. Importantly, also the core topics discovered in the literature were all found to be associated with the HAB. No literature was found on the association between personality characteristics, gender differences, the role of parents and peers,

maltreatment, neural and biological underpinnings, and the role of emotion and the HIB, HPB or HEB. The same was the case for studies on the role of attention allocations and interventions targeting the HPB or the HEB. The HPB was the only bias that was not studied in relation to media violence. Taken together, these results suggest that all hostility biases are important phenomena's in understanding aggressive behavior, but future research should elucidate whether all biases are related to similar domains.

3.11 Interim Summary

The reviewed studies show that there is a robust relationship between the HAB, the HIB, and aggressive behavior, but that there is a problematic lack of studies on the HEB and HPB. Another important issue is that hostility biases were measured on continuous scales. Hence, it is only possible to refer to hostility biases in terms of high/low levels. No cut-off and/or norm scores are available to determine whether an individual displays a bias or not. Additionally, another robust effect that was found for the HAB and HIB is the association with high emotionality and poor emotion regulation. The relative over-focus on the HAB makes it difficult to discover characteristics that may be unique to each of the biases. It is important to keep in mind that the current review only included articles about the HIB published from 2015 onwards, and it could be argued that the time constraint may have restricted the scope of our review. But, the current review on the HIB builds upon a previous review on this topic (Mellentin et al., 2015), and the conclusions are based on the aggregated results. Still, the current knowledge on the biases has provided sufficient fuel for the development of interventions. In the next section, we will discuss interventions studies that aimed to understand how hostility biases can be altered and whether a reduction in a hostility bias is associated with a decrease in aggression.

3.12 Intervention

Our systematic literature search indicated that two cognitive bias modification programs (CBM) have been developed that aimed to alter the HAB, and four CBM were focused on altering the HIB. We also identified one intervention to alter the HAB that was based on self-persuasion techniques instead of CBM. The notion behind CBM is to expose individuals to an experimentally established contingency during performance of a simple task which is designed to attenuate the target selective processing bias (Koster, Fox, & MacLeod, 2009). In a study by Vassilopoulos, Brouzos, and Andreou (2015), participants were randomly allocated to the training or control condition. Participants in the training condition read 45 descriptions of hypothetical social events across three sessions (15 descriptions in each session). During the session, each subject received a pack of 15 flashcards with the event descriptions printed on them and was asked to read one description at a time and answer the question that followed. After circling their chosen response, participants turned the card over and saw the required response (benign attribution) printed on the back with a "correct" feedback message on top of it. No explanation for the correct response was provided. Before turning to the next card, children were asked to "take a moment to think about the correct explanation". They then repeated this procedure for the rest of the cards. The results showed that the levels of the HAB decreased while benign attributions increased during the training condition (r = .74), but not in the control condition (r = .57). Furthermore, a reduction of perceived anger, aggressive behavior, and an increase in self-control after training was found (r ranging from .39 to .50). This pilot study showed that the HAB can be altered by using a CBM procedure among children relatively high in aggressive behavior, and that a reduction of this bias was associated with a reduction in anger and aggression.

The CBM developed by AlMoghrabi, Huijding, and Franken (2018) consisted of a single session with 40 images of hypothetical ambiguous situations presented on a computer screen.

The scenarios depicted how one individual was harmed by another. All images were preceded by a short description of the situation. The image was followed by the question: why did this happen? After that, a hostile as well as a non-hostile attribution appeared on the screen. Participants were asked to choose the attribution they considered to be most likely. The training was offered in two conditions: positive versus negative training. In the positive condition, the non-hostile attribution was reinforced as "correct" by presenting the word correct on the screen after the non-hostile attribution was chosen, whereas in the negative condition the hostile attribution was "correct". The CBM was conducted among 40 male and 40 female undergraduate students. Aggressive behavior was measured using self-report measures and a behavioral computer task. The results showed that a single session of positive attribution training resulted in an increase in prosocial interpretation bias, less anger and aggression and more happiness (r = .44; r = -.33; r = -.34, respectively). The individuals trained in the negative attribution condition did not show a change in their attribution bias. But, the better the participants performed in this negative condition, the more aggressive responses they showed in an aggression computer task. The authors concluded that their study provided evidence that an attribution bias can be modified in a positive direction by using a CBM, and that this procedure had positive effects on aggressive behavior and on mood.

Hawkins and Cougle (2013) developed and investigated a CBM to alter the HIB in order to influence anger reactivity. Sixty-four ambiguous and 24 control (filler) scenarios were created for this purpose. Scenarios appeared on a computer screen. Participants were asked to read them and imagine themselves in these situations. All scenarios were one sentence long, e.g., "You are walking down the hallway and someone bumps into you." Following the first sentence sketching the scenario, another sentence appeared on the screen to provide a less ambiguous interpretation of the scenario. However, one letter was missing from the key word of the second sentence. In the negative interpretation condition, the second sentence was:

"This person is aggre_sive", and in the positive interpretation condition the sentence was: "This person is clu_sy." The participants were asked to fill in the missing letter of the word (to form the words "aggressive" or "clumsy") in order to assign the positive or negative interpretation to the situation. Subsequently, this interpretation was reinforced by requiring the participants to correctly answer "yes" or "no" to a comprehension question ("Did this person intend to bump you?") before they proceeded to the next scenario.

The results showed that positive interpretations increased from pre- to post-training for the positive training group compared to the negative training group (r = .20). Negative interpretations increased from pre- to post-training for the negative training group relative to the positive training and control group (r = .30 and r = .21, respectively). Furthermore, it was found that participants in the positive condition reported less anger during an insult paradigm compared to the participants in the control condition, and they showed less irritation relative to the participants in the negative training condition. Only the participants in the positive training condition showed a change in the levels of the HAB that was negatively associated with state anger during (r = -.37) and immediately after an insult paradigm (r = -.39). The authors concluded that the positive training might be a promising treatment option for individuals with anger problems.

Cougle et al. (2017) investigated a CBM among individuals with high levels of trait anger and diagnosed with an alcohol use disorder. For the training, participants were instructed to read and imagine themselves in ambiguous scenarios presented on their computer screen; each session consisted of 64 trials (512 total trials over 8 sessions). Each trial started by showing an ambiguous, anger-related scenario (e.g., "You speak to someone and they do not respond."), followed by another sentence to reinforce a benign interpretation. In this sentence, the key word appeared with a missing letter (e.g., "This person is unaw_re"). Participants were then asked to type the missing letter (e.g., "a" for "unaware"). Benign/non-hostile

interpretations were further reinforced via a comprehension question (e.g., “Did this person hear you?”), to which participants responded with “yes” or “no.” If their answer confirmed the benign interpretation or rejected the hostile interpretation, they proceeded to the next trial. Otherwise, they were asked to “try again.” Each of the trials used unique scenarios that were presented randomly. The scenarios in each session covered a wide range of anger-related themes, including feeling ignored, feeling criticized, having one's path blocked, feeling unappreciated, driving situations, feeling disrespected, physical missteps (e.g., being bumped into), disagreements, and annoying traits of others.

Participants received weekly emails with two treatment/training sessions during a month. The patients followed the CBM, the control participants had to watch videos that discussed self-care and healthy habits. The general results showed that the levels of the HIB reduced in both groups. However, the reduction in the CBM condition was significantly larger (r = .39) at post- and follow-up measurement. The effect of the intervention was still present at one-month follow-up (r = .38). Furthermore, it was found that a benign interpretation bias increased in the CBM condition at post- and follow-up measurement. Again, this change was greater as compared to the change in the other condition (r = .49). In addition, the intervention had an effect on anger expressions and a marginally significant effect on trait anger. No direct effects on drinking outcomes were found at one month follow-up. However, it was found that reductions in trait and state anger accounted for reductions in drinking to cope with anger. Based on these findings, the authors concluded that the CBM seems to be a promising intervention.

Another study by Wilkowski, Crowe, and Ferguson (2015) investigated whether a CBM to recruit cognitive control (defined as the down-regulation of retaliatory impulses) was associated with having lower levels of the HIB. A version of the flanker task was used in which the central letter was always perceptually incongruent with the flanker letters (i.e., they

were always different letters). A hostile or non-hostile prime word was presented before each stimulus. Participants were asked to determine which category the prime belonged to by pressing either the left or right arrow key. Participants in the CBM condition were then told that whenever they saw a cleaning-related (i.e. non-hostile) word, it was highly likely that a congruent trial (where the letters both indicate the same response) would follow. They were also told that whenever they saw an aggression-related word it was highly likely that an incongruent trial (where the letters indicate different responses) would follow next on the flanker task. Participants in the non-modification condition were told that there would be absolutely no relationship between the type of word they saw on a trial.

The findings showed that participants in the CBM condition displayed a smaller flanker effect following hostile primes (r = .22), suggesting that the CBM successfully encouraged participants to exert cognitive control. Furthermore, it was found that the CBM reduced aggression in participants prone to show the HIB. A slight increase in aggressive behavior after CBM was found in participants with low levels of the HIB. An association was also found between hostile interpretations and higher revenge motivation, but the CBM did not affect this association.

Hiemstra, Orobio de Castro, and Thomaes (2018) investigated whether a CBM reduced the HIB in a clinical group of aggressive boys. A computer task was used in which photos of angry and happy facial expressions of boys (aged 10-15 years) were morphed, creating images of different ambiguity levels. These images were randomly presented on a computer screen for 500ms and participants were asked to indicate whether the face in an image was either happy or angry. This task was used as baseline measurement. Subsequently, participants received computer-generated feedback to signal either “correct” or “incorrect” answers during the training phase, which lasted five consecutive days. The feedback was based on the baseline measurement. Feedback was not provided in the control condition, and

participants were randomly allocated to one of the conditions. It was found that the CBM procedure significantly decreased the levels of the HIB (r = .79), and these results were replicated in a second study (r = .71). The second study was identical to the first one, except that it included a shorter training period (three days instead of five).

Finally, van Dijk, Thomaes, Poorthuis, and Orobio de Castro (2018) investigated whether self-persuasion could decrease the HAB. Self-persuasion involves asking individuals to publicly advocate against their own beliefs. Children assigned to the experimental condition were asked to endorse non-hostile attributions of ambiguous provocations in a video message. Children in the control condition were asked to describe the ambiguous provocations. Assessments took place one month before the recording of the video messages and immediately thereafter. The results showed that children in the experimental condition displayed a reduction of the HAB from pre- to post-intervention (r = .45), but they did not display less aggressive behavior (r = .28). A second study included an additional experimental condition: other-persuasion, in which children were asked to describe the ambiguous provocations. Subsequently, the children were told that it would be better if other children would see how the experimenter explained benign attributions. The child then observed the experimenter endorse two non-hostile attributions of ambiguous provocations in a video message. The results of this study showed that children in the other-persuasion condition had lower levels of the HAB than children in the control condition (r = .57), but also did not display less aggressive behavior (r < .10). Taken together, the results indicated that both self- and other-persuasion of benign attributions were effective in reducing the HAB.

In summary, the reviewed studies suggest that CBM seems to be a suitable paradigm to alter the HAB as well as the HIB. However, more research is needed on whether this paradigm is also applicable to the HPB and HEB as well, as there is a need for replication. Next, we will integrate the findings concerning associations between the biases and

aggression that was found in the present review.

4. Discussion

4.1. Hostility biases and aggression

In line with previous meta-analyses and reviews (Bushman, 2016; Dodge, 2006; Mellentin et al., 2015; Orobio de Castro et al., 2002; Tuente et al., 2019), we suggest that the association between aggressive behavior and hostility biases is robust. However, because the HAB is the most researched bias, the HIB, HPB and especially the HEB remain heavily understudied. The relative over-focus on the HAB makes it difficult to discover characteristics that may be unique to each of the biases. Nevertheless, the general patterns of associations between all hostility biases and aggressive behavior could also indicate the existence of a general hostile construct instead of distinct biases. Also, the similarities between the techniques used to alter hostility biases might provide some preliminary evidence for the existence of a general construct. The contingency used in the CBM paradigms were similar across studies: positive vs. negative, or non-hostile/benign vs. hostile. Training in the positive/benign domain resulted in reductions of the HAB, HIB, anger, aggressive behavior and irritation, but also an increase in prosocial interpretations.

We suggest that the hostility bias in social information processing could be an important cause and contributing factor to the development and persistency of aggressive behavior. It is important to note that most of the studies included in the current and previous reviews were cross-sectional. Hence, no causal conclusions can be drawn, and future research should elucidate the nature of this relationship in further detail. Nonetheless, hostility biases are thought to be important constructs for the understanding and treatment of aggression. Before hostility biases can be targeted in clinical settings, it is of importance to understand how they are acquired. Therefore, it is necessary to mechanistically comprehend their underpinnings,

which are currently unknown. This is surprising, as these biases are thought to be cognitive phenomena which cannot be considered as stand-alone processes. Also, the CBM procedures developed to alter hostility biases aim to restructure the incorrect cognitive processing of social stimuli without awareness of underlying elements that facilitate, strengthen or contribute to the occurrence of the biases. The treatment of hostility biases may just be symptom management without a proper understanding of these underlying elements. An enhanced understanding of the mechanism behind hostility biases is also needed to comprehend its association with aggressive behavior. Even the most frequently used model to explain the role of biased social information processing in aggression inadequately explains this underlying mechanism.

The Social Information Processing model (SIP; Crick & Dodge, 1994) assumes that, based on early experiences, hostile schemas are stored in memory and affect the way social information is processed. Furthermore, social cues have to be processed in consecutive steps in order to react appropriately in social situations: 1) encoding of cues; 2) interpretation of cues; 3) goal clarification; 4) generating response alternatives; 5) evaluation of response alternatives and selection of an optimal response; and 6) enactment of the optimal response. The SIP model was developed as a framework using cognition as the main level of explanation, disregarding the role of emotions. Based on the results of the present review, we suggest that emotion could be one of the elements of the mechanism behind hostility biases. The current review revealed that ten studies have been conducted about the relation between emotion and the HAB. All of the studies found results pointing in the same direction; that the HAB had a negative association with emotion understanding and a positive association with emotion sensitivity. The positive association was further supported by the finding that hostile attributions increased after a negative mood induction. Moreover, aggressive individuals not

only exhibited the HAB, but were also thought to experience more anger and engage less in emotion regulation strategies.

More evidence for the role of emotion is provided by studies that focused on hostility biases and aggression subtype. The HAB and HIB, in particular, were both repeatedly found to be associated with reactive aggression. Reactive aggressive behavior is defined as impulsive, angry or defensive responses to threat, frustration or provocation (Dodge & Coie, 1987). Importantly, reactive aggression is also referred to as spontaneous and emotionally driven forms of aggression, i.e. affective aggression (e.g. Cima & Raine, 2009). The results of the current review, therefore, suggest that the emergence of this hostile construct is more likely to occur in individuals who display impulsive behavior rather than in individuals who display planned, manipulative, callous, and deliberate behavior. One explanation for how emotions may affect social information processing is offered in the revised version of the SIP model (Lemerise & Arsenio, 2000), which will be discussed next.

4.2. Revised SIP model and schema inconsistency hypothesis

In their revised version of the SIP model, Lemerise and Arsenio (2000) hypothesized that each step of social information processing can be affected by individual differences in emotionality and emotion regulation. Besides one's own emotions, they stressed the necessity to encode and interpret others affective signals as they, in combination with one's own affective cues, provide ongoing information about how the social encounter is proceeding. This information allows us to make subtle adjustments to behavior.

Additionally, one's own emotions may influence encoding and interpretation processes. Current emotions, mood or arousal may influence what is noticed about a social encounter, which in turn make the recollection of mood-congruent information more likely. For instance, if someone is in an irritated or aggressive mood, one is more likely to attribute or interpret

social stimuli or interactions in a hostile way. Lemerise and Arsenio (2000) also suggested that the type of goals that will be selected can be affected by the intensity with which one experiences emotions and the ability to regulate emotions. If one is overwhelmed by his/her own and/or others' emotions, avoidant or hostile goals might be chosen to reduce arousal. Furthermore, poor regulatory abilities might interfere with the assessment of the situation from a different affective and cognitive perspective. This may hamper a flexible approach to goal selection, for which contextual factors have to be taken into account.

In addition, it was proposed that individuals who experience strong emotions might be too overwhelmed and too self-focused to generate a diversity of responses and to evaluate responses from all possible perspectives. Furthermore, the ability to flexibly display emotions appropriate to the situation requires control over one's expressivity as well as over one's sensitivity to the situation. Individuals with deficits in reading and sending affective signals may be dependent of relatively rigid approaches to situations.

Based on the original and revised SIP model, it seems plausible that social information processing occurs in sequential steps, in order to react appropriately in social situations. An individual's database of memories of past experiences affects each of these steps. Past experiences may contribute to the development of hostile schemas which are stored in one's memory. An early maladaptive schema is defined as a broad, pervasive theme or pattern, comprised of memories, emotions, bodily sensations, and cognitions, and regard to oneself and one's relationships with others, that are dysfunctional to a significant degree (Young, Klosko, & Weishaar, 2003). Furthermore, these schemas evolve throughout a lifetime into an automatic and unconscious set of tendencies. One's emotionality and ability to regulate emotions may strengthen these schemas. This hampers reappraisal, ultimately leading to an overreliance on immediate appraisal and impulsive responses.

How exactly hostile schemas result in the incorrect processing of neutral or ambiguous stimuli, however, is inadequately explained by the original and revised SIP model. An understanding of this mechanism is of importance as hostility biases are particularly thought to occur in response to ambiguous stimuli. The 'schema inconsistency hypothesis' proposed by Horsley et al. (2010) complements both SIP models by introducing an explanation for this mechanism.

This hypothesis postulates that hostile schemas direct attention towards schema-inconsistent information, away from the expected hostile cues. Subsequently, it is assumed that the ambiguity of social cues and the strength of hostile schemas determine whether these cues are interpreted accurately. Paying attention towards schema-inconsistent information often requires more time. This longer attention allocation is thought to reflect an attempt to verify unexpected information considering an already existing interpretation of the situation (Horsley et al., 2010). As a consequence, schema-based perception makes it more difficult to encode and process schema-inconsistent information because it does not fit the expectations. This, in turn, makes it harder to take this information into account in interpreting the situation. A benign interpretation, attribution, perception or expectation of the social encounter will then be highly unlikely, especially when emotions are involved. Some first evidence in favor of this hypothesis was provided by a few studies that used eye-tracking and found that aggressive individuals had longer gaze duration for non-hostile cues and that they looked back longer at more non-hostile information, provided on pictures that were presented earlier but remained visible. In contrast, non-aggressive individuals looked back longer at more hostile information (Horsley et al., 2010; Lin et al., 2016; Wilkowski et al., 2007).

Taken together, the revised SIP model and the schema inconsistency hypothesis add additional explanatory levels to understand the occurrence of hostility biases. However, the models have not been able to account for all the evidence and it is still unknown how exactly

hostility biases are acquired. Other issues are that 1) there is a tendency to study hostility biases separately, as if they are non-interacting phenomena and, 2) the fact that current approaches cannot directly quantify the latent cognitive processes pertaining to the hostility biases, thus creating an explanatory gap. To bridge this gap, we introduce the Computations of Hostile Biases (CHB) model which we will elaborate in the next section.

4.3. CHB-model: Grounding hostility biases in a unified neuroscientific framework

We propose that the HAB, HIB, HPB, and HEB are not distinct phenomena, but are manifestations of one general hostility bias mechanism. This mechanism can be seen as part of the encoding step of the original SIP-model and is accountable for distortions in several social information processes. It is proposed that individuals prone to react aggressively enter a social situation with a database of memories of past experiences, social schemas and social knowledge. This knowledge affects the encoding of social cues. Additionally, someone will rely more on these schemas in ambiguous situations as in such environments there is 1) a high uncertainty about the meaning of social cues, and 2) there is a possibility that the meaning of these cues changes with time. Once hostile schemas exist, the likelihood of hostile encoding increases. Cues that are inconsistent with the hostile schemas are not taken into account and are (partly) filtered out. High emotionality acts as a moderator that facilitates and strengthens this process. As a consequence, hostile encoding becomes predominant, leading to biased attention, sensation, and perception of social cues. This introduces biases in subsequent higher-order interpretive processes, such as intent attributions and interaction evaluations. Once the encoding step is biased, every other subsequent stage will be affected too (see Figure 2). In other words, interactions, intentions, goals, and eventually behavior will all be distorted. Depending on what kind of measurements are used, e.g. vignettes or morphed face tasks, different aspects of the interpretive processes will be targeted. Thus, there is a

hierarchical relationship between the stages of processing, with information cascading from sensory processing to higher order cognitive control processes.

Previous models to explain aggressive behavior by social information processing deficits were influential but descriptive, and lack a firm embedding in general cognitive frameworks. The latter is necessary to systematically study and understand the mechanisms driving aggression. Current approaches provide information about which social cues lead to what kind of bias, but there is a lack of insight into the latent cognitive processes and the contribution of each of these processes to observed behavior. To completely understand the development of an individual's aggression we need to understand the unique contribution of each explaining characteristic, implying that a clear mapping across different levels of description is required. We argue that a novel model is direly needed to further understand and study hostility biases, which will be discussed in the next section.

4.3.1. A computational account for hostility biases.

An emerging approach with an increasing impact on the field of cognitive neuroscience is computational psychiatry (Wiecki et al., 2015). The notion is to construct models based on integrated evidence from psychology and neuroscience to explain cognitive processes, behavior, and neural activity. Such an approach describes information processing in terms of basic principles of cognition, and uses mathematical approaches to directly quantify the engagement of cognitive processes. This also enables us to quantify fundamental elements of cognition that usually are difficult to observe (i.e. latent variables), which makes it extremely useful for studying psychiatric populations (Stephan & Mathys, 2014). This is also highly consistent with the most recent endeavors in psychiatry, such as the Research Domain Criteria (RDoC) initiative of the National Institute of Mental Health (Insel et al., 2010). Furthermore, such models enable researchers to generate precise predictions and to quantitatively test

competing hypotheses (Busemeyer & Diederich, 2010; Forstmann & Wagenmakers, 2015; Lewandowsky & Farrell, 2010). The advantages of using this approach to explain the hostility bias are that 1) the bias is anchored in a well-defined theoretical framework that integrates separate cognitive processes and their interactions, 2) it offers a clearly defined mathematical translation of how the corresponding cognitive computations take place and interact, 3) and it is supported by the most recent neuroscientific insights.

One example of such a framework is the hierarchical Gaussian filter (HGF; Mathys, Daunizeau, Friston, & Stephan, 2011; Mathys et al., 2014). The HGF is a generic hierarchical Bayesian model of learning under perceptual uncertainty and environmental changes using time-series data (Mathys et al., 2011; Mathys et al., 2014). This model is based on the idea that the brain continuously creates a generative (i.e. predictive) model of its sensory inputs and tries to optimize this model by reducing uncertainty (i.e., increasing the accuracy) about the beliefs of the world (see e.g. Brazil, Mathys, Popma, Hoppenbrouwers, & Cohn, 2017; Mathys et al., 2014). Uncertainty can be reduced by either the adjustment of beliefs about the world (perception) or by changing the way the world is sampled (Friston, 2010; Stephan & Mathys, 2014). In light of the evidence pointing towards hierarchical stages of cognitive processing during social encounters, we propose that considering the hostility bias in the context of the HGF offers a clear and well-defined framework for explaining and studying distorted social information processing. Our integrative view on the hostility bias (described in the aforementioned CHB-model) can be directly mapped to the theoretical framework formalized through the HGF (see Figure 3).

Consider the following scenario: John is walking down a shopping street. He had a rough night and is in an irritated mood. In the distance, a man is walking in his direction. “Why is he looking at me?” John thinks. “He grins at me! What does he want from me?” When the man passes John by, he bumps into John quite firmly. John yells: “Watch out! Why did you do

that?! You did that on purpose! Do you want to fight?". In this case, the decision whether the intent of someone's action was hostile and the verbal aggressive response of John are observable whereas the underlying cognitive processes are latent. The notion behind the HGF is that choice-behavior is driven by a hierarchy of interacting latent cognitive processes. The hostility bias needs to be disentangled into smaller cognitive units that need to interact so that we have an optimal representation of events that occur in the world in order to learn to make optimal choices. Some of these hidden computations represent an individual's expectations about the likelihood of an event occurring based on past experiences ("latent computation 1"), for instance, how often someone who grinned at you in the past had bad intentions or how often someone's intentions when bumping into you were hostile vs. benign. This also depends on the uncertainty about the meaning of the social cues occurring in the environment (i.e. situations are not always straightforward and often consist of ambiguous information; "latent computation 2") and the individual's perceptual uncertainty (i.e. how confident is someone about what he/she is perceiving; "latent computation 3"). When a subject is uncertain about the correctness of his/her representation of the world, one may be highly likely to use previous beliefs to guide the current representation (e.g. based on hostile schemas). Additionally, these contingencies are prone to change. This means that, previously, the person who bumped into you had a hostile intention but someone else may have not. To keep making optimal choices we need to learn how likely it is that these contingency changes may occur ("latent computation 4"), but also at which rate these changes may occur ("latent computation 5").

Additional hidden computations represent an individual's schema-based perception tendency. For instance, how persistent is this tendency ("latent computation 6") and is an individual capable to explore and rely on schema inconsistent information ("latent computation 7")? The final hidden computations proposed by the CHB-model represent an

individual's emotionality. For example, how much does John's mood affect his decision that the intention of the man was hostile? This question basically concerns the relative weight that John's irritated mood had on how information was processed ("latent computation 8").

The HGF has received recognition in (social) neuroscience and has already been applied to investigate cognitive processes that are engaged during social information processing. For instance, in two studies Diaconescu et al. (2014; 2017) found that individuals employ hierarchical processing to make inferences about the changing intentions of others, and that we use knowledge about these changes in intentionality (i.e. volatility) to inform decision-making during social interactions.

Thus, we propose a novel approach for understanding hostility biases in which the behavioral correlates of hostility stem from a single bias mechanism. This mechanism can be differentially triggered and expressed by several factors. For instance, asking to indicate the intention of someone in an ambiguously written vignette will capture the attribution bias instead of any of the other expressions of the core bias. The uniqueness of this model is that it bridges different levels of explanations (e.g. schema-based perception, emotion). The strength of including computational modelling is that it allows to unravel the sources of the hostility bias. These sources can be captured with such a high precision that it subsequently could lead to the development of more targeted interventions for each individual.

5. Conclusion

The current systematic review confirms the robust association between hostility biases and aggressive behavior. This review complements previous ones by providing an overview of literature on all hostility biases. Despite the large amount of studies dedicated to this topic, still a lot remains unclear. Such as how hostility biases develop over age and whether aggressive behavior results from the hostility bias or that aggression contributes to the

development of the hostility bias. Hence, there is a critical need for replication studies and studies investigating all hostility biases in one design.

Based on previous models and the results of the current review, the CHB-model is proposed. In this model, we suggest that a general hostile mechanism in the encoding of social information introduces biases in subsequent higher-order interpretive processes. The uncertainty of social cues, the volatility of these cues and emotionality all are proposed to be elements of this hostile mechanism. To examine this notion prospectively, we propose computational modelling (e.g. using the HGF) as a novel approach to understanding and studying hostility biases, thus offering the unique opportunity to quantify the latent cognitive processes sub-serving hostility biases. Using this approach may result in a better understanding of the underlying causes of aggressive behavior. This, in turn, might be beneficial for clinical practice as it would be possible to develop and target interventions more specifically towards the individual patient's needs (Brazil et al., 2018), ultimately resulting in the understanding of an individual's aggression at the level of the underlying causes.

References

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. doi: 10.1016/j.jbtep.2017.08.003

Bailey, C. A., & Ostrov, J. M. (2007). Differentiating forms and functions of aggression in emerging adults: Associations with hostile attribution biases and normative beliefs. *Journal of Youth and Adolescence, 37*(6), 713-722. doi:10.1007/s10964-007-9211-5

Bartholow, B. D., & Heinz, A. (2006). Alcohol and aggression without consumption: Alcohol cues, aggressive thoughts, and hostile perception bias. *Psychol Sci, 17*(1), 30-37. doi: 10.1111/j.1467-9280.2005.01661.x

Bègue, L., & Muller, D. (2006). Belief in a just world as moderator of hostile attributional bias. *British journal of social psychology, 45*(1), 117-126. doi: 10.1348/014466605X37314

Bondü, R. (2018). Is bad intent negligible? Linking victim justice sensitivity, hostile attribution bias, and aggression. *Aggress Behav, 44*(5), 442-450. doi:10.1002/ab.21764

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. *Front Psychol, 7*. doi:10.3389/fpsyg.2016.00795

Boulton, M. J. (2012). Children's hostile attribution bias is reduced after watching realistic playful fighting, and the effect is mediated by prosocial thoughts. *Journal of experimental child psychology, 113*(1), 36-48. doi:10.1016/j.jecp.2012.02.011

Bowen, K. N., Roberts, J. J., & Kocian, E. J. (2016). Decision making of inmates: Testing social information processing concepts using vignettes. *Applied Psychology in Criminal Justice, 12*(1), 1-17.

Bratton, H., O'Rourke, S., Tansey, L., & Hutton, P. (2017). Social cognition and paranoia in forensic inpatients with schizophrenia: A cross-sectional study. *Schizophrenia research, 184*, 96-102. doi:10.1016/j.schres.2016.12.004

Brazil, I. A., Mathys, C. D., Popma, A., Hoppenbrouwers, S. S., & Cohn, M. D. (2017). Representational uncertainty in the brain during threat conditioning and the link with psychopathic traits. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(8), 689-695. doi:10.1016/j.bpsc.2017.04.005

Brazil, I. A., van Dongen, J. D., Maes, J. H., Mars, R., & Baskin-Sommers, A. R. (2018). Classification and treatment of antisocial individuals: From behavior to biocognition. *Neuroscience & Biobehavioral Reviews, 91*, 259-277. doi:doi.org/10.1016/j.neubiorev.2016.10.010

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Thousand Oaks, CA: Sage.

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

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. *Aggress Behav, 38*(1), 47-63. doi:doi.org/10.1002/ab.21407

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. doi:10.1037/a0032293

Choe, D. E., Shaw, D. S., & Forbes, E. E. (2015). Maladaptive social information processing in childhood predicts young men's atypical amygdala reactivity to threat. *Journal of Child Psychology and Psychiatry, 56*(5), 549-557. doi:10.1111/jcpp.12316

Cillessen, A. H., Lansu, T. A., & Van Den Berg, Y. H. (2014). Aggression, hostile attributions, status, and gender: A continued quest. *Dev Psychopathol, 26*(3), 635-644. doi:10.1017/S0954579414000285

Cima, M., & Raine, A. (2009). Distinct characteristics of psychopathy relate to different subtypes of aggression. *Personality and Individual Differences, 47*(8), 835-840. doi: 10.1016/j.paid.2009.06.031

Combs, D. R., Penn, D. L., Michael, C. O., Basso, M. R., Wiedeman, R., Siebenmorgan, M., . . . Chapman, D. (2009). Perceptions of hostility by persons with and without persecutory delusions. *Cognitive neuropsychiatry, 14*(1), 30-52. doi: 10.1080/13546800902732970

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. *Behav Res Ther, 99*, 57-66. doi:10.1016/j.brat.2017.09.004

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological bulletin, 115*(1), 74. doi:10.1037/0033-2909.115.1.74

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child development, 67*(3), 993-1002. doi:10.1111/j.1467-8624.1996.tb01778.x

Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child development, 73*(4), 1134-1142. doi:10.1111/1467-8624.00462

Darrell-Berry, H., Bucci, S., Palmier-Claus, J., Emsley, R., Drake, R., & Berry, K. (2017). Predictors and mediators of trait anger across the psychosis continuum: The role of attachment style, paranoia and social cognition. *Psychiatry Res, 249*, 132-138. doi: 10.1016/j.psychres.2017.01.007

Davies, P. T., Coe, J. L., Hentges, R. F., Sturge-Apple, M. L., & Ripple, M. T. (2018). Temperamental Emotionality Attributes as Antecedents of Children's Social Information Processing. *Child development, Advanced online publication*. doi:10.1111/cdev.13191

De La Osa, N., Penelo, E., Navarro, J.-B., Trepat, E., Domenech, J. M., & Ezpeleta, L. (2018). Oppositional Defiant Disorder dimensions and aggression: The moderating role of hostile bias and sex. *Psicothema, 30*(3), 264-269. doi:10.7334/psicothema2017.363

DeWall, C. N., Twenge, J. M., Gitter, S. A., & Baumeister, R. F. (2009). It's the thought that counts: The role of hostile cognition in shaping aggressive responses to social exclusion. *J Pers Soc Psychol, 96*(1), 45. doi:10.1037/a0013196

Diaconescu, A. O., Mathys, C., Weber, L. A., Daunizeau, J., Kasper, L., Lomakina, E. I., . . . Stephan, K. E. (2014). Inferring on the intentions of others by hierarchical Bayesian learning. *PLoS computational biology, 10*(9), e1003810. doi: 10.1371/journal.pcbi.1003810

Diaconescu, A. O., Mathys, C., Weber, L. A., Kasper, L., Mauer, J., & Stephan, K. E. (2017). Hierarchical prediction errors in midbrain and septum during social learning. *Soc Cogn Affect Neurosci, 12*(4), 618-634. doi:10.1093/scan/nsw171

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Dev Psychopathol, 18*(03), 791-814. doi:10.10170S0954579406060391

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *J Pers Soc Psychol, 53*(6), 1146.

Dodge, K. A., & Crick, N. R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin, 16*(1), 8-22. doi: 10.1177/0146167290161002

Dodge, K. A., Malone, P. S., Lansford, J. E., Sorbring, E., Skinner, A. T., Tapanya, S., . . . Al-Hassan, S. M. (2015). Hostile attributional bias and aggressive behavior in global context. *Proceedings of the National Academy of Sciences, 112*(30), 9310-9315. doi: 10.1073/pnas.1418572112

Edwards, R., & Bond, A. J. (2012). Narcissism, self-concept clarity and aggressive cognitive bias amongst mentally disordered offenders. *Journal of Forensic Psychiatry & Psychology, 23*(5-6), 620-634. doi:10.1080/14789949.2012.715180

Ellis, M. L., Weiss, B., & Lochman, J. E. (2009). Executive functions in children: Associations with aggressive behavior and appraisal processing. *Journal of Abnormal Child Psychology, 37*(7), 945-956. doi:10.1007/s10802-009-9321-5

Forstmann, B. U., & Wagenmakers, E.-J. (2015). *Model-based cognitive neuroscience*. New York, YW: Springer.

Freeman, K., Hadwin, J. A., & Halligan, S. L. (2011). An experimental investigation of peer influences on adolescent hostile attributions. *Journal of Clinical Child & Adolescent Psychology, 40*(6), 897-903. doi:10.1080/15374416.2011.614582

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127. doi:10.1038/nrn2787

Gagnon, J., Aubin, M., Emond, F. C., Derguy, S., Brochu, A. F., Bessette, M., & Jolicoeur, P. (2017). An ERP study on hostile attribution bias in aggressive and nonaggressive individuals. *Aggress Behav, 43*(3), 217-229. doi:doi.org/10.1002/ab.21676

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

Galán, C. A., Choe, D. E., Forbes, E. E., & Shaw, D. S. (2017). The interaction between monoamine oxidase A and punitive discipline in the development of antisocial behavior: Mediation by maladaptive social information processing. *Dev Psychopathol, 29*(4), 1235-1252. doi:10.1017/S0954579416001279

Gentile, D. A., Coyne, S., & Walsh, D. A. (2011). Media violence, physical aggression, and relational aggression in school age children: a short-term longitudinal study. *Aggress Behav, 37*(2), 193-206. doi:10.1002/ab.20380

Godleski, S. A., & Ostrov, J. M. (2010). Relational aggression and hostile attribution biases: Testing multiple statistical methods and models. *Journal of Abnormal Child Psychology, 38*(4), 447-458. doi:10.1007/s10802-010-9391-4

Godleski, S. A., Ostrov, J. M., Houston, R. J., & Schlienz, N. J. (2010). Hostile attribution biases for relationally provocative situations and event-related potentials. *International Journal of Psychophysiology, 76*(1), 25-33. doi:10.1016/j.ijpsycho.2010.01.010

Halligan, S. L., Cooper, P. J., Healy, S. J., & Murray, L. (2007). The attribution of hostile intent in mothers, fathers and their children. *Journal of Abnormal Child Psychology, 35*(4), 594-604. doi:10.1007/s10802-007-9115-6

Halligan, S. L., & Philips, K. J. (2010). Are you thinking what I'm thinking? Peer group similarities in adolescent hostile attribution tendencies. *Developmental Psychology, 46*(5), 1385. doi:10.1037/a0020383

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. doi:10.1016/j.beth.2013.04.005

Helfritz-Sinville, L. E., & Stanford, M. S. (2014). Hostile attribution bias in impulsive and premeditated aggression. *Personality and Individual Differences, 56*, 45-50. doi:10.1016/j.paid.2013.08.017

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. doi:10.1007/s10802-015-0027-6

Heppner, W. L., Kernis, M. H., Lakey, C. E., Campbell, W. K., Goldman, B. M., Davis, P. J., & Cascio, E. V. (2008). Mindfulness as a means of reducing aggressive behavior: Dispositional and situational evidence. *Aggress Behav, 34*(5), 486-496. doi: 10.1002/ab.20258

Hiemstra, W., Orobio de Castro, B. O., & Thomaes, S. (2018). Reducing Aggressive Children's Hostile Attributions: A Cognitive Bias Modification Procedure. *Cognitive Therapy and Research*, 1-12. doi:10.1007/s10608-018-9958-x

Horsley, T. A., Orobio 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. doi: 10.1007/s10802-009-9361-x

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *The American Joural of Psychiatry, 167*, 748-751. doi: 10.1176/appi.ajp.2010.09091379

Jahoda, A., Pert, C., & Trower, P. (2006). Frequent aggression and attribution of hostile intent in people with mild to moderate intellectual disabilities: An empirical investigation. *American journal on mental retardation, 111*(2), 90-99. doi:10.1352/0895-8017(2006)111[90:FAAAOH]2.0.CO;2

Jin, X., Eagle, M., & Keat, J. E. (2008). Hostile attributional bias, early abuse, and social desirability in reporting hostile attributions among Chinese immigrant batterers and nonviolent men. *Violence and Victims, 23*(6), 773-786. doi:10.1891/0886-6708.23.6.773

Jusyte, A., & Schönenberg, M. (2017). Impaired social cognition in violent offenders: perceptual deficit or cognitive bias? *Eur Arch Psychiatry Clin Neurosci, 267*(3), 257-266. doi:10.1007/s00406-016-0727-0

Karadenizova, Z. M., & Dahle, K.-P. (2018). It is written in your eyes: hostile attributions and self-directed gaze perception in incarcerated violent adolescent male offenders. *Int J Offender Ther Comp Criminol, 62*(12), 3623-3638. doi:10.1177/0306624X17746292

Kay, C. L., & Green, J. M. (2016). Social cognitive deficits and biases in maltreated adolescents in UK out-of-home care: relation to disinhibited attachment disorder and psychopathology. *Dev Psychopathol, 28*(1), 73-83. doi:10.1017/S0954579415000292

Kokkinos, C. M., Karagianni, K., & Voulgaridou, I. (2017). Relational aggression, big five and hostile attribution bias in Adolescents. *Journal of Applied Developmental Psychology, 52*, 101-113. doi:10.1016/j.appdev.2017.07.007

Koster, E. H., Fox, E., & MacLeod, C. (2009). Introduction to the special section on cognitive bias modification in emotional disorders. *J Abnorm Psychol, 118*(1), 1. doi: 10.1016/j.appdev.2017.07.007

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. doi:10.1371/journal.pone.0187080

Leff, S. S., Baker, C. N., Waasdorp, T. E., Vaughn, N. A., Bevans, K. B., Thomas, N. A., . . . Monopoli, W. J. (2014). Social cognitions, distress, and leadership self-efficacy: Associations with aggression for high-risk minority youth. *Dev Psychopathol, 26*(3), 759-772. doi:10.1017/S0954579414000376

Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child development, 71*(1), 107-118. doi:10.1111/1467-8624.00124

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. Thousand Oaks, CA: Sage.

Li, C., Sun, Y., Ho, M. Y., You, J., Shaver, P. R., & Wang, Z. (2016). State narcissism and aggression: The mediating roles of anger and hostile attributional bias. *Aggress Behav, 42*(4), 333-345. doi:10.1002/ab.21629

Lin, P.-I., Hsieh, C.-D., Juan, C.-H., Hossain, M. M., Erickson, C. A., Lee, Y.-H., & Su, M.-C. (2016). Predicting aggressive tendencies by visual attention bias associated with hostile emotions. *PLoS One, 11*(2), e0149487. doi:10.1371/journal.pone.0149487

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. doi:10.1521/pedi.2013.27.1.53

MacBrayer, E. K., Milich, R., & Hundley, M. (2003). Attributional biases in aggressive children and their mothers. *J Abnorm Psychol, 112*(4), 698-708. doi:10.1037/0021-843X.112.4.698

Mammen, O., Kolko, D., & Pilkonis, P. (2003). Parental cognitions and satisfaction: Relationship to aggressive parental behavior in child physical abuse. *Child maltreatment, 8*(4), 288-301. doi:10.1177/1077559503257112

Martins, N. (2013). Televised relational and physical aggression and children's hostile intent attributions. *Journal of experimental child psychology, 116*(4), 945-952. doi: 10.1016/j.jecp.2013.05.006

Matheny, N. L., Smith, H. L., Summers, B. J., McDermott, K. A., Macatee, R. J., & Cougle, J. R. (2017). The role of distress tolerance in multiple facets of hostility and willingness to forgive. *Cognitive Therapy and Research, 41*(2), 170-177. doi:10.1007/s10608-016-9808-7

Mathieson, L. C., Murray-Close, D., Crick, N. R., Woods, K. E., Zimmer-Gembeck, M., Geiger, T. C., & Morales, J. R. (2011). Hostile intent attributions and relational aggression: The moderating roles of emotional sensitivity, gender, and victimization. *Journal of Abnormal Child Psychology, 39*(7), 977. doi:10.1007/s10802-011-9515-5

Mathys, C., Daunizeau, J., Friston, K. J., & Stephan, K. E. (2011). A Bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci, 5*, 39. doi: 10.3389/fnhum.2011.00039

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. *Front Hum Neurosci, 8*, 825. doi:10.3389/fnhum.2014.00825

Matthews, B. A., & Norris, F. H. (2002). When Is Believing "Seeing"? Hostile Attribution Bias as a Function of Self-Reported Aggression 1. *Journal of Applied Social Psychology, 32*(1), 1-31. doi:10.1111/j.1559-1816.2002.tb01418.x

McDermott, K. A., Smith, H. L., Matheny, N. L., & Cougle, J. R. (2017). Pain and multiple facets of anger and hostility in a sample seeking treatment for problematic anger. *Psychiatry Res, 253*, 311-317. doi:10.1016/j.psychres.2017.04.006

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. doi:10.1016/j.avb.2015.09.001

Miller, N. V., & Johnston, C. (2019). Social threat attentional bias in childhood: Relations to aggression and hostile intent attributions. *Aggress Behav, Advanced online publication*, 1-10. doi:10.1002/ab.21813

Möller, I., & Krahé, B. (2009). Exposure to violent video games and aggression in German adolescents: A longitudinal analysis. *Aggress Behav, 35*(1), 75-89. doi: 10.1002/ab.20290

Nasby, W., Hayden, B., & DePaulo, B. M. (1980). Attributional bias among aggressive boys to interpret unambiguous social stimuli as displays of hostility. *J Abnorm Psychol, 89*(3), 459. doi:10.1037/0021-843X.89.3.459

Nelson, D. A., & Coyne, S. M. (2009). Children's intent attributions and feelings of distress: Associations with maternal and paternal parenting practices. *Journal of Abnormal Child Psychology, 37*(2), 223. doi:10.1007/s10802-008-9271-3

Nentjes, L., Bernstein, D., Arntz, A., van Breukelen, G., & Slaats, M. (2015). Examining the influence of psychopathy, hostility biases, and automatic processing on criminal offenders' Theory of Mind. *Int J Law Psychiatry, 38*, 92-99. doi: 10.1016/j.ijlp.2015.01.012

Neumann, D., Malec, J. F., & Hammond, F. M. (2017). Negative attribution bias and anger after traumatic brain injury. *The Journal of head trauma rehabilitation, 32*(3), 197-204. doi:10.1097/HTR.0000000000000259

Orobio 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. doi:10.1207/s15374424jccp3401_10

Orobio de Castro, B., 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. doi: 10.1207/S15374424JCCP3201_06

Orobio de Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child development, 73*(3), 916-934. doi:10.1111/1467-8624.00447

Peets, K., Hodges, E. V., Kikas, E., & Salmivalli, C. (2007). Hostile attributions and behavioral strategies in children: Does relationship type matter? *Developmental Psychology, 43*(4), 889. doi:10.1037/0012-1649.43.4.889

Quan, F., Yang, R., Zhu, W., Wang, Y., Gong, X., Chen, Y., . . . Xia, L.-X. (2019). The relationship between hostile attribution bias and aggression and the mediating effect of anger rumination. *Personality and Individual Differences, 139*, 228-234. doi:10.1016/j.paid.2018.11.029

Rothmann, S., & Coetzer, E. P. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology, 29*(1), 68-74.

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. *Aggress Behav, 43*(4), 386-397. doi:10.1002/ab.21697

Smith, H. L., Summers, B. J., Dillon, K. H., Macatee, R. J., & Cougle, J. R. (2016). Hostile interpretation bias in depression. *Journal of affective disorders, 203*, 9-13. doi: 10.1016/j.jad.2016.05.070

Stephan, K. E., & Mathys, C. (2014). Computational approaches to psychiatry. *Current opinion in neurobiology, 25*, 85-92. doi:10.1016/j.conb.2013.12.007

Teige-Mocigemba, S., Hölzenbein, F., & Klauer, K. C. (2016). Seeing More Than Others. *Social Psychology, 47*, 136-149. doi:10.1027/1864-9335/a000266

Thomas, R. A., & Weston, R. (2019). Exploring the Association between Hostile Attribution Bias and Intimate Partner Violence in College Students: Romantic Relationships and Friends with Benefits. *Journal of Aggression, Maltreatment & Trauma*, 1-20. doi: 10.1080/10926771.2019.1587561

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. doi:10.1016/j.avb.2019.01.009

van Dijk, A., Thomaes, S., Poorthuis, A. M., & Orobio de Castro, B. O. (2018). Can self-persuasion reduce hostile attribution bias in young children? *Journal of Abnormal Child Psychology*, 1-12. doi:10.1007/s10802-018-0499-2

Vassilopoulos, S. P., Brouzos, A., & Andreou, E. (2015). A multi-session attribution modification program for children with aggressive behaviour: Changes in attributions, emotional reaction estimates, and self-reported aggression. *Behav Cogn Psychother, 43*(5), 538-548. doi:10.1017/S1352465814000149

Vassos, E., Collier, D., & Fazel, S. (2014). Systematic meta-analyses and field synopsis of genetic association studies of violence and aggression. *Mol Psychiatry, 19*, 471-477. doi:10.1038/mp.2013.31

Wang, Y., Zhu, W., Xiao, M., Zhang, Q., Zhao, Y., Zhang, H., . . . Xia, L.-X. (2018). Hostile attribution bias mediates the relationship between structural variations in the left middle frontal gyrus and trait angry rumination. *Front Psychol, 9*, 526. doi:10.3389/fpsyg.2018.00526

Wegrzyn, M., Westphal, S., & Kissler, J. (2017). In your face: the biased judgement of fear-anger expressions in violent offenders. *BMC psychology*, *5*(1), 16. doi: 10.1186/s40359-017-0186-z

Werner, N. E. (2012). Do hostile attribution biases in children and parents predict relationally aggressive behavior? *The Journal of genetic psychology, 173*(3), 221-245. doi: 10.1080/00221325.2011.600357

Wiecki, T. V., Poland, J., & Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. *Clinical Psychological Science, 3*(3), 378-399. doi:DOI: 10.1177/2167702614565359

Wilkowski, B. M., Crowe, S. E., & Ferguson, E. L. (2015). Learning to keep your cool: Reducing aggression through the experimental modification of cognitive control. *Cognition and Emotion, 29*(2), 251-265. doi:10.1080/02699931.2014.911146

Wilkowski, B. M., & Robinson, M. D. (2012). When aggressive individuals see the world more accurately: The case of perceptual sensitivity to subtle facial expressions of anger. *Personality and Social Psychology Bulletin, 38*(4), 540-553. doi: 10.1177/0146167211430233

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*(3), 650-666. doi:10.1016/j.jrp.2006.07.003

Wong, M. S., Chen, X., & McElwain, N. L. (2019). Emotion understanding and maternal sensitivity as protective factors against hostile attribution bias in anger-prone children. *Social Development, 28*(1), 41-56. doi:10.1111/sode.12336

Yeager, D. S., Miu, A. S., Powers, J., & Dweck, C. S. (2013). Implicit theories of personality and attributions of hostile intent: A meta-analysis, an experiment, and a longitudinal intervention. *Child development, 84*(5), 1651-1667. doi:10.1111/cdev.12062

Yeung, R. S., & Leadbeater, B. J. (2007). Does hostile attributional bias for relational provocations mediate the short-term association between relational victimization and

aggression in preadolescence? *Journal of Youth and Adolescence*, *36*(8), 973-983. doi:10.1007/s10964-006-9162-2

Young, J. E., Klosko, J. S., & Weishaar, M. E. (2003). *Schema therapy: A practitioner's guide*. New York, YW: Guilford Press.

Zajac, L., Bookhout, M. K., Hubbard, J. A., Carlson, E. A., & Dozier, M. (2018). Attachment Disorganization in Infancy: A Developmental Precursor to Maladaptive Social Information Processing at Age 8. *Child development*. doi:10.1111/cdev.13140

Figure 1. Flowchart of the systematic review study selection process (adapted from Moher et al., 2009).

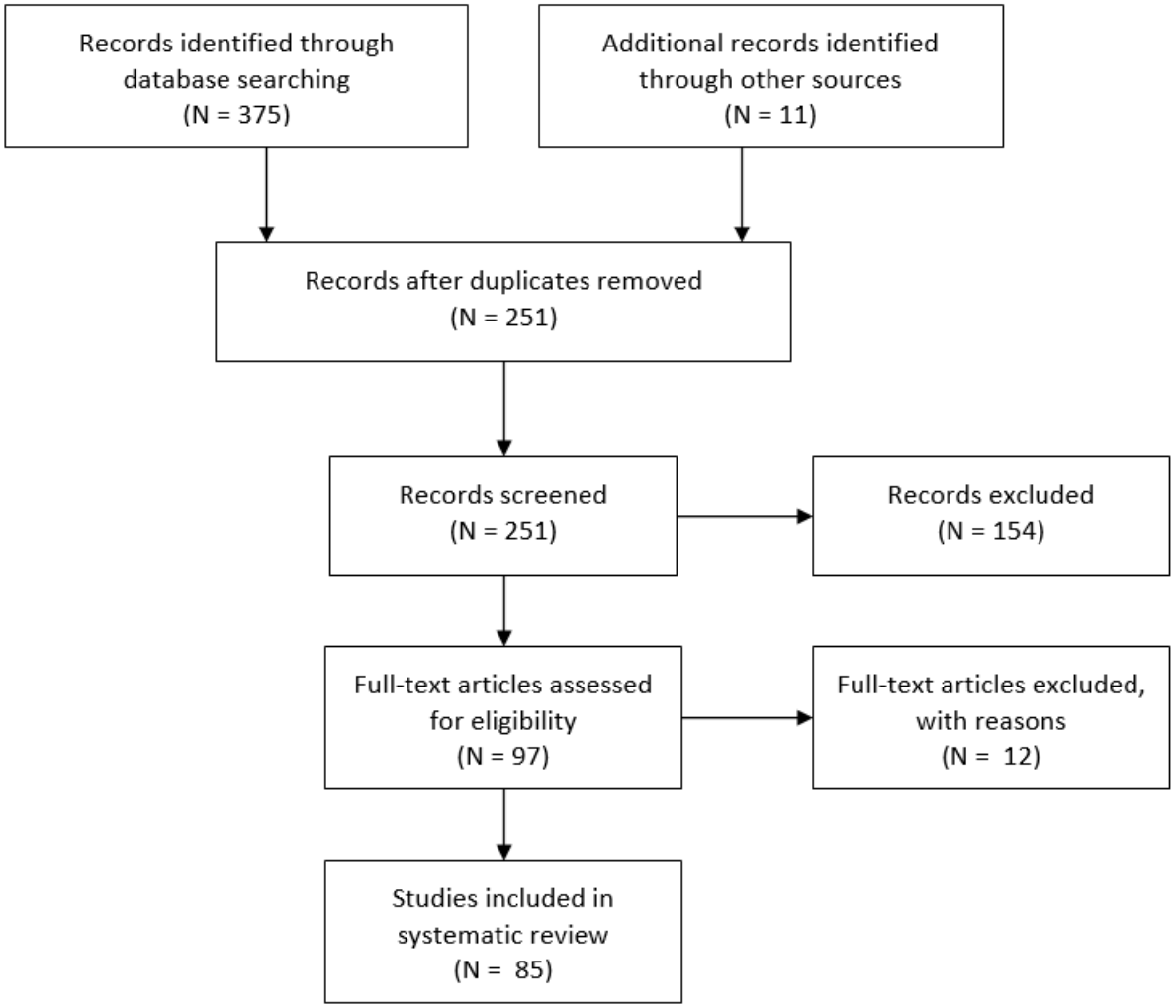

Figure 2. The Computations of Hostile Biases (CHB) model

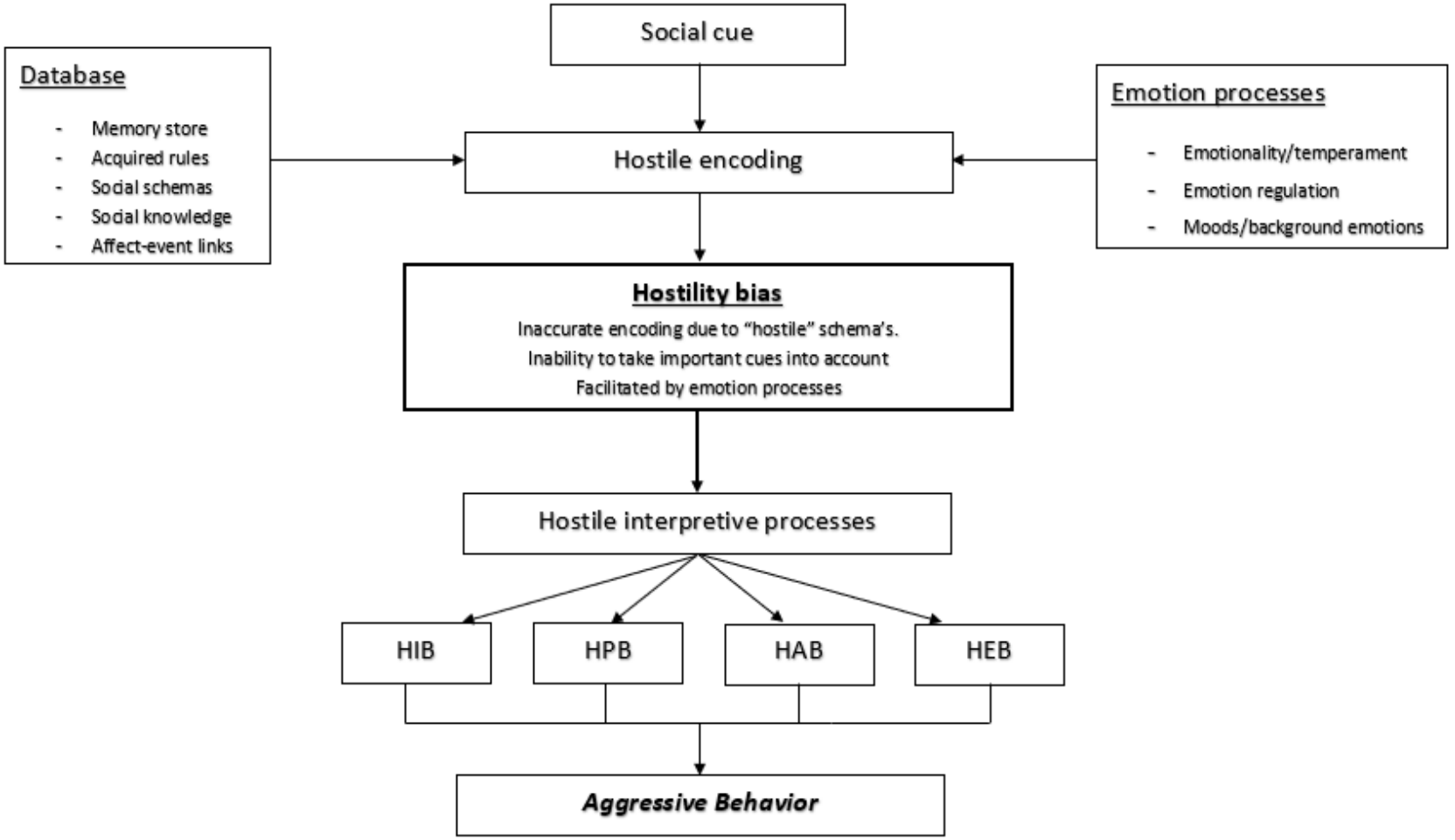

Figure 3.1. The Hierarchical Gaussian filter (HGF)

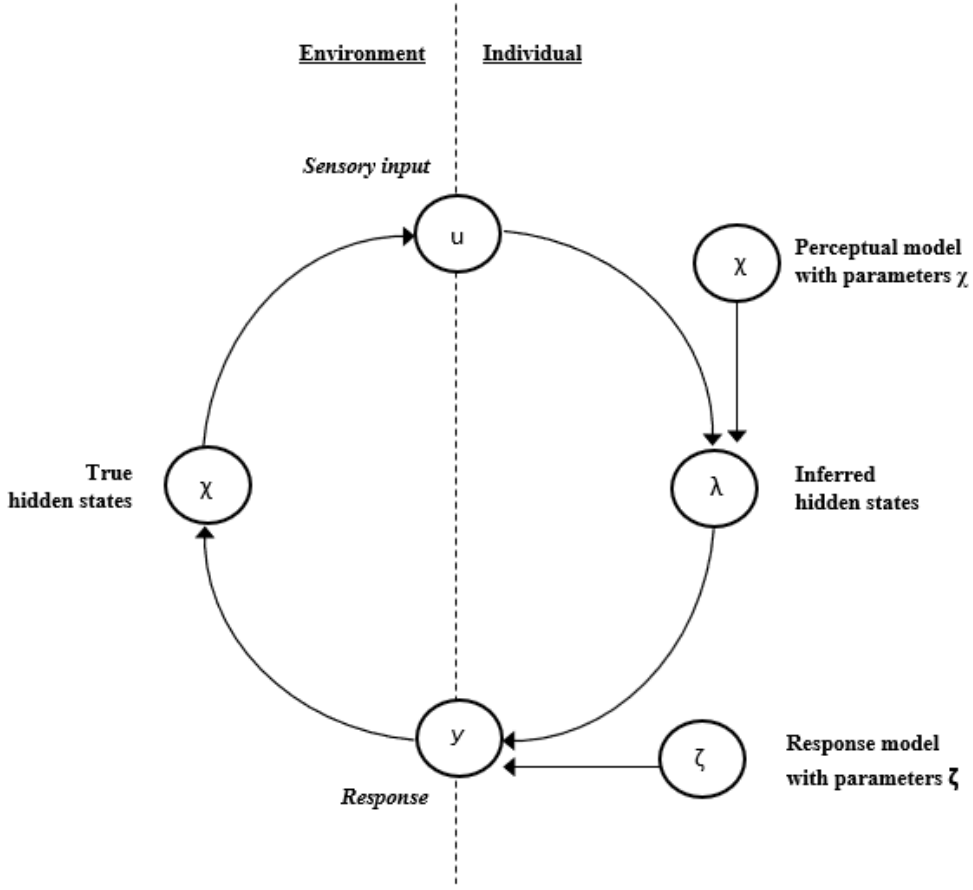

Figure 3.2. The CHB-model mapped onto the HGF

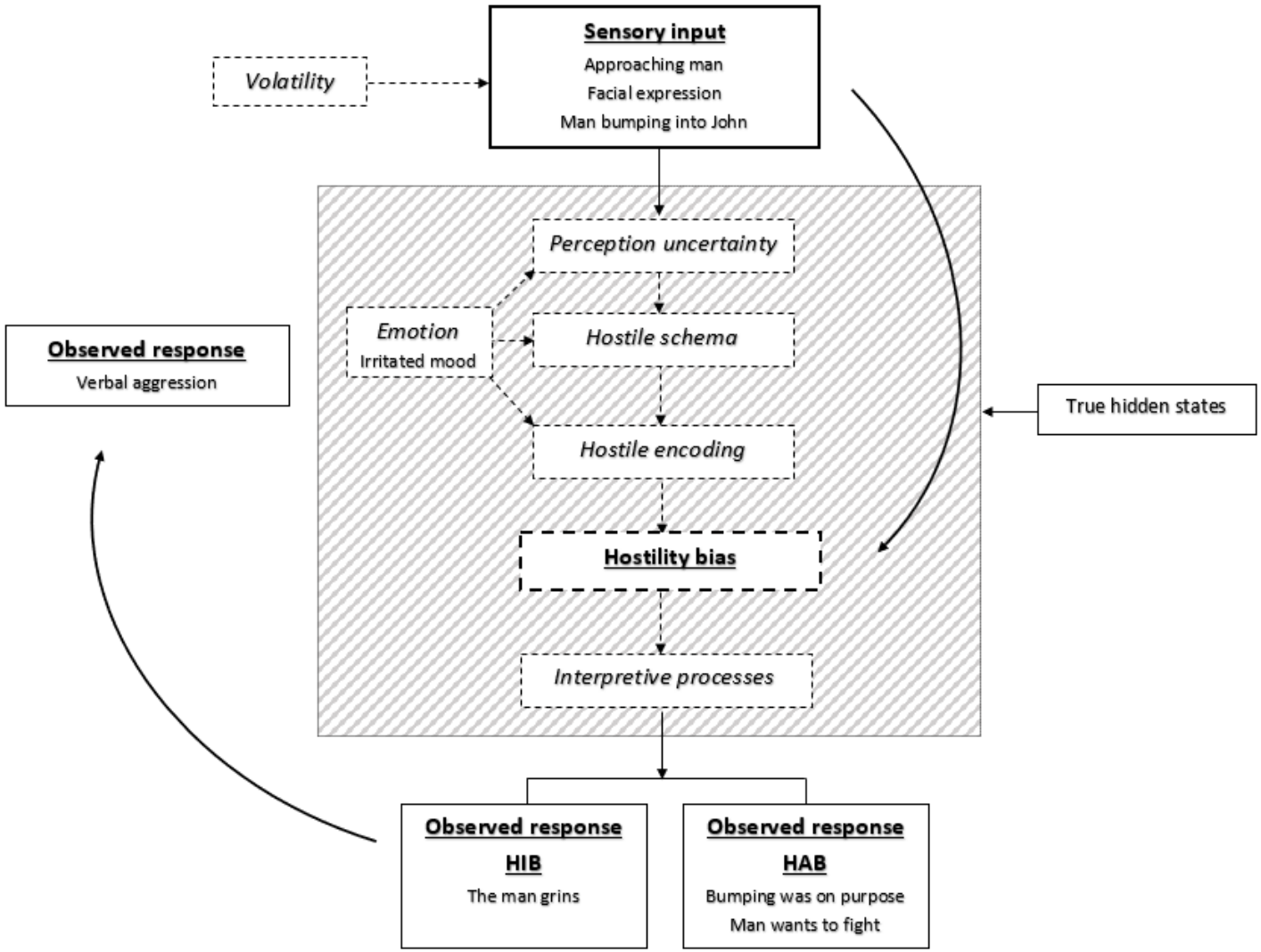

Table 1. *Overview included studies (Appendix)*

Author	Hostile bias	Measurement	Population	Country	N
Almoghrabi et al. (2018)	HAB	CBM	General population, undergraduate students	Netherlands	80
Bailey et al. (2007)	HAB	Vignettes	General population, undergraduate students	USA	165
Bartholow et al. (2006)	HPB	Donald paragraph	General population, undergraduate students	USA	246
Davies et al. (2018)	HAB	Vignettes	General population, children	USA	243
Begue et al. (2006)	HAB	Vignettes	General population, adolescents	France	379
Bondü et al. (2016)	HAB	Vignettes	General population, adults	Germany	349
Bondü et al. (2018)	HAB	Vignettes	General population, children + adolescents	Germany	279
Boulton (2012)	HAB	Vignettes	General population, children	UK	242
Bowen et al. (2016)	HAB	Vignettes	Incarcerated males	USA	330
Bratton et al. (2017)	HAB	Vignettes	Forensic psychiatric adult inpatients with schizophrenia	UK	27
Bushman (2016)	All	Review	-	-	-
Chen et al. (2012)	HAB	Vignettes	General population, adults	USA	2749
Choe et al.(2015)	HAB	Vignettes	General population, children (longitudinal)	USA	310
Choe et al. (2013)	HAB	Vignettes	General population, children	USA	231
Cillessen et al. (2014)	HAB	Vignettes	General population, children	Netherlands	366
Combs et al. (2009)	HAB	Vignettes	General population, adults + psychiatric patients with/without delusions	USA	110
Cougle et al. (2017)	HIB	CBM	General population, adults with alcohol use disorder + high trait anger	USA	58
Crick et al. (2002)	HAB	Vignettes	General population, children	USA	662
Darrell-Berry et al. (2017)	HAB	Vignettes	General population, adults + psychiatric patients with psychosis	UK	174
De la Osa et al. (2018)	HAB	Vignettes	General population, children	Spain	1341
DeWall et al. (2009)	HPB	Word pairs	General population, undergraduate students	USA	78
Dodge (2006)	HAB	Review	-	-	-
Dodge et al. (2015)	HAB	Vignettes	General population, children	World wide	1299
Edwards et al. (2012)	HAB	Vignettes	Mentally disordered offenders	England	62
Ellis et al. (2009)	HAB	Vignettes	General population, children	USA	83
Freeman et al. (2011)	HAB	Vignettes	General population, adolescents	UK	134
Gagnon et al. (2017)	HAB	Vignettes	General population, adults + undergraduate students	Canada	176
Gagnon et al. (2017)	HAB	Vignettes	General population, adults + aggressive psychiatric patients	Canada	87
Galan et al. (2017)	HAB	Vignettes	General population, children (longitudinal)	USA	310
Gentile et al. (2011)	HAB	Vignettes	General population, children	USA	430
Godleski et al. (2010)	HAB	Vignettes	General population, children	USA	840
Godleski et al. (2010)	HAB	Vignettes	General population, undergraduate students	USA	112
Haligan et al. (2007)	HAB	Vignettes	General population, children + parents	UK	134
Haligan et al. (2010)	HAB	Vignettes	General population, adolescents	UK	910
Hawkins et al. (2013)	HIB	CBM	General population, undergraduate students	USA	135
Helfritz-Sinville et al. (2014)	HAB	Vignettes	General population, undergraduate students	USA	867
Helseth et al. (2015)	HAB	Vignettes	General population, children with CU + CP traits	USA	60

Heppner et al. (2008)	HAB	Vignettes	General population, undergraduate students	USA	175
Hiemstra et al. (2018)	HIB	CBM	Children with behavioral problems	Netherlands	134
Horsley, et al. (2010)	HAB	Vignettes	General population, children	Netherlands	60
Jahoda et al. (2006)	HAB	Vignettes	Aggressive + non-aggressive individuals	Schotland	89
Jin et al. (2008)	HAB	Vignettes	Chinese immigrant batterers	USA	126
Jusyte et al. (2017)	HIB	Morphed face task	General population, adults + violent offenders	Germany	69
Karadenizova et al. (2018)	HAB	Vignettes	Violent adolescent offenders	Germany	27
Kay et al. (2016)	HAB	Vignettes	Adolescents in out-of-home care	UK	132
Kokkinos et al. (2017)	HAB	Vignettes	General population, adolescents	Greece	347
Kuin et al. 2017)	HIB	Morphed face task	General population, adults + violent non-violent offenders	Netherlands	117
Leff et al. (2014)	HAB	Vignettes	General population, adolescents + parents	USA	109
Li et al. (2016)	HAB + HIB	Questionnaire	General population, undergraduate students	China	162
Lin et al. (2016)	HIB	Eye-tracking	General population, adults	USA	36
Lobbestael et al. (2013)	HAB	Vignettes	General population, adults + psychiatric + forensic patients	Netherlands	66
MacBrayer et al. (2003)	HAB	Vignettes	General population, children + psychiatric patients	USA	100
Mammen et al. (2003)	HAB	Vignettes	Abusive parents	USA	52
Martins (2013)	HAB	Vignettes	General population, children	USA	150
Matheny et al. (2017)	HIB	WSAP-H	Adults seeking anger treatment	USA	131
Mathieson et al. (2011)	HAB	Vignettes	General population, children	USA	635
Matthews et al. (2002)	HAB	Vignettes	General population, adults	USA	163
McDermott et al. (2017)	HIB	WSAP-H	Adults seeking anger treatment	USA	131
Mellentin et al. (2015)	HIB	Review	-	-	-
Miller et al. (2019)	HAB	Vignettes	General population, children	Canada	211
Möller et al. (2009)	HAB	Vignettes	General population, adolescents	Germany	295
Nelson et al. (2009)	HAB	Vignettes	General population, children + parents	USA	219
Nentjes et al. (2015)	HIB	RMET	General population, adults + psychopathic + non-psychopathic offenders	Netherlands	102
Neumann et al. (2017)	HAB	Vignettes	General population, adults + individuals with brain injury	USA	95
Orobio de Castro (2002)	HAB	Review	-	-	-
Orobio de Castro (2003)	HAB	Vignettes	Highly, moderate and non-aggressive children	Netherlands	57
Orobio de Castro (2005)	HAB	Vignettes	General population, children + referred aggressive children	Netherlands	84
Peets et al. (2007)	HAB	Vignettes	General population, children	Estonia	144
Quan et al. (2019)	HIB	WSAP-H	General population, undergraduate students	China	942
Smeijers et al. (2017)	HIB	Morphed face task	General population, adults + psychiatric + forensic psychiatric patients	Netherlands	212
Smith et al. (2016)	HIB	WSAP-H	General population, undergraduate students + individuals seeking anger treatment	USA	203
Teige-Mocigemba et al. (2016)	HIB	Morphed face task	General population, adults + undergraduate students	Germany	65
Tuente et al. (2019)	HAB	Review	-	-	-
Thomas et al. (2019)	HAB	Vignettes	General population, undergraduate students	USA	241
van Dijk et al. (2018)	HAB	Vignettes	General population, children	Netherlands	104
Vassilopoulos et al. (2015)	HAB	CBM	General population, children	Greece	34
Wegrzyn et al. (2017)	HIB	Morphed face task	General population, adults + violent and sexual offenders	Germany	62
Werner (2013)	HAB	Vignettes	General population, children + parents	USA	91

Wilkowski et al. (2015)	HIB	CBM	General population, undergraduate students	USA	108
Wilkowski et al. (2012)	HIB	Morphed face task	General population, undergraduate students	USA	213
Wilkowski et al. (2007)	HIB	Eye-tracking	General population, undergraduate students	USA	45
Wong et al. (2019)	HAB	Vignettes	General population, children	USA	128
Yeager et al. (2013)	HAB	Vignettes	General population, adolescents	USA	1758
Yeung et al. (2007)	HAB	Vignettes	General population, children	Canada	142
Zajac et al. (2018)	HAB	Vignettes	General population, children	USA	77

Note: Vignettes refer to videos, cartoons or written stories, CBM refers to cognitive bias modification, CU refers to callous unemotional, CP refers to conduct problems, WSAP-H refers to Word Sentence Association Paradigm – Hostility, RMET refers to Reading the Mind in the Eyes Task