STUDIA UBB PSYCHOL.-PAED., LXXI, 1, 2026, p. 5-50
(Recommended Citation)
DOI:10.24193/subbpsyped.2026.1.01

Reading Threat Everywhere: A Scoping Review of Hostility Biases and Dark Personality Traits in Offenders

Laura OPRIŞ[1,2], Laura VISU-PETRA[1,3*]

ABSTRACT. Despite significant interest and relevance to society, the triggers that lead to antisocial behavior are still insufficiently investigated, especially in key populations such as convicted offenders.

This study aimed to summarize the current literature on *hostility biases* and aggressive behavior in incarcerated individuals, in relation to their dark personality traits. A thorough search across multiple electronic citation databases (Google Scholar, PsycInfo, Scopus, Science Direct, and Web of Science) was performed, focusing on articles published from January 1990 to August 2025. Out of the 344 articles initially identified, only 14 were deemed suitable for inclusion in the scoping review following the eligibility screening process. The studies that were included measured hostility biases (HBs) occurring at different stages of social information processing (perception, interpretation, attribution, expectation). The research had to be conducted on offenders and include a measure of aversive personality traits alongside other relevant variables.

A diverse population of convicted individuals displaying aggressive propensities (proactive/reactive) or mental health problems, as well as malevolent dimensions in their personality structure, was analyzed for extracting potential triggers of law breaking.

Our review revealed that delinquent behavior emerges as a result of a multilevel processing of social cues, from facial recognition of emotions to interpretive and intentional attributions, differentially flavored by antagonistic personality traits. Analyzing research carried out in correctional facilities

[1] Research in Individual Differences and Legal Psychology (RIDDLE) Lab, Department of Psychology, Babeş-Bolyai University, Republicii, 37, 400015, Cluj-Napoca, Romania
[2] Oradea Penitentiary, Parcul Traian, 3, 410033, Oradea, Romania
[3] Department of Social and Human Research, Romanian Academy, 400015 Cluj-Napoca, Romania
* Corresponding author: laurapetra@psychology.ro

offers valuable insights into the core of the information decoding mechanism that precipitates offenders' propensity to engage and persist in committing antisocial behaviors.

Keywords: social information processing, hostility bias, aversive personality, psychopathy, aggression, incarcerated offenders.

INTRODUCTION

Effective social interaction and engagement rely on the accurate processing of social information. Individuals who exhibit disruptive behaviors are believed to perceive, interpret, and react to social cues in ways that elevate the likelihood of acting aggressively (Dodge & Crick, 1990). Considerable research has so far examined the intricate links between aggressive behavior and challenges in understanding and processing social information. Aggression can be amplified when an individual perceives a threat, anger, or malevolent intentions even in neutral contexts. This phenomenon is conceptualized as a ***hostility bias*** (Orobio de Castro et al., 2002; Dodge, 2006; Smeijers et al., 2019). The Social Information Processing (SIP) framework (Dodge & Crick, 1994) suggests that distortions in social information processing can appear at various stages, including cue encoding, interpretation, and response evaluation. From this theoretical standpoint, individuals who exhibit aggression are more likely to ascribe hostile intentions to others, compared to those who are nonaggressive (Dodge, 1980; Huesmann, 1998).

Empirical studies consistently demonstrated a positive correlation between *hostility biases* (HBs) and *aggressive behaviors* in children and adolescents (Crick & Dodge, 1996; Crick et al., 2002; Orobio de Castro et al., 2002), as well as in the adult population (Epps & Kendall, 1995; Matthews & Norris, 2002; MacBrayer et al., 2003; Basquill et al., 2004; Miller & Lynam, 2006; Bailey & Ostrov, 2008; Zajenkowska et al., 2021; Ursulet et al., 2022). Additionally, comprehensive reviews and meta-analytic research indicate that HBs are consistently linked to reactive aggression and disruptive behaviors throughout childhood, adolescence (Crick & Dodge, 1994; Verhoef et al., 2019), and adulthood (Smeijers et al., 2019; Tuente et al., 2019). HBs significantly impact the interpretation of unclear social situations and serve as cognitive triggers for reactive or impulsive behavior, emerging as the most robust cognitive correlates of aggression and antisocial conduct (Orobio de Castro et al., 2002; Dodge, 2006).

Beyond cognitive factors, ***personality traits*** also significantly influence the development and manifestation of delinquent behavior. According to the General Aggression Model (Anderson & Bushman, 2002; Allen et al., 2018), the combination of individual and situational elements activates a person's internal state, including cognition, emotion, and physiological arousal, as well as their interaction. This subsequently leads to specific behavioral responses (Anderson & Bushman, 2002). Consequently, a significant amount of research has been carried out to determine which personality characteristics are linked to the manifestation of aggression. So far, research based on the Five-Factor model of personality (Big Five, Costa & McCrae, 1992) has consistently associated low levels of agreeableness and conscientiousness with externalizing behaviors and antisocial outcomes (Malouff et al., 2005; Jones et al., 2011). Additionally, agreeableness, conscientiousness, extraversion, and openness negatively correlated with the hostile attribution bias (HAB, see Table 1) and with aggressive behavior levels, whereas neuroticism showed a positive correlation with the latter outcomes (Kokkinos et al., 2017). A connection was also found between relational aggression and a specific combination of Big Five traits, including high neuroticism and low levels of agreeableness, conscientiousness, extraversion, and openness, which was mediated by the HAB (Kokkinos et al., 2017). The authors proposed that individuals with this specific blend of traits may be more inclined to engage in hostile attributions, which is associated with a greater likelihood of exhibiting aggression. Beyond typical personality traits, a variety of features known as "dark" personality traits: callous-unemotional, psychopathy, narcissism, Machiavellianism, and sadism emerge as correlates for aggressive responses, antisocial behavior, and even offending (Law & Falchenbach, 2017; Subra, 2023; Jiang et al., 2024; Kersten et al., 2024; Potter, 2024). Apparently, HBs are functioning as cognitive pathways linking antagonistic dispositions and interpersonal aggression, in the general adult population (Jiang et al., 2024), as well as in delinquents (Slaby & Guerra, 1988; Potter, 2024).

Hostility biases: conceptual and methodological clarification

According to SIP (Crick & Dodge, 1994), HBs manifest when negative cognitive schemas and past experiences related to others and situations are triggered by current adverse events that evoke conscious or unconscious links to previous incidents (Guerra & Huesmann, 2004). The model integrates the social, cognitive, behavioral, and emotional factors that contribute to the emergence of aggression. Accordingly, an individual's response to a perceived provocation or threat is impacted not only by objective social cues but is also significantly shaped by the manner in which that social information is unraveled.

Based on the SIP framework, the mechanism of social information processing requires the following six phases: (1) encoding cues; (2) interpreting signals; (3) clarifying a goal; (4) generating a response; (5) selecting a response and evaluating its effects; and ultimately (6) exhibiting behavior (Crick & Dodge, 1994). The first two stages represent early social information processing and are thought to lead to reactive aggressive behavior due to the misinterpretation of unclear and ambiguous situations, which trigger feelings and thoughts of threat. The final four stages are categorized as late information processing and are mostly defined by cognitive processing. Analyzing this multiphasic mechanism reveals that various deviations can occur along the complex path of decoding social cues.

However, aside from the hostility bias in intent attribution – the most investigated construct – additional hostility-related biases may emerge during the different phases of processing social information. Thereby, three other hostility biases have been emphasized: hostile interpretation bias, perception bias, and expectation bias, respectively (see Table 1).

Table 1. *Hostility biases (perception/ attribution/ interpretation/ expectation) – Conceptual clarifications*

Concept	Definition and conceptual clarifications
Social information processing (SIP)	= decoding and interpreting social cues in others and in situations
Hostility bias (HB)	= *the `a priori tendency(es) to perceive and interpret social information in a hostile manner`* (Smeijers, 2022)
Attributional bias / style	= pattern in which individuals explain the causes, or the significance for social interactions and contexts
Hostile perception bias (HPB)	= identifying and assigning a hostile meaning to recognition of emotions based on facial/non-facial indicators or bodily movements
Hostile interpretation bias (HIB)	= deciphering social cues as negative (e.g., misinterpretation of words, facial expressions, or tone)
Hostile attribution bias (HAB)	= the tendency to ascribe a provocative intent to others' actions in ambiguous situations
Hostile expectation bias (HEB)	= anticipating a harmful outcome from what appears to be neutral or unclear stimuli.

The ***hostile perception bias (HPB)*** pertains to the inclination to perceive ambiguous social exchanges as hostile (Bushman, 2016). For example, when observing two people speaking loudly to each other, a hostile perception could imply that they are engaged in an argument or preparing to fight. The ***hostile interpretation bias (HIB)*** is defined as the preexisting tendency to view social stimuli as hostile. For instance, if someone is gazing at you, a hostile interpretation might suggest that their facial expression indicates anger, when it is actually neutral. The distinction between the HIB and the HPB is that interpretation bias focuses solely on understanding social stimuli, while the perception bias has a wider scope, encompassing entire social interactions interpreted based on facial/bodily movement cues. The ***hostile attribution bias (HAB)*** is defined as a cognitive tendency to process social cues by assigning a threatening intent to neutral, unclear, or vaguely expressed actions. Therefore, HAB requires a higher level of reasoning about intentions behind actions, while HPB and HIB occur at a lower level of decoding and attaching a malevolent meaning to emotional expressions. Lastly, ***hostile expectation bias (HEB)*** refers to the tendency to anticipate that someone will respond to potential conflicts with hostility (Bushman, 2016). For instance, if you bump into someone accidentally in a crowded place, a hostile expectation might lead you to believe that the person will think you did it intentionally and, due to this, react aggressively. It is essential to underline that all these concepts referring to biases are occasionally used interchangeably in the literature, which makes it difficult to clearly delineate the links, similarities, and specific cognitive mechanisms.

As a general *mechanism* responsible for the relation between HBs and aggression, it has been posited that a person is more inclined to behave aggressively when they perceive, anticipate, or interpret hostility in others. Consequently, this elicits a more aggressive response from those nearby. Misinterpreting situations as hostile significantly contributes to the development and persistence of aggressive behavior. This tendency is associated with the overarching view that hostility is a fundamental aspect of social interactions (Guerra & Huesmann, 2004). Thus, both HBs and aggression will be perpetuated through this vicious cycle. However, it remains unclear whether hostility biases are mainly a cause or a consequence of aggressive behavior. Due to the robust and reciprocal relationship between HBs and aggressive behavior, these distortions are important targets for understanding and treating aggression and delinquency.

Hostility biases and (aversive) personality traits

The persistent decoding of social information in a hostile rather than positive attributional style can evolve into a consistent personality trait (Dodge,

2006). Conversely, exploring the role of certain dimensions of personality in precipitating antisocial outcomes associated with HBs is also of significant scientific and applied interest.

A substantial body of research on the relationship between personality traits and antisocial behavior has found that individuals who engage in criminal activities often exhibit self-centeredness, hostility, distorted beliefs, and deficits in impulse control (Bate et al., 2014; Blair & Mitchell, 2009; Blonigen et al., 2012; Hare, 1996; Stalenheim, 2004; Tharshini et al., 2021). Therefore, considering callous-unemotional (CU) traits, psychopathy, and antisocial personality was a natural step in delving into the mechanisms of HBs.

CU traits shape the manifestation of unique socio-cognitive characteristics. Research on children aged between 8 and 17 years with behavioral problems revealed that these characteristics, which involve a lack of empathy, guilt, and remorse, are correlated with an increased propensity to decode unclear social indicators as hostile (Hartmann et al., 2020). A subsequent study on middle-school children aged between 11 and 14 years found that the accuracy of emotion recognition for anger was low in those with higher CU scores, regardless of the level of conduct problems. Conversely, children who scored low on CU traits exhibited less accuracy in recognizing emotions as the severity of conduct problems increased (Ciucci et al., 2024). Consequently, different paths to aggression have been identified by the fact that while some individuals with elevated scores on CU traits demonstrate diminished emotional responsiveness, others still manifest exacerbated HBs when provoked (Ciucci et al., 2024).

Antisocial personality and *psychopathic traits* have been associated with deficits in the recognition of facial affect (Marsh & Blair, 2008; Dawel et al., 2012). However, impairments in recognition may arise from an a priori proclivity to interpret facial expressions in a distorted stance, which is aligned to the concept of HB (Smeijers et al., 2017). Previously, researchers proposed that HBs are more prevalent among individuals with psychopathic traits, since their lack of emotional responsiveness might impede their ability to interpret subtle social signals and differentiate between ambiguous and explicit hostile scenarios (Maccoon & Newman, 2006). Lastly, it was found that secondary psychopathy is associated with attributing hostile intent and reactive aggression (Law & Falchenbach, 2017).

Across adults from community samples, new findings reveal that HAB mediates the link between *dark traits* (especially psychopathy, sadism, and Machiavellianism), and relational and reactive aggression (Jiang et al., 2024; Kersten et al., 2024). Positive correlations have been revealed between all aspects of the Dark Tetrad (narcissism, Machiavellianism, psychopathy, and everyday sadism) and aggressive behaviors (Barlett, 2016; Paulhus et al., 2018).

However, recent research highlights that HBs are not only associated with clinical levels of antagonistic personality dimensions, mainly represented by psychopathy, but the explicative frame needs to be extended to other `dark` or `aversive` personality traits (Moshagen et al., 2025). Due to the findings that reveal that these subclinical personality dimensions predict engagement in criminal behavior (Bonfá-Araujo et al., 2022; Hurezan et al., 2024), exploring the role of HBs in predicting law violations and persistent offending is required. For instance, *narcissism* (both vulnerable and grandiose) is a strong predictor of HAB (Subra, 2023). Prior studies have shown that individuals with high levels of narcissism tend to perceive more hostility in social situations and react aggressively when they perceive threats to their ego (Baumeister et al., 1996). Therefore, HAB is seen as a crucial factor interfering with the correlation between narcissism and reactive aggression (Bushman & Baumeister, 1998). Also, a temporary increase in state narcissism made individuals more angry and aggressive after unexpected provocation in a sample of college students (Li et al., 2016). These results are in line with prior findings, which highlight that activating narcissism can lead to negative emotions (e.g., anger, hostility) and aggression (Back et al., 2013). A recent study (Eriksson & Schmidt, 2026) identified a subgroup of adolescents with moderate symptoms of social anxiety, for whom a combination of vulnerable and grandiose narcissism traits specifically predisposed them to aggressive behaviors and impulsivity, rather than to shyness and withdrawal.

Although HBs were frequently identified as robust correlates of reactive aggression in general adult samples, the findings in offender and delinquent populations are notably less consistent, mostly focused on evoking the manifestation of the HAB, with inconsistent attention to other HBs. The present scoping review addresses how antagonistic personality traits interact with attributional processes occurring at different levels of information decoding, to trigger and maintain delinquent behaviors, particularly among adults in prison.

Hostility biases and offending: links and explanations

Research on criminal behavior focused on the role of attribution distortions, attempting to decipher how individuals interpret the causes or intentions behind others` actions. Among these biases, especially HAB has emerged as one of the most consistent cognitive correlates of behaviors that defy social norms (Orobio de Castro et al., 2002; Dodge, 2006). However, the research examining the impact of social cognition on violent offending remains limited (Hutchings et al., 2010). Meta-analytic evidence highlights that HAB is consistently associated with reactive aggression and conduct problems across childhood and adolescence (Crick & Dodge, 1994; Zajenkowska et al., 2020). More recent longitudinal research

on juvenile offenders confirms that HAB often mediates the relationship between early environmental risk factors (e.g., harsh parenting, peer rejection) and subsequent delinquent behavior (Lin et al., 2023).

Despite the correlation already posited between aggressive behavior and HAB, the question arises as to whether all HBs are a consequence of the extreme aggressive behavior or whether they mainly lead to difficulties in managing aggressive tendencies. This is especially concerning because a deeper insight into the social, cognitive, and emotional processes contributing to aggression is crucial for enhancing assessment methods and interventions (Coccaro et al., 2017).

Besides HAB, in a sample of psychopathic and non-psychopathic offenders, HIB was found not to significantly correlate with psychopathy, although it was related to an inclination towards proactive rather than reactive aggression. It was also linked to a heightened ability to recognize angry-looking eyes and a greater accuracy in attributing mental states compared to non-psychopathic offenders (Nentjes et al., 2015). This latter observation indicates that individuals exhibiting high levels of psychopathy may not demonstrate a bias; rather, they seem to possess an enhanced awareness of hostile mental states.

The majority of existing studies carried on with adults have indicated a small to medium positive correlation between aggression and HAB (see review by Tuente et al., 2019). One seminal research in highlighting the specific manifestation of HBs among adults in general and offenders in particular, analyzed 25 studies assuming that the link between HAB and aggression is more pronounced in groups that exhibit elevated levels of aggression (e.g., violent offenders and forensic patients). Several studies have found a positive correlation between HAB and aggression (Hornsveld et al., 2007; Lim et al., 2011). Specifically, Hornsveld et al. (2007) indicated a significant relationship between the Buss-Perry Aggression Questionnaire and the Physical Fear Scale - Aggression Violence in a sample of inpatients and outpatients in forensic psychiatric settings. Furthermore, both physical and verbal aggression showed significant positive correlations with assessed HAB.

Lim et al. (2011) reported that an overall sample of both violent and non-violent inmates exhibited a significant positive correlation between aggression expectancy and HAB. Lastly, individuals who commit violent offenses perceive others' actions as more deliberately hostile than those who commit nonviolent offenses (Lim et al., 2011).

Edwards & Bond (2012) revealed that 62 offenders with mental disorders produced considerably more aggressive statements after reading narratives depicting physical aggression. The same study found that high narcissism and low self-concept clarity predicted the HAB in mentally disordered offenders.

Furthermore, forensic outpatients with antisocial and borderline personality disorders showed not only more elevated levels of reactive and proactive aggression compared to healthy controls and non-forensic patients with the same personality disorders, but they exhibited more pathological forms of aggression (Smeijers et al., 2017).

While attributing hostile intent to the actions of others does, in fact, play a role in fostering aggressive behavior in adults, the initial hypothesis presumed by Tuente and collaborators (2019) was not substantially supported by empirical findings. Of the seven investigations involving forensic or offender samples, three showed an absence or mixed relationships between HAB and aggression. In particular, Bowen et al. (2016) found that HAB did not serve as a predictor of aggressive responses in male offenders and, in some instances, was inversely related to these responses; Lobbestael et al. (2013) found that HAB was only associated with reactive, not proactive, aggression. Coccaro et al. (2017) discovered unexpected inverse correlations in their SIP-based model, showing that low inhibitory control and increased emotional reactivity may coexist with intact or even enhanced socio-cognitive abilities.

These results imply that, unlike community samples, HBs do not consistently serve as a reliable indicator of aggression among delinquent adults. Some possible explanations for these contrasting results may be attributed to the increased prevalence of instrumental aggression, to the overlapping characteristics of HBs, or to certain personality characteristics among offenders, and to the variability in methods used to measure HBs within these groups (Tuente et al., 2019).

Heterogeneous findings from the aforementioned studies led to the conclusion that further investigation into the relationship between HBs and aggression in adults should utilize multiple methods (such as interviews, surveys, and scenarios) and multi-informant sources (combining self-reports and behavioral observations). Moreover, HBs are a constellation of biases that operate at several stages of social information processing rather than a single process therefore, it is crucial to address them in an integrative manner. This approach would lead to an unifying model in which all HBs are manifestations of one general hostility bias mechanism (Smeijers, 2019).

PURPOSE OF THE REVIEW

The causal factors and mechanisms of aggression are intricate and multidimensional. Aggressive behavior generally occurs when a mix of individual and environmental triggers leads to increased distorted thoughts, „hot" emotions, and physical arousal. An abundance of studies consider that HBs aid in both the

formation and the ongoing persistence of aggressive behavior (Crick & Dodge, 1996). Besides that, integrative models propose that personality traits and attribution processes interact in a dynamic interplay. In addition, extensive studies have been carried out to identify personality characteristics associated with aggressive behavior. Findings have indicated positive correlations between all dimensions of the Dark Tetrad (narcissism, Machiavellianism, psychopathy, and everyday sadism) and aggressive tendencies maintained by HBs (Barlett, 2016; Paulhus et al., 2018).

To the best of our knowledge, no review has been published regarding the HBs of convicted offenders with elevated levels of aversive personality traits. Initially, we focused on existing studies conducted with inmates to map the current state of research on HBs and aggressiveness. However, the aforementioned association`s interference with aversive personality dimensions led us to extend our analysis of information processing distortions to include offenders characterized by dark personality traits. We considered studies that measured antagonistic personality traits to refine the investigation of potential explanatory models of crime occurrence.

This scoping review summarizes the types of HBs investigated and the instruments used to measure them. It provides an overview of the main findings and examines the aversive personality traits of incarcerated offenders (mainly psychopathy and narcissism), as well as other involved variables (e.g., proactive/ reactive aggression, cognitive abilities, thinking styles, social desirability, etc.).

METHOD

In conducting this scoping review, we followed the six-step framework provided by Arksey and O'Malley (2005) and the general guidelines outlined by PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, Liberati et al., 2009).

Search strategy and included studies

A complex search was performed in September 2025 for articles in English, published between January 1990 and August 2025, and focusing on the link between hostility biases and aggressive behavior in multiple electronic citation databases: Google Scholar, PsycInfo, Scopus, Science Direct, and Web of Science. The following keywords were used with all databases: 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 prison*; hostile (attribution, interpretation, perception, expectation) AND bias AND inmate; hostile (attribution, interpretation, perception, expectation) AND bias AND personality AND delinquen*; hostile (attribution, interpretation, perception, expectation) AND bias AND personality AND psychopathy.

This initial search resulted in 344 references, of which 25 were duplicates. The remaining 319 articles were evaluated for eligibility, leading to a final selection of ***14 relevant*** references (see Figure 1 for the flowchart depicting the study selection process). As a first step, the titles and abstracts were separately screened by the first author to establish if they reported studies regarding HBs and were (1) measuring HBs through self-report questionnaires or experimental tasks, (2) conducted on a forensic population, (3) exploring dimensions of aversive personality (e.g., psychopathy, narcissism). All articles meeting the mentioned criteria were included in the scoping review. Further, we obtained the full-text version for each article selected after the title and abstract screening. For this step, we established a set of inclusion and exclusion criteria.

Relevant articles were selected using the following inclusion criteria:

(a) Research should be written in English;

(b) Studies should use for measuring self-report questionnaires, experimental tasks, regarding HBs (attribution, interpretation, perception, expectation) and dark personality traits;

(c) Studies should be conducted on convicted offenders (currently serving a prison sentence inside a prison/ correctional facility) or forensic inpatients in a secure clinical setting;

(d) All research should focus on adult samples, including participants with psychiatric disorders;

(e) Participants of all genders (men, women, and non-binary individuals) were considered eligible";

(f) The date of publication should be between January 1990 and August 2025;

(g) Grey literature (unpublished theses and dissertations).

The screening process of the relevant articles had certain exclusion criteria:

(a) Articles focused only on emotion recognition or on different variables related to hostility attribution bias (e.g., personal space intrusion);

(b) Research with a different age category than adults (e.g., juvenile delinquents);

(c) Samples composed by offenders that were not serving a sentence in a correctional/ psychiatric facility (e.g., under surveillance of the probation service);

(d) Articles describing research focused solely on outcomes of programs aimed at improving attribution bias or reducing aggression rate (e.g., pilot studies, pre/post-measures of HBs);

(e) Articles that were focused solely on the instruments used to measure HBs and their psychometric properties;

(f) Qualitative studies (e.g. thematic analysis), systematic reviews, scoping reviews, books, and book chapters,

(g) Articles written in a language other than English.

Articles categorization. Following the articles' screening process, three main categories emerged, based on the type of HBs that were investigated: a) hostile perception bias (HPB), b) hostile interpretation bias (HIB), c) hostile attribution bias (HAB), and d) multiple HBs. Each article was included in only one category that best characterized the purpose for which the research instruments were used to investigate the HBs. For example, although some studies seemed to measure the HPB due to the morphing task or other perceptual-focused instruments, the inclusion of the article was made accordingly to the purpose pursued. The present scoping review focuses on all of the categories mentioned above and reports the outcomes regarding individual differences in offenders' processing information bias.

Data summary and analysis

For data summarization, a structured table was developed, containing the following information: author(s), year of publication, type of HBs investigated, measure(s) for HBs (vignettes or experimental tasks), personality instruments and other variables, study population (type – prison group and/or community sample, sample size, average age) and main findings. Further, we collated all extracted details and produced a table, mapping the basic characteristics of the selected studies: HBs assessed in the prison setting, with convicted adult offenders, the type of personality instrument used, as well as other related variables, and the main findings. Lastly, we discussed the outcomes according to the category of the investigated HBs, underlying the particularities related to the prison setting.

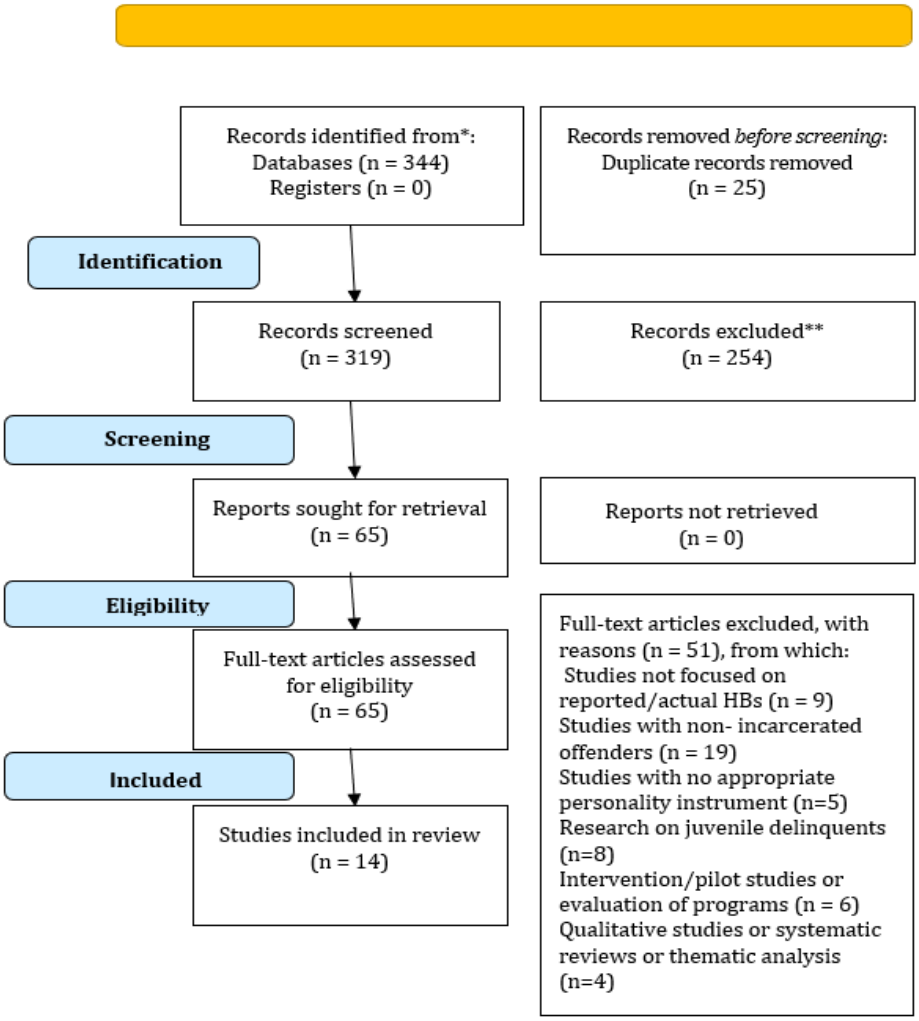
RESULTS

The 14 selected articles were divided into four categories, according to the HB measured: exclusively HPB (n = 3 articles), exclusively HIB (n = 2 articles), exclusively HAB (n = 5 articles, out of which 2 were a doctoral thesis

and a dissertation), and mixed HPB/HIB/HAB (n = 4 articles). No articles on measuring HEB were found to meet the established criteria.

Besides measuring various HBs assessment in offenders, the research had to include a measure of aversive personality traits placed in the relation between HAB and aggressive behavior. *Table 2* presents an overview of the study design, population, and main findings for each article.

Figure 1. *The PRISMA Flow Diagram*



*Databases: Google Scholar (first 344 hits), Scopus, Science Direct, Web of Science, and PsycINFO
**The records were excluded by the authors

Table 2. *Summary of studies measuring HBs in offenders*

Author(s), year of publication / Participants (N/n= number, M=mean age)	Type of hostile bias and task(s) used	Assessment of aversive personality (I) and other variables (II)	Main findings
Faith et al., 2022 / Offenders: N=280 (M = 26.66)	**HPB** Monomorph Facial Affect Task	I. PCL-R; II. SILS-2; WRAT-4	- psychopathy was negatively correlated with d` (sensitivity) for anger, according to Signal Detection Theory (SDT); - noticeable impact of psychopathy on the response criterion (c) for fear and on the response criterion (c) for anger were observed; - HPB is not only a perceptual impairment, but also decision-based mechanism.
Gillespie et al., 2015 / Sexual offenders: N=13 (M = 50.5), Violent offenders: N=16 (M = 37.8) Non-offending controls: N=19 (M = 48.2)	**HPB** NimStim for expression recognition	I. LSRP; II. MC-C;	- sexual offenders showed HPB (reduced sensitivity to detect emotional expressions compared with non-offenders); - sexual and violent offenders had a reduced sensitivity to perceive fearful expressions; - compared with controls, sexual and violent offenders exhibited impaired sensitivity to high-intensity female fearful expressions.
Philipp- Wiegmann et al., 2017 / Violent offenders: N=41 assigned to the group of proactive (n=24; M = 34.04) or reactive violent offenders (n=17; M = 34.94); Control group: N = 29 students; (M = 31.55)	**HPB** FEEL	I. PCL-R II. MWT-B; PROREA	- emotion recognition in reactive offenders was significantly lower compared to proactive violent offenders and controls (in general and particularly in recognition of negative emotions such as anxiety, sadness, and disgust); - reactive violent offenders were more likely to show HPB by misreading non-anger emotions as anger; - proactive violent offenders performed similar to controls.
Kuin et al., 2017 / Offenders: N = 85, of which violent offenders (n = 71; M = 36.56); non-violent offenders (n = 14; M = 37.43); Control group (n = 32; employees of the prison service; M = 41.75)	**HIB** Emotion perception task	I. PPI-R; II. RPQ,; BPAQ; IPAS-30; SDAS-11; RSPM; TMT; The beads-in-a-jar task	- no significant data to support that violent offenders perceived facial emotional expressions as more angry than non-violent offenders or healthy controls; - age and a self-report measure of hostility successfully predicted results on the emotion perception task (older offenders made fewer hostile ratings assigned to the facial expressions; -psychopathic traits, IQ, attention and a tendency to jump to conclusions were not correlated with interpretation of anger in facial emotional expressions.

Author(s), year of publication / Participants (N/n= number, M=mean age)	Type of hostile bias and task(s) used	Assessment of aversive personality (I) and other variables (II)	Main findings
Schönenberg & Jusyte, 2014 / Violent offenders (N=55; M = 33.35) Control group (N=55; M = 30.38)	**HIB** Ambivalence task	I. PPI-R; II. BPAQ; MINI	- aggression was associated with HIB, respectively with a strong preference to interpret ambiguous facial expressions as hostile; - no biased interpretation of distress cues under conditions of uncertainty.
Czajkowska-Łukasiewicz et al., 2025 / Violent offenders: N = 125 (M = 38.67), from which 61 women Non-offenders: N = 183 (M = 38.29), 94 women	**HAB** AIHQ	I. HSNS; II. ARSQ	- vulnerable narcissism was positively associated with HAB, with rejection sensitivity as mediator; - rejection sensitivity had a significant indirect effect on HAB among controls, but not in the inmates.
Edwards & Bond (2012) / Offenders: N = 62 male mentally disordered offenders, (M = 39)	**HAB** Computerized Stories Task; SIP-AEQ	I. NPI; II. SCC	- significant correlation between the physical aggression category of the story and HAB; - HAB associated with writing more aggressive sentences after offensive stories displaying physical aggressive behavior.
Vitale et al., 2005 / Offenders: N=150, from which Caucasians (n=73; M = 28.20), African-Americans (n=77; M = 27.28)	**HAB** Ten hypothetical vignettes; Questions about attribution of intent	I. PCL-R II. ISQ; SCL–R 90	- PCL-R and ISQ scores were significant predictors of HAB; - HAB did not mediate the links between psychopathy and violent crime.
Hendry, 2013 (Dissertation thesis) / Offenders: N = 56 (M = 33.48), 34% on probation, 28 women; Civil psychiatric patients: N = 118 (M = 34.14), from which 55 women	**HAB** EHAS	I. PPI-R, PCL-SV; PCL-R II. SCID-I RV; STAXI-2; BPRS-E; CSS-M;	- HAB was a dynamic risk factor for violence and recidivism; - HAB was a significant predictor of violence and recidivism, especially in the short-term.

Author(s), year of publication / Participants (N/n= number, M=mean age)	Type of hostile bias and task(s) used	Assessment of aversive personality (I) and other variables (II)	Main findings
Dhalival, 2002 (Doctoral thesis, Study 2) / Offenders: (N= 70; M = 31.5)	**HAB** HAI	I. PCL-R II. SAR-FC2; My ETV; PDS,	- significant positive correlation between HAB and psychopathy in the case of ambiguous scenarios was observed; - HAB was indirectly related to aggression, with anger as a mediator.
Zajenkowska et al., 2025 / Offenders: homicide offenders - premeditated (N =74; M=36.79) and impulsive (N =39; M=29.25)	**HIB+HAB** Visual scenes; Emotion recognition on morphed faces; Emotion Preferences task; AIHQ	I. TriPM; II. Eye-tracking	- higher disinhibition and meanness - more hostile attributions in ambiguous relational harm scenarios. - higher levels of disinhibition, meanness, and boldness - reduced recognition of anger in ambiguous facial expressions; - higher meanness - delayed initial fixations on angry faces compared to other emotions, and longer gaze durations on male fearful faces compared to female fearful faces.
Stein et al., 2024 / Violent offenders: N = 65 (M = 40.86); Control group: N = 60, age-matched control participants (M = 43.40)	**HPB+HIB** Visual search tasks; Ambivalence task; Morphing task	I. PCL-R II. BPAQ; WMT	- no evidence for a fear deficit in violent offenders, nor for an association of psychopathy or aggression with impaired processing of fearful faces; - no evidence for a HPB for angry faces linked to psychopathy or aggression; - violent offenders showed a categorization bias (HIB) for anger, and this anger bias correlated with self-reported trait aggression (but not with psychopathy).
Wegrzyn, et al., 2017 / Offenders (N=45), convicted for violent offences (n = 30; M = 42), child sexual abuse (n = 15; M = 42) Control group (N=17, M = 43)	**HPB+HIB** Morphed faces experiment; Basic emotion recognition task; KDEF; Basic expression recognition task	I. PPI-R; II. SCID II; AFAS	- violent offenders rate ambiguous fear-anger face expressions as more angry (HPB), compared to both the control population and perpetrators of child sexual abuse; - a lowered threshold to detect anger in violent offenders compared to the general population was revealed.

Author(s), year of publication / Participants (N/n= number, M=mean age)	Type of hostile bias and task(s) used	Assessment of aversive personality (I) and other variables (II)	Main findings
Jusyte & Schönenberg, 2016 / Violent offenders (N = 34; M = 37.79) Control group (N=35; M = 30.51)	**HPB + HIB** Emotion sensitivity task; Ambiguous expressions task	I. SRP-III; II. BPAQ; WMT	- experiment 1 (emotional sensitivity) showed no group differences and no evidence for HPB; - experiment 2 revealed the presence of HIB in violent offenders, in the form of a deficit in the categorization of ambiguous fearful blends; - the impairment was associated with self-reported psychopathy and aggression only in violent offenders.

Notes. HBs measures: Visual search tasks (Jusyte et al., 2019); Ambivalence task (Jusyte & Schönenberg, 2017; Schönenberg & Jusyte, 2014); Morphing task (Schönenberg et al., 2013); Monomorph Facial Affect Task (facial affect recognition, Blair et al., 2001); Emotional Expression Labeling Test (FEEL, Kessler et al, 2002); Morphed faces experiment (GIMP and the GAP toolbox, www.gimp.org); Basic emotion recognition task; Face stimuli (KDEF, Lundqvist et al., 1998); NimStim, Tottenham et al., 2009); Basic expression recognition task (Hager et al., 2002); Emotion perception task (Penton-Voak et al., 2013); Emotion sensitivity task (FantaMorph, Abrosoft, Beijing, China); Ambiguous expressions task (FantaMorph, Abrosoft, Beijing, China); Ambiguous intentions and hostility questionnaire (AIHQ, Combs et al., 2007); Visual scenes (Wilkowski et al., 2007; Zajenkowska & Rajchert, 2020); Emotion recognition on morphed faces (Schönenberg & Jusyte, 2014); Emotion Preferences task; Ambiguous intentions and hostility questionnaire (AIHQ, Combs et al., 2007; Polish adaptation, Zajenkowska et al., 2020); Ambivalence task (Radboud faces database, Langner et al., 2010); Computerised Stories Task (Bond et al., 2004); SIP-AEQ (Coccaro et al., 2009); Ten hypothetical vignettes (adapted from Serin, 1991 & Rose et al., 1994); External Hostile Attributions Scale (EHAS, McNiel et al., 2003); Hostile Attributions Inventory (HAI, Epps & Kendall, 1995); NimStim for expression recognition (Tottenham et al., 2009); Levenson Self Report Psychopathy Scale (LSRP; Levenson et al.,1995);
Personality: Psychopathy Checklist- Revised (PCL-R, Hare, 2003); Psychopathy Checklist- Revised (PCL-R, Hare, 1991); Psychopathy Checklist- Revised (PCL-R, Hart et al., 1995); SCID II (First & Gibbon, 2004); Psychopathic Personality Inventory Revised (PPI-R, Lilienfeld & Widows, 2005); Self-reported Psychopathy Scale (SRP-III, Neumann et al., 2012); Hypersensitive narcissism scale (HSNS, Hendin & Cheek, 1997; Polish version: Czarna et al., 2014); Triarchic Psychopathy Measure (TriPM, Patrick et al., 2009; Polish adaptation, Pilch et al., 2015); Narcissistic Personality Inventory (NPI: Raskin & Terry, 1988); Psychopathy Checklist – Screening Version (PCL-SV; Hart et al., 1995); Structured Clinical Interview for the DSM-IV – Axis I Disorders Research Version (SCID-I RV; First et al., 1997)
Other variables: Buss-Perry Aggression Questionnaire (BPAQ, Buss & Perry, 1992); Scale of proactive and reactive violence (PROREA, Müller, 2014); State-Trait Anger Expression Inventory -2 (STAXI-2; Spielberger, 1999); Scenario Aggression Response Ratings-Forced Choice 2 (SAR-FC2, VanOostrum and Horvath, 1997); My Exposure to Violence (My ETV, Selner- O`Hagan et al., 1998); Appetitive and Facilitative Aggression Scale (AFAS, Köbach et al., 2015); Reactive-Proactive Aggression Questionnaire (RPQ, Raine et al., 2006); Impulsive/ Premeditated Aggression Scales (IPAS-30, Stanford et al., 2003); Social Dysfunction and Aggression Scale (SDAS-11, Wistedt, 1990); Rejection sensitivity questionnaire for adults (ARSQ, Berenson et al., 2009; Downey et al., 2006; Polish version: Bodecka et al., 2021);Self-Concept Clarity scale (SCC; Campbell et al., 1996); Inferential Styles Questionnaire (ISQ, Rose et al. 1994); Symptom Checklist–90–Revised (SCL–R 90, Derogatis, 1992); Mini International Neuropsychiatric Interview (MINI, Lecrubier et al., 1997, German version); Brief Psychiatric Rating Scale – Expanded Version (BPRS-E; Lukoff, et al., 1986); Criminal Sentiments Scale – Modified (CSS-M; Shields & Simourd, 1991); Paulhus Deception Scales (PDS, Paulhus, 1984); Marlowe-Crowne Form C (MC-C; Reynolds, 1982);Wiener Matrizen Test – short version,(WMT, Formann, Waldherr, & Piswanger, 2011); Shipley Institute of Living Scale-2 (SILS-2, Shipley, Gruber, & Martin, 2009); Wide Range Achievement Test (Reading Test), 4th edition (WRAT-4, Wilkinson & Robertson, 2006); Multiple-Choice Vocabulary Intelligence Test (MWT-B, Lehrl et al., 1995); Raven Standard Progressive Matrices (RSPM, Raven et al., 2000); Trail Making Test (TMT, Partington & Leiter, 1949); The beads-in-a-jar task (Huq et al., 1988); Eye-tracking (Real Eye – web-based software).

General considerations regarding the investigation of HBs

The analysis of the 14 studies revealed important conclusions about the manner in which we should understand each HB and the determinants involved in producing these distortions.

Although this scoping review did not intent to explain the variables that contribute to HB formation, however, it is still unclear how each HB is acquired and related to antisocial behaviors. This led researchers to study HBs separately, as if they were non-interacting phenomena. However, different papers used overlapping terminology for HPB, HIB and HAB without being very explicit about the level of processing they investigate and thus, creating confusion in conceptually delineating between the actual HB assessed.

In their desire to explore as comprehensively as possible the deficits that arise in information processing, researchers themselves came to consider that certain instruments which required participants only to name basic emotions or decode mixed emotional expressions actually measured the attribution of intention. However, for this stage to occur, higher-order cognitive processing must intervene.

It is extremely difficult to separate the phases of decoding the cues from the social information we permanently receive, and to refer to them as only perception, strictly interpretation, or just attribution of intent. In real-time situations, it is hard to clearly set limits on how long each sequence lasts; therefore, in recent research, complex methodologies have been developed to measure multiple HBs within the same study.

In drawing conclusions from the analyzed papers, we faced the same challenges. This prompted us to first present how specific tasks can be mapped to the measurement of different HBs, in close alignment with the sequences involved in decoding social information. Our approach is not motivated by a desire to create an artificial segregation of the phenomenon, but rather to foster a multidimensional understanding of the process and subsequently offer an integrative perspective on it. Therefore, from the outset we decided that important conceptual distinctions are necessary.

Differentiating between studies that tackled HPB and HIB was more difficult. To clearly differentiate the level of information processing at which a potential distortion occurs it was required the clarification of the following concepts: perceptual encoding, interpretation of information (assigning a meaning) and emotion sensitivity.

When impairments which generate HBs occur at a more basic level, involving simple perceptual processes involved in scanning the emotional landscape and intentions of others, this is translated in evaluating HPBs. To

effectively determine HPB in offenders, researchers used tasks exposing emotions through *facial expression* since processing facial indicators is one of the most direct sources of information in social interactions, (e.g. basic expression recognition task, emotion sensitivity task, etc.). Previously, it was posited that higher order HBs are mainly caused by impairments in emotion decoding, especially of anger and fear. When individuals perceive more anger or fear than is objectively present, a HPB is inferred. Therefore, for capturing HPB, the instruments used require participants to recognize and name basic emotions (*perceptual level*). For example, participants are asked to label the presence of a certain emotion when non-ambiguous stimuli (faces) are shown. The threshold for accurately perceiving the presence of a certain emotion offers hints about an important perceptual parameter – ***emotion sensitivity,*** which refers to how much emotional signal is required before a person detects or recognizes an emotion.

The *interpretation* bias occurs after the perceptual phase, when the person assigns a meaning to the perceived stimuli either by interpreting uncertain expressions as angry or misclassifying them as such. HIB emerges in a context of conflicting signals, which creates difficulty in accurately interpreting the facial cues. Ambiguity forces the assignment of a meaning, thus intervening interpretation of the stimuli, despite the completion of the perceptual phase. This leads to the production of HIB, by assigning a distorted meaning to what they previously perceived. In this regard, in the ***ambivalence task*** two distinct emotions coexist and are blended together in various proportions to create ambiguous stimuli, from low ambiguity pairs (containing 90% and 10% of each emotion), mid ambiguity (70%/30%), to high ambiguity pairs (50%/50%). The instrument requires to the participants to evaluate the emotional expression (e.g. happy, angry or fearful) in a forced-choice manner. Subsequently, subjects rate the intensity level of the identified emotion for each face on a rating scale ranging from 0 (not present at all) to 10 (full-blown emotion). Additionally, the ***morphing tasks*** are ideal candidates for exposing HIB by manipulating the emotion intensity through a continuum varying from one emotion to another. A morphing task is based on a perceptual categorization paradigm in which an emotion is gradually transformed along a continuum between two fully-saturated extremes in a succession of multiple sequences (from neutral 0% to 100%). Participants have to announce the recognition of the emotional expression as soon as they detect it while the researchers manipulate its intensity. In summary, HIB arise on ambiguous cues (ambivalent or morphed faces with unclear emotional signals) which change in seconds and require directing attentional resources for solving the task. Differing from HPB instruments, these tasks involve active changes and unclear cues which also allow measuring perceptual sensitivity as well as thresholds of detecting several emotions.

The investigation of HAB was more easily detectable through ***vignettes*** depicting ambiguous social scenarios that explicitly required (hostile) attributing intent to an actor's actions. However, *attributions of intentions* were assessed through ***self-reported questionnaires*** designed to measure the attributional style. For the latter, the items typically describe ambiguous interpersonal interactions and require respondents to rate how hostile the other person's intent is.

Specific considerations regarding measurements of HBs in offenders

As previously presented**,** several instruments are used to assess HBs, including questionnaires, vignettes, as well as computerized tasks (especially for displaying ambivalence and morphing tasks).

The present scoping review aims to provide a comprehensive analysis of instruments used for measuring HBs in offenders to reflect the level of manifestation of each distortion, starting from the perceptual stage, to interpretative and finally to cognitive processing. For every study included in the review, we preliminarily clarified the purpose for which it was used a certain methodology and subsequently classified the type of bias examined (see Table 2).

The use of vignettes, for example, did not automatically ensure inclusion of the study in the review, despite an initial tendency to consider it would meet the criteria. Several studies were not considered eligible after the content analysis, due to the use of social scenarios for investigating variables other than HBs, respectively criminal thinking patterns (Walters, 2007) or neural activation during personal space intrusion (Schienle et al., 2016).

Fourteen studies were selected that investigated the manifestation of HBs in the criminal population. Three of these studies used measurements to elicit perceptual impairments through signal detection theory (SDT; Stanislaw & Todorov, 1999) and recognition of facial emotional expressions. Two papers investigated HIB with ambivalence and morphing tasks, and five studies highlighted HAB through self-reported questionnaires and vignettes. Lastly, four studies employed a complex methodological approach to expose multiple HBs, with the purpose of clearly delineating the different phases of social information processing.

Specifically, the application of SDT in revealing HPB underscores the importance of distinguishing between perception sensitivity and response bias (i.e. the decision to report perceiving a certain emotion) when analyzing the perception deficits in forensic population (Gillespie et al., 2015; Faith et al, 2022). Furthermore, the ability to recognize facial expressions depicting basic emotions highlights the significance of emotion misinterpretation in explaining maladaptive behaviors across different categories of aggressive offenders (Philipp-Wiegmann et al., 2017).

The dynamics of facial expressions reflect immediate changes in internal emotional states (Eckman & Friesen, 1975), providing cues for adjusting social behaviors. With regard to this, the ability to recognize facial expressions in others has been demonstrated to influence aggressive behaviors, being impaired in individuals displaying antisocial behaviors (Sato et al., 2009; García-Sancho et al., 2015; Gillespie et al., 2015; Schönenberg et al., 2016). Therefore, the instruments that require the correct identification of emotions within ambivalence or morphing task were deemed appropriate for revealing HIBs. A large body of the selected studies investigated potential impairments in the accurate interpretation of emotions and thus detected HIB using morphed facial expression tasks (Wegrzyn et al., 2017; Faith et al., 2022; Stein et al., 2024; Zajenkowska et al., 2025) and ambivalence tasks (Schönenberg & Jusyte, 2014; Stein et al., 2024). These tasks require participants to identify the first/main emotion they observe when two primary emotions are combined, or when they are presented on a continuum from neutral to fully salient.

Finally, for assessing HAB, the analyzed papers used self-report instruments, such as External Hostile Attributions Scale (EHAS; Hendry, 2013), and also vignettes, which present uncertain and unclear social interactions. Regarding the self-report measures, they assessed distortions occurring in processing social information, providing valuable insights into an individual's subjective experience of depicting others' actions. Compared to other methods, they are easier to administer to larger populations and the results can be interpreted more efficiently, using standardized ratings.

One of the studies that used vignettes to evaluate the propensity to ascribe threatening purpose to others` intentions (Czajkowska-Łukasiewicz et al., 2025) displayed a commonly used instrument for HAB, respectively Ambiguous Intention Hostile Attribution Questionnaire (AIHQ; Combs et al., 2007). This task measures the extent to which the respondent attributes a hostile evaluation to certain stories with an unclear connection between characters or a diffuse social context. To capture specific cultural differences and refine the interpretation behind the provocateur's intention, this task was used in an adapted Polish version in a recent study (Zajenkowska et al., 2025). From the vignettes-category, we also mention the studies that used SIP-AEQ task and Stories Task (Edwards & Bond, 2012), Hostile Attribution Inventory (HAI; Dhalival, 2002), as well as an adapted version of Serin`s vignettes (Vitale et al., 2005).

Prior research investigated the presence of HBs by mainly relying on overlapping and confusing terminology, as well as a conceptual mixture of the different phases of SIP. Using a vignette-based methodology exclusively fails to consider the significance of basic perceptual stages and the influence of facial expressions as relevant emotional cues on the production of HBs in the early

phases of information processing. Nevertheless, facial expressions are crucial in social interactions and may trigger aggressive intentions and actions if misinterpreted. To correctly address these shortcomings and globally assess the multi-level mechanism of SIP, a more comprehensive explicative model was developed. The computational model (Smeijers, 2019) allows for the measurement of more than one HB by using multiple tasks in the same research for an integrative understanding of the process (e.g. emotion recognition on facial morphs, AIHQ, visual scenes, emotion preferences task, and eye tracking; Zajenkowska et al., 2025). Relatively similar methodologies have been used in other studies that measured multiple HBs in offenders, with results that led to significant conclusions about this mechanism of information processing, which cannot be isolated solely at the perceptual, interpretive, or cognitive level, but must be viewed as a whole (Jusyte & Schönenberg, 2016; Wegrzyn et al., 2017; Stein et al., 2024).

Findings from measuring HPB in offenders

Our summary revealed significant HPBs across the three analyzed studies, especially in aggressive offenders who are characterized by a marked tendency to perceive anger in ambiguous faces.

For understanding the prerequisites of HPB production, it should be acknowledged that altered recognition of emotional states is regarded as a primary cause of aggressive behavior, which is an essential characteristic of disorders like psychopathy and antisocial personality disorder (Marsh & Blair, 2008). Two significant yet different models have been postulated to link disruptive behavior to impairments in social information processing. On one side, in the *integrated emotion system theory* (Blair, 2005), an accurate perception and detection of distress cues ensures the suppression of aggressive impulses (Blair, 2005). Conversely, the *HAB model* (Milich & Dodge, 1984) accounts for aggression through a tendency to depict unclear expressions as angry, due to a cognitive misinterpretation that occurs during subsequent information processing.

In order to effectively assess whether impairments in facial affect recognition are attributable to a decisional process or are simply consequences of a perceptual sensitivity, SDT was employed, as it provides a more nuanced approach to the study of affect recognition in offenders characterized by high levels of aggression and psychopathy.

One study found that sexual offenders exhibited lower sensitivity (d′) to facial emotional expressions in general (across various intensities, genders of faces, and types of emotions) when compared to non-offenders (Gillespie et al., 2015). The results indicate that difficulties in recognizing facial emotions are

found not just in individuals who commit violent crimes, but also in those who commit sexual offenses. The displayed impairment with intense expressions of female anger and disgust indicates that their deficiencies are complex, emotion- and gender-related. This study was conducted with sexual perpetrators, connecting emotion recognition with HAB frameworks through the lens of SDT, distinguishing between sensitivity (d′) and decision bias (c). The findings support the idea that deficits in recognizing fear may significantly contribute to the likelihood of involvement in both violent and sexual offenses — potentially by diminishing their awareness of vulnerability or distress in others (Gillespie et al., 2015).

The previous conclusions are in line with another analyzed paper (Faith et al., 2022), which revealed that deficits in fear recognition were better explained by another mechanism than IES, which involves a hostile attribution, aside from that which causes low control of aggressive impulses in psychopathy.

Taken together, the previous studies claim the importance of SDT as a practical modality for elucidating inconsistent findings pertaining to the existence of a HPB arising from emotion recognition by disentangling perceptual sensitivity from the response bias.

Trying to capture the fear-anger sensitivity to emotional facial perception, Philipp-Wiegmann et al. (2017) examined differences among reactive offenders (with impulsive responses to provocation, marked by anger and lack of control, aimed at attacking the provocateur) and proactive perpetrators (with planned and goal-directed actions, seeking a desired outcome). As posited, in comparison with proactive violent offenders and the control group, the reactive offender group demonstrated a significantly lower capacity for emotion recognition, both in general and with regard to negative emotions such as sadness and disgust. Reactive violent offenders demonstrated a tendency to misperceive non-anger emotions as anger. In contrast, proactive violent offenders demonstrated a performance that was similar to that of the control group. Surprisingly, in visual search tasks, offenders and controls exhibited no substantial differences in their processing of fearful faces. Lastly, in the tasks involving perceptual sensitivity, individuals with high levels of psychopathy did not demonstrate a decreased threshold for recognizing anger earlier than other emotions.

These outcomes add a skeptical stance to the theories that posited that fear processing is impaired both in aggressive individuals and in psychopathy, and provide the perspective that aggression is much more linked to HAB that unfolds after the perceptual processing stage. In fact, someone can perceive the stimulus accurately, but interpret them as hostile due to beliefs, or personality dimensions.

This more comprehensive perspective indicates a step ahead from solely perceptual-deficit models (such as *integrated emotion system theory*, Blair, 2005) towards a more refined approach where cognitive biases play a central role in understanding aggression in individuals with psychopathy or other antagonistic personality traits.

Findings from measuring HIB in offenders

Considering that aggressive individuals struggle with identifying facial emotions, it can be inferred that they may encounter greater uncertainty in social situations and are consequently prone to adopting their maladaptive tendency of assuming the worst about others.

Our summary included two studies which aimed to explore if impairments in the interpretation of ambiguous social cues (i. e., facial expressions) are linked to aggressive behavior and psychopathy in violent and non-violent offenders (Schönenberg & Jusyte, 2014; Kuin et al., 2017).

Among the stimuli that are constantly decoded in current interactions, facial expressions convey a wide range of socially significant information that reflects the emotional states of other individuals. Therefore, difficulties in accurately interpreting subtle variations in facial signals during social interactions may contribute to a negative perception of others' emotions and intentions. Besides this, emotional facial expressions play a crucial role in modulating interpersonal connections. Indeed, substantial evidence indicates that deficits in recognizing clear facial emotions are associated with socially deviant and aggressive behavior (Marsh & Blair, 2008).

Although individuals who committed violent offenses showed a mild inclination to interpret ambiguous (neutral) faces as "*angrier*" compared to non-violent offenders or healthy individuals, the differences were not statistically significant (Kuin et al., 2017). These controversial outcomes indicate that the absence of highlighting HIB among violent inmates could be less consistent, or that its impact may be nuanced due to the characteristics of the sample (e.g., heterogeneity of the violent offences, low statistical power) or to methodological issues (e.g., tasks with binary choice).

In an effort to expose HIB production, Schönenberg & Jusyte (2014) demonstrated that aggressive offenders not only that they (mis)interpreted ambiguous facial signals as hostile, but also exhibited a pronounced inclination to consistently overrate the perceived intensity of anger compared to the control group. Their findings support the assumption that uncertain facial signals with elements of anger can act as strong indicators of provocations or imminent threats for aggressive individuals. Consequently, these evidences reveal the strong link between antisocial behavior and an enhanced perceptual sensitivity

of angry facial cues. However, the study found no evidence to support the hypothesis of a link between the degree of self-reported psychopathic traits and the propensity to interpret ambiguous cues as hostile, which conflicts with earlier research (Serin, 1991; Vitale et al., 2005).

The aforementioned research (mostly) shows that when cues are ambiguous, offenders, especially violent ones, are more prone to exhibit biases. Specifically, they might react aggressively by overestimating danger in situations where clarity is lacking rather than misreading clear cues.

Findings from measuring HAB in offenders

Defined as the tendency to consider the intentions of others as hostile in situations where social cues are unclear and unpredictable (Milich & Dodge, 1984), HAB is considered a significant factor in the development of problematic behaviors (Orobio de Castro et al., 2002), being the most investigated HB.

A consistent body of research was conducted on capturing the impact of ascribing a malevolent intention to others` actions, instigating problematic social interactions. Our summary analyzed five studies which explored the links between HAB and narcissism (Edwards & Bond, 2012; Czajkowska-Łukasiewicz et al., 2025), respectively psychopathy (Dhalival, 2002; Vitale et al., 2005; Hendry, 2013) mainly by using vignettes, but also tasks with pictorial stimuli.

Two decades ago, Vitale et al. (2005) investigated HAB in approximately 150 incarcerated adult males and revealed that, particularly in offenders, this bias seems to stem from at least two separate pathways: (1) one associated with psychopathy (marked by interpersonal hostility and deficits in response modulation) and (2) a more depressogenic/negative-schema pathway (encompassing broader negative beliefs about oneself and the world). These pathways demonstrate different relationships with personality traits, attributional styles, and criminal actions. Despite the promising findings, this study has one important limitation: it attempted to capture the role of HAB in offenders through only a few hypothetical situations and questions about the attribution of intent. These may not have accurately reflected a real-life context; therefore, the use of vignettes in this study may have influenced the non-significant outcomes of HAB mediating the link between psychopathy and violence.

For exposing the specific manifestation of HAB in offenders, in this scoping review, two unpublished research studies were also included alongside peer-reviewed articles, aligned to the eligibility criteria – a doctoral thesis (Dhalival, 2002) and a dissertation thesis (Hendry, 2013). Although not formally peer-reviewed, these works serve as supplementary evidence to enrich the published literature and to complete the existing research with new potential directions. Due to the changeability of HAB over time, it was explored its role as

a dynamic risk factor in producing violent behavior and reoffending is also a target in clinical interventions (Hendry, 2013). Previously, Dhaliwal (2002) found a significant association between HAB and aggression in ambiguous social scenarios, as well as a positive correlation between psychopathy and HAB. Dhaliwal proposed a model in which psychopathy and HAB contribute to anger, which then leads to affective (emotional) aggression. Specifically, state anger mediates the relationship between HAB and affective aggression, indicating that offenders who attribute hostile intent experience increased anger, which in turn predict aggressive behavior.

Expanding the area of research to other personality traits, Czajkowska-Łukasiewicz and colleagues (2025) compared Polish inmates and a sample from the general population, which revealed nuanced findings. It was hypothesized that individuals with higher scores in vulnerable narcissism tend to make more hostile attributions and that this link might be explained by sensitivity to rejection. The claimed link between HAB and vulnerable narcissism was validated, a positive correlation being emphasized, with rejection sensitivity acting as a mediator. However, only in the general population, rejection sensitivity demonstrates a significant indirect impact on HAB. The absence of mediation in the inmates group suggests that the path might operate differently and not exclusively through sensitivity to rejection, but could be explained by other contextualized mechanisms (differences in social cognition, history of violence, the incarceration experience etc.).

Based on the idea postulated by Bushman & Baumeister (1998) that narcissistic individuals react violently to threats because they seek to punish others and restore their sense of superiority, Edwards & Bond (2012) also investigated the attribution processes that mediate the relationship between narcissism and aggression in a sample of 62 mentally disabled offenders. A key finding was that narcissism made a smaller but still significant contribution to the prediction of HAB, while self-concept clarity was a strong predictor of HAB.

Taken together, these studies expand the avenues for exploring the links between multiple antagonistic personality traits and aggression, triggered by inappropriate information processing strategies. The path is already paved with outcomes that support HAB as a mediator between Dark Triad and relational aggression (Jiang et al., 2024).

Findings from measuring multiple HBs in offenders

In their efforts of disentangling the level in which HBs occur, researchers initiated complex methodologies for measuring more than one HB. In this regard, our summary identified four studies that addressed different distortions, with the aim of accurately pointing whether the bias is sensory, interpretative or attributional.

Following Smeijer's computational model (2019), research in the field of HBs has reached a turning point by diversifying the methodology in order to more clearly determine the levels at which impairments occur, adopting an integrative view of the whole information processing mechanism.

Previously to this innovative model, a consistent approach in examining both HPB and HIB was made by Jusyte and Schönenberg (2016). Their study emphasizes the idea that when assessing social-cognitive deficits in violent offenders, it's crucial to distinguish between *detection* vs. *categorization* tasks – just because someone detects emotional cues doesn't guarantee accurate interpretation. The aim was to investigate how aggression relates to the perception of facial expressions of anger and fear in violent offenders compared to healthy controls in two separate experiments. Actually, the findings from both experiments suggest that the fear deficit noted extensively in research on psychopathy, antisocial personality disorder, and aggression is likely due to an impairment in adequate categorizing (i.e., assigning a verbal label to) emotional expressions.

Actually, the aforementioned research enhances comprehension of social cognition impairments in violent individuals – it's not that they perceive danger in all situations (perception sensitivity), but they might incorrectly identify specific emotions (categorization), especially fear, when facial expressions are unclear (Jusyte & Schönenberg, 2016).

Investigating also HPB and HIB, Wegrzyn et al. (2017) showed that the HPB is particularly pronounced for male faces, suggesting that the effect depends on face gender. Violent offenders exhibited a diminished threshold for anger detection. These findings are consistent with those reported by Schönenberg & Jusyte (2014), who observed that when exposed to ambiguous face stimuli, impaired performances arise when anger is at the end of the emotion spectrum. Additionally, all participants showed a similar pattern of confusion, with fear being persistently confused with surprise, or disgust with anger. To note, though, fear is undeniably more challenging for even emotionally healthy individuals to identify compared to feelings like happiness and sadness (Russell, 1994; Elfenbein & Ambady, 2002).

More recently, using combined measures (vignettes, morphing task, eye-tracking) for revealing how hostile intent attribution appear from an early perceptual stage**,** Zajenkowska and colleagues (2025) investigated how the three facets of psychopathy (Disinhibition, Meanness, Boldness – The Triarchic Psychopathy Model, TriPM; Patrick et al, 2009) are associated to different types of HBs in homicide offenders (impulsive and premeditated). Surprisingly, individuals with high levels of psychopathy have reduced recognition of anger; rather than overestimating its presence, they fail to identify it in unclear facial expressions. This might indicate a complicated emotional-cognitive profile:

they might not be excessively alert to anger in facial expressions, yet still assess hostile intent through different methods (such as vignette-based or relational approaches).

This research enhances our understanding of psychopathy and hostile attribution, but it is simplistic to presume that individuals with psychopathy consistently perceive only threats around them. Certain dimensions of psychopathy (such as meanness and disinhibition) are associated with attributing hostility, yet these same traits may also correlate with a reduced awareness of facial expressions of anger.

Once the integrative models of HBs mechanism appeared, testing the consistency of previous ones increased. For example, one of the studies included in the present scoping review and investigating mixed HBs does not support the perspective (from IES) that aggression and psychopathy are primarily caused by an inability to recognize fear (Stein et al., 2024). Instead, the findings align better with the HAB model (Orobio De Castro et al., 2002): individuals who display aggression may not misinterpret emotions themselves, but rather view ambiguous emotional cues (particularly "mixed" signals) as more hostile or angry. To investigate the bias, the tasks used in the experimental phase were designed to separate bottom-up perceptual processing from higher-level cognitive labeling and/or interpretation. HAB seems to develop during later cognitive processes (after perception), during the labeling or interpretation of emotion, not during the initial perceptual detection.

Nevertheless, mixed-method approaches in exploring unique associations between personality facets and HBs can reveal surprising profiles of attributing hostility by those with dark sides of psychopathy.

Findings from measuring HBs in relation with (aversive) personality traits, aggression and other relevant dimensions

Exploring the flavored specificity that aversive personality traits can add to the manifestation of HBs in perpetrators, the most prevalent conclusion is that psychopathy significantly influences offenders' perceptions of social cues, often leading to an inflated perception of hostility (Vitale et al., 2005). It became clearer that offenders higher in psychopathy are more likely to perceive, interpret, and attribute malevolent meanings to social cues. These biases operate predominantly in the context of ambiguity, from misreading facial expressions to impaired interpretation of a social scenario, and over-attribution of hostile intention.

Consequently, aggressive offenders with high psychopathy scores expressed more anger when faced with hypothetical provocative scenarios and

assigned a higher degree of hostile intent to the ambiguous acting character when they imagined themselves as the victim in comparison to violent individuals without psychopathy (Serin, 1991).

Therefore, psychopathy emerges as a crucial factor that shapes hostile perception through reduced sensitivity to distress cues, lowered thresholds for detecting anger (Faith, 2022).

Furthermore, there have been limited studies that have comprehensively examined whether psychopathy is associated with diminished facial mimicry or reduced physiological responses to the dynamic facial expressions of others.

For example, psychopathy was not significantly associated with anger categorization bias (Stein et al., 2024). Moreover, there was no significant relationship between psychopathic traits (measured with PPI-R), intelligence, attention, or the tendency to jump to conclusions and how individuals perceived ambiguous facial expressions (Kuin et al., 2017). Nevertheless, these findings leave open questions about the robustness, specificity, and underlying mechanisms of this ability in psychopathic individuals.

Altogether, these mixed conclusions reinforce the hypothesis that HBs in offenders are intertwined with antagonistic personality traits. Still, antisocial manifestations do not rely exclusively on a psychopathic foundation. Despite the clinical levels of certain personality disorders, there are other facets, themes, or constellations of traits that can converge to aggression and violence, even from a non-clinical intensity.

Although investigating aversive personality traits related to HBs in offenders was mainly oriented towards psychopathy, seen as a crucial marker of antisocial personality, for our summary were found only two studies that explored the occurrence of HBs in relation to narcissism (Edwards & Bond, 2012; Czajkowska-Łukasiewicz et al., 2025).

One of the possible reasons for the low interest in expanding research to other personality traits might be that the majority of meta-analyses indicate that individuals with high levels of psychopathy exhibit lower accuracy in categorizing or recognizing typical facial emotional expressions of others (Wilson et al., 2011; Dawel et al., 2012).

Beyond the two aforementioned studies, there is a lack of research exploring the role of HBs in the relation between several other aversive personality traits that interfere with aggressive behavior, such as sadism, Machiavellianism, or even Dark Factor of personality (Moshagen et al., 2018). Although previous research has indicated theoretical support for a link between the dimensions of Dark Triad (Machiavellianism, psychopathy, and narcissism) and relational aggression, the mediating factors influencing the connection still require in-depth exploration (Jiang et al., 2024).

Inventorying the measurements employed for assessing *antagonistic personality traits*, the Psychopathy Checklist-Revised (PCL-R, Hare, 2003) was used in six studies (Dhalival, 2002; Vitale et al., 2005; Hendry, 2013; Philipp-Wiegmann et al., 2017; Faith et al., 2022; Stein et al., 2024). The self-reported level of psychopathy was assessed through the Psychopathic Personality Inventory–Revised (Hendry, 2013; Schönenberg & Jusyte, 2014; Wegrzyn, 2017), the Self-Reported Psychopathy Scale (Jusyte & Schönenberg, 2016), Levenson Self-Report Psychopathy Scale (Gillespie et al., 2015). One of the studies explored the different facets of psychopathy through The Triarchic Psychopathy Measure (Zajenkowska et al., 2025), deciphering the particular profiles that disinhibition, meanness, and boldness make in association with HAB

Aside from the influence of personality, research involving forensic population has shown conflicting results about the impact of HBs in determining aggressive responses. While some of the findings reinforced the impact of hostile processing of information in producing disruptive behaviors (Jusyte & Schönenberg, 2016; Kuin et al., 2017; Wegrzyn, 2017), other outcomes were contradictory (Vitale, 2005; Lobbestael et al., 2013). Possible explanations include confounded variables (e.g., age, IQ, social desirability, etc.), small and heterogeneous samples of offenders, or a lack of diversity in using appropriate tasks to detect potential links.

In measuring *aggressive behavior* and the propensity to develop diverse forms of hostile manifestations, the BPAQ (Buss & Perry, 1992) was used, a widely known instrument. Four studies describe using this task in revealing links between HBs and aggression (Schönenberg & Jusyte, 2014; Jusyte & Schönenberg, 2016; Kuin et al., 2017; Stein et al., 2024). All of them reported links between different facets of aggression and HBs. One of the analyzed studies used multiple tasks for deciphering the role of several facets of aggression in relation to HBs (Kuin et al., 2017).

For a more refined methodological approach, some studies used multiple instruments in order to capture the aggression spectrum (Dhalival, 2002; Kuin et al, 2017). The latter study employed instruments to measure also reactive vs. proactive aggression, impulsive vs. premeditated aggression, and a scale for social dysfunction and aggression. This study provides evidence that HIB is a risk mechanism for emotionally driven violence and that violent offenders exhibit a stronger correlation between HIB and reactive aggression.

Also, when the methodological frame was diversified (including self-reported tasks – vignettes, and perceptual – morphing tasks), the results outlined evident links between the measured variables (Edwards & Bond; Gillespie et al., 2015; Zajenkowska et al., 2025).

Gender differences observed in the association between HBs – aggression appear to be of scientific interest, although in this scoping review, twelve studies featured exclusively male samples, and only two examined particular features

in women offenders. Specifically, among female inmates, there was a medium correlation between vulnerable narcissism and HAB (Czajkowska-Łukasiewicz et al., 2025), but with no differences compared to the male offenders. Hendry (2013) showed that while men had considerably higher scores in HAB task than women, there was no significant correlation between gender and any of the outcomes (violence or recidivism).

The tendency to endorse socially acceptable traits was addressed in three of the included studies (Dhalival, 2002; Hendry, 2013; Gillespie et al., 2015) by using tasks designed to measure *social desirability*. Additionally, to control for covariates, *IQ* tasks were employed, both verbal and figural (Jusyte & Schönenberg, 2016; Kuin et al., 2017; Philipp-Wiegmann et al., 2017; Faith et al., 2022; Stein et al., 2024).

Regarding the samples` composition, most of the studies focused on revealing differences between violent and non-violent offenders, but mixed numerous offences under the same category (heterogeneous offences with no cut-off point regarding the level of violence used). Still, two studies of sexual offenders (Wegrzyn, 2017; Gillespie, 2015) revealed particular features of their HBs. For example, it was found that sexual offenders exhibit reduced sensitivity to emotional facial expressions (fear, anger, disgust) compared to non-offenders. This incorporates the idea of a broader impairment in decoding facial cues involved in emotion regulation and inhibition of violent behavior.

Therefore, a considerable focus is needed on ensuring greater sample homogeneity by offence type, as well as a more in-depth analysis of multiple separate categories of offences, to better detect variance in HBs manifestation.

In summary, the type of aggression is a key factor in the investigation of HBs which have been consistently linked to reactive aggression, a manifestation triggered by a perceived threatening and emotional disturbance. However, HBs are not associated with proactive or instrumental aggression, which is more typical to psychopathy. Thus, when aggression is split into subtypes, the association between HBs and proactive aggression becomes insignificant (Vitale et al., 2005; Philipp-Wiegmann et al., 2017).

DISCUSSION

For all the growing literature from the last two decades, there are still many challenges and controversies in understanding HBs, especially in incarcerated offenders, mainly because different research approaches tend to use distinct methodologies to explore various sides of information processing, while at the same time being limited to specific developmental stages, populations or to some psychiatric conditions.

Although some studies underline the role of HBs in producing antisocial behavior (Andrews & Bonta, 2010; Walters & DeLisi, 2013), there is still a lack of systematic research on the relationship between facets of aversive personality and aggressiveness determined by any kind of HBs, or on the potential differences between offenders and non-offenders.

Despite the aforementioned gaps, the inconsistent findings could be explained by the variety of methods used to capture different HBs manifestations, overlooking the role of individual differences in shaping personal and situational factors related to information processing. Also, the challenges and limitations inherent to the prison environment make the behavioral study of HBs less feasible. In this regard, the present scoping review clarifies the overall methodology and outcomes that were counted so far from studies that linked HBs in incarcerated inmates to dark personality features.

Several conclusions emerge from the analyses performed, aspects that will serve as future research directions in exploring the various variables that intervene across the path of information processing prior to the crime.

First, the studies selected for inclusion suggest that offenders have *higher levels of HBs than the general population*, which may trigger higher levels of aggressiveness, with the association being stronger for those with psychopathic traits (Schönenberg & Jusyte, 2014; Jusyte & Schönenberg, 2016), or for those exhibiting higher levels of narcissism (Edwards & Bond, 2012; Zajenkowska et al., 2025). Contrary to common assumptions, two of the analyzed studies did not support the hypothesis of HBs being stronger in offenders with higher scores in psychopathy (Vitale, 2005; Kuin et al., 2017). The inconsistencies rely on the different mechanisms underlying both psychopathy and HBs, which need to divide the variables involved into primary or secondary psychopathy or the type of aggression (proactive, reactive).

Secondly, despite these contradictory outcomes, a specific impairment seems to be particular to offenders even in *detecting basic emotions, with fear and anger as the most problematic*, being a relevant predictor for violent behaviors and reoffending (Hendry, 2013), although not all studies supported this basic deficit (Kuin et al. 2017; Stein et al., 2024).

Thirdly, combining findings from studies that investigated the perceptual level oh HBs, a core conclusion is that in forensic populations exhibiting aggressive behavior (especially reactive aggression), there is a *propensity to interpret ambiguous emotional cues as hostile* (specifically, angry). This hostile bias does not merely reflect a difficulty in recognizing emotions; instead, it seems to manifest more during the interpretative or decision-making stages of processing facial expressions. Consequently, rather than being "*blind*" to the emotions of others,

offenders may assign threats where others might not — and this interpretive bias could potentially fuel aggressive actions.

Then, another comprehensive finding across various studies is that particularly *HAB is notably higher among impulsive, reactive, and emotionally dysregulated offenders* rather than those committing instrumental or proactively planned violence. Recent research by Zajenkowska and colleagues (2025) indicated that impulsive homicide offenders possess significantly greater HAB compared to premeditated offenders, correlating this bias with secondary psychopathy and reactive aggression. Previously, Vitale et al. (2005) identified multiple pathways of HAB, revealing that offenders with high anger sensitivity or emotional dysregulation exhibit stronger tendencies for hostile intent. Collectively, these studies emphasize HAB as a critical factor distinguishing types of offenders, elucidating why some respond violently to ambiguous provocations while others act for strategic purposes.

Lastly, *a strong link between HAB, rejection sensitivity, and vulnerable narcissism was highlighted* within the study of Czajkowska-Łukasiewicz et al. (2025), emphasizing that HAB also stems from emotional fragility rather than exclusively from antisocial traits. Previously, Schönenberg & Jusyte (2014) illustrated that offenders with antisocial traits often misinterpret ambiguous facial expressions as threatening, suggesting that HAB is deeply rooted in emotion processing and threat detection systems. These findings expand the understanding of the explanatory frame of offender behavior by including emotional disturbance and interpersonal sensitivity, beyond already posited antisocial traits. We can conclude that especially HAB influences how offenders interpret interpersonal conflicts and ambiguous social interactions, operating beyond the perceptual level of threat perception.

This multi-faceted approach shows that HBs are operating as a cognitive-emotional filter, shaping how offenders perceive, interpret, and react to social cues. Thus, incorporating HBs in the explanatory model of offending enhances understanding of why some individuals escalate ambiguous or even neutral situations into conflict, while others do not.

Nevertheless, the mixed findings in the explored research may be attributed also to the specificity of the prison environment, which can sometimes interfere with the manifestation of HBs due to the adaptive strategies and the dynamics of the relationships between offenders. In some cases, the prison setting may affect the reduced level of hostile attributions, due to the process of habituation, which determines inmates to purposefully avoid hostile interpretations of others` actions to avoid more conflict and confrontation (Czajkowska-Łukasiewicz et al., 2025).

Likewise, social desirability has to be addressed, as it is known that offenders tend to align with the internal regulations that promote socially acceptable behaviors, fostered by a system of rewards and sanctions. Thus, the declared non-hostile approach of threatening social scenarios should be treated with caution, as they may try to portray themselves in a positive stance. Additionally, the lack of a connection between HBs and proactive aggression indicates that addressing proactive or instrumental violence may require focusing on different targets that could interfere (e.g., moral reasoning, reward evaluation, and theory of mind).

In conclusion, methodological issues, such as the use of heterogeneous tasks, small samples, and inadequate control over the confounding variables, may have led to the inconsistent results obtained in the area of how offenders decode social cues before acting in a socially unacceptable manner.

IMPLICATIONS

All research conducted within the correctional system leads to valuable insights into the mechanisms by which social information is processed by convicted individuals and plays a crucial role in shaping explanatory models on which effective rehabilitation strategies will be designed. Nevertheless, the findings should be regarded as preliminary, but with great potential for rehabilitation practices.

The multi-faceted block of research encompasses a broad range of topics, but in addition to this variety, given the limited number of studies, it may be interpreted that HBs have been approached inconsistently in understanding the biases that affect the information decoding process by offenders. Although it appears to employ the same principles, only a few studies examine similar dimensions and provide an in-depth look at specific aspects of HBs, making it challenging to compare outcomes and clarify contradictory findings.

More than that, in relation to the prison population, most research has focused on incarcerated males, particularly those involved in violent crimes. Although the majority of prisoners are male, limited research on female inmates and juvenile offenders creates a notable gap in understanding HBs of convicted individuals. While violent offenses carry significant negative repercussions, they do not represent the entirety of the prison demographic. Therefore, to achieve a nuanced and accurate understanding of the biases in attributing intent among convicted offenders, future research should explore these distortions in multiple types of offenses.

Acknowledging HBs as harmless deficits in interpreting ambiguous social scenarios can prompt more aggressive responses. Therefore, we stress the importance of understanding the mechanisms involved in HBs and Social

Information Processing more broadly, as this insight could aid in the creation of training programs focused on aggression regulation training. Additionally, individuals exhibiting severe aggressive behavior issues, such as forensic patients, should recognize their inclination to view ambiguous social situations as threatening. When individuals gain a better understanding of their HB, interventions can concentrate on teaching them to assess ambiguous social contexts more appropriately.

A greater focus on HBs and the methods of their assessment is crucial, as understanding the cognitive mechanisms and connections related to aggression can improve the development of more effective aggression training programs. For example, recent research posits that CBM-I can significantly increase the rate of positive interpretation bias and reduce the self-reported level of physical aggression in youth offenders (Ren et al., 2021). Also, HAB can be targeted using five-session training techniques with effects in reducing reactive aggression in adolescents (Van Bockstaele et al., 2020).

Hence, further investigation is needed to determine if the hostile bias in facial recognition serves as a crucial precursor (for instance, in youths prone to aggression), how it connects to aggressive reactions, and if it can be altered by traditional or innovative techniques. A recent study suggests that adjusting the categorical threshold between angry and happy facial expressions through biased feedback may enhance the perception of happiness in ambiguous expressions. This adjustment has been linked to a potential reduction in aggressive behavior among high-risk youths (Penton-Voak et al. 2013), with promising outcomes in designing future interventions, adult-tailored.

CONCLUSIONS

This scoping review reaffirms the strong link between HBs and aggressive manifestations in general, but especially with antisocial behavior in particular. This endeavor aimed to adopt an inclusive stance on previous research by summarizing literature on all types of HBs conducted with incarcerated offenders identified as having antagonistic personality facets.

In summary, this review has compiled the existing research investigating the link between HBs and various aversive personality traits, primarily psychopathy and narcissism, in adult offenders. The majority of studies reflected a strong correlation between the distortion of information in a hostile manner of processing, which characterizes incarcerated individuals, with aggression or cognitive biases. Additionally, the aversive personality structure contributed as a catalyst to the overall spiteful depiction of others' intentions.

Concluding, the research indicates that HBs develop through a variety of personality-based routes, some of which are directly linked to aggression, others to psychopathy. To analyze the distinction that proactive/reactive aggression and primary/secondary psychopathy bring to the apparently contradictory outcomes, both pathways need to be disaggregated. Moreover, the methodological variations lead to conflicting results due to the stages of processing that are measured. While tasks measuring emotion recognition or perception accuracy capture primary stages of processing, the social scenario tasks depict higher-order inference. Combining these types of measurements would lead to a more comprehensive approach to identifying significant differences in the occurrence of HBs. Finally, exploring the multidimensional aversive stance of personality may reveal particular features of social data processing that stand as pillars in repeated offending. An integrative approach to analyzing various paths to aggression that considers the roles of cognition, emotion, and personality in processing social information will bring more clarity in scientific research and impact the design of effective intervention models.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

Co-author Laura Visu-Petra was supported by a SNSF grant IZ11Z0_230823/F-RO-CH-2024-0328 during the writing of this paper.

REFERENCES

Allen, J. J., Anderson, C. A., & Bushman, B. J. (2018). The General Aggression Model. *Current Opinion in Psychology*, *19*, 75–80. https://doi.org/10.1016/j.copsyc.2017.03.034

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*(1), 27–51. https://doi.org/10.1146/annurev.psych.53.100901.135231

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law: An Official Law Review of the University of Arizona College of Law and the University of Miami School of Law*, *16*(1), 39–55. https://doi.org/10.1037/a0018362

Back, M. D., Küfner, A. C. P., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. A. (2013). Narcissistic admiration and rivalry: disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013–1037. https://doi.org/10.1037/a0034431

Bailey, C. A., & Ostrov, J. M. (2008). Differentiating forms and functions of aggression in emerging adults: Associations with hostile attribution biases and normative beliefs. *Journal of Youth and Adolescence*, *37*(6), 713–722. https://doi.org/10.1007/s10964-007-9211-5

Barlett, C. P. (2016). Exploring the correlations between emerging adulthood, Dark Triad traits, and aggressive behavior. *Personality and Individual Differences*, *101*, 293–298. https://doi.org/10.1016/j.paid.2016.05.061

Basquill, M. F., Nezu, C. M., Nezu, A. M., & Klein, T. L. (2004). Aggression-related hostility bias and social problem-solving deficits in adult males with mental retardation. *American Journal of Mental Retardation*, *109*(3), 255–263. https://doi.org/10.1352/0895-8017(2004)109%253C255:AHBASP%253E2.0.CO;2

Bate, C., Boduszek, D., Dhingra, K., & Bale, C. (2014). Psychopathy, intelligence and emotional responding in a non-forensic sample: an experimental investigation. *The Journal of Forensic Psychiatry & Psychology*, *25*(5), 600–612. https://doi.org/10.1080/14789949.2014.943798

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, *103*(1), 5–33. https://doi.org/10.1037/0033-295x.103.1.5

Berenson, K. R., Gyurak, A., Ayduk, O., Downey, G., Garner, M. J., Mogg, K., Bradley, B. P., & Pine, D. S. (2009). Rejection sensitivity and disruption of attention by social threat cues. *Journal of Research in Personality*, *43*(6), 1064–1072. https://doi.org/10.1016/j.jrp.2009.07.007

Blair, R. J. R. (2005). The neurobiology of antisocial behaviour and psychopathy. In A. Easton & N. J. Emery (Eds.), *The cognitive neuroscience of social behaviour* (pp. 291–324). Psychology Press. https://doi.org/10.4324/9780203311875_chapter_10

Blair, R. J. R., Colledge, E., Murray, L., & Mitchell, D. G. V. (2001). A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. *Journal of Abnormal Child Psychology*, 29(6), 491–498. https://doi.org/10.102 3/A:1012225108281

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

Blonigen, D. M., Sullivan, E. A., Hicks, B. M., & Patrick, C. J. (2012). Facets of psychopathy in relation to potentially traumatic events and posttraumatic stress disorder among female prisoners: the mediating role of borderline personality disorder traits. *Personality Disorders*, *3*(4), 406–414. https://doi.org/10.1037/a0026184

Bodecka-Zych, Marta, Jonason, P. K., & Zajenkowska, A. (2021). Hostile attribution biases in vulnerable narcissists depends on the Socio-relational context. *Journal of Individual Differences*, *43*(2), 70–78. https://doi.org/10.1027/1614-0001/a000354

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

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of Personality and Social Psychology*, *75*(1), 219–229. https://doi.org/10.1037/0022-3514.75.1.219

Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, *63*(3), 452–459. https://doi.org/10.1037//0022-3514.63.3.452

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, *70*(1), 141–156. https://doi.org/10.1037/0022-3514.70.1.141

Ciucci, E., Baroncelli, A., Facci, C., Righi, S., & Frick, P. J. (2024). Callous-unemotional traits and emotion perception accuracy and bias in youths. *Children (Basel, Switzerland)*, *11*(4), 419. https://doi.org/10.3390/children11040419

Coccaro, E. F., Fanning, J. R., Fisher, E., Couture, L., & Lee, R. J. (2017). Social emotional information processing in adults: Development and psychometrics of a computerized video assessment in healthy controls and aggressive individuals. *Psychiatry Research*, *248*, 40–47. https://doi.org/10.1016/j.psychres.2016.11.004

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

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): a new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, *12*(2), 128–143. https://doi.org/10.1080/13546800600787854

Costa, P. T., Jr, & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

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), 74–101. https://doi.org/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. https://doi.org/10.2307/1131875

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. https://doi.org/10.1111/1467-8624.00462

Czajkowska-Łukasiewicz, K., Iwon, K., Zajenkowska, A., & Smoleń, S. (2025). Exploring the links between rejection sensitivity, vulnerable narcissism, and hostile attributions in inmates and non-incarcerated individuals. *Personality and Individual Differences*, *240*(113171), 113171. https://doi.org/10.1016/j.paid.2025.113171

Czarna, A. Z., Dufner, M., & Clifton, A. D. (2014). The effects of vulnerable and grandiose narcissism on liking-based and disliking-based centrality in social networks. *Journal of Research in Personality*, *50*, 42–45. https://doi.org/10.1016/j.jrp.2014.02.004

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

Derogatis, L. R. & Psychometric, R. C. (1992). SCL-90-R: Administration, scoring & procedures manual-II, for the revised version and other instruments of the psychopathology rating scale series. Clinical Psychometric Research, Towson

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development*, *51*(1), 162–170. https://doi.org/10.1111/j.1467-8624.1980.tb02522.x

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

Dodge, Kenneth A., Price, J. M., Bachorowski, J.-A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology*, *99*(4), 385–392. https://doi.org/10.1037/0021-843x.99.4.385

Downey, G., Berenson, K. R., & Kang, J. (2006). *The adult rejection sensitivity questionnaire (ASRQ)*. Columbia University

Eckman, P., & Friesen, W. (1975). *Unmasking the Face*. Prentice Hall.Edwards, R., & Bond, A. J. (2012). Narcissism, self-concept clarity and aggressive cognitive bias amongst mentally disordered offenders. *The Journal of Forensic Psychiatry & Psychology*, *23*(5–6), 620–634. https://doi.org/10.1080/14789949.2012.715180

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037/0033-2909.128.2.203

Epps, J., & Kendall, P. C. (1995). Hostile attributional bias in adults. *Cognitive Therapy and Research*, *19*(2), 159–178. https://doi.org/10.1007/bf02229692

Eriksson, M. J., & Schmidt, L. A. (2026). Characterizing the dark side of social anxiety in adolescence: A replication and extension study. *Personality and Individual Differences*, *251*(113581), 113581. https://doi.org/10.1016/j.paid.2025.113581

Faith, R. N., Miller, S. A., & Kosson, D. S. (2022). Facial affect recognition and psychopathy: A signal detection theory perspective. *Journal of Psychopathology and Behavioral Assessment*, *44*(3), 738–749. https://doi.org/10.1007/s10862-022-09969-5

First MB, Gibbon M. (2004). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). New York: Biometric Research Department, New York Psychiatric Hospital

First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W. (1997). Structured clinical interview for DSM-IV (axis I disorders) clinician version (SCID-CV). American Psychiatric Press, Washington, DC

Formann, A. K., Waldherr, K., & Piswanger, K. (2011). Wiener Matrizen–Test 2. Manual. Beltz Test GmbH

García-Sancho, E., Salguero, J. M., & Fernández-Berrocal, P. (2015). Déficits en el reconocimiento facial de las emociones y su relación con la agresión: Una revisión sistemática [Deficits in facial affect recognition and aggression: A systematic review]. *Ansiedad y Estrés, 21*(1), 1–20. https://www.ansiedadyestres.es/sites/default/files/rev/ucm/2015/anyes2015a1.pdf

Gillespie, S. M., Rotshtein, P., Satherley, R.-M., Beech, A. R., & Mitchell, I. J. (2015). Emotional expression recognition and attribution bias among sexual and violent offenders: a signal detection analysis. *Frontiers in Psychology*, *6*, 595. https://doi.org/10.3389/fpsyg.2015.00595

Hager, J. C., Ekman, P., & Friesen, W. V. (2002). *Facial action coding system. Salt Lake City: A Human Face*. ISBN 0-931835-01-1

Hare, R.D. (2003). Manual for the revised psychopathy checklist. 2nd Edition, *Multi-Health Systems*, Toronto

Hart SD, Cox DN, Hare RD (1995) Manual for the psychopathy checklist: Screening version (PCL: SV). *Multi-Health Systems*,Toronto

Hart, S. D., & Hare, R. D. (1996). Psychopathy and antisocial personality disorder. *Current Opinion in Psychiatry*, *9*(2), 129–132. https://doi.org/10.1097/00001504-199603000-00007

Hartmann, D., Ueno, K., & Schwenck, C. (2020). Attributional and attentional bias in children with conduct problems and callous-unemotional traits: a case-control study. *Child and Adolescent Psychiatry and Mental Health*, *14*(1), 9. https://doi.org/10.1186/s13034-020-00315-9

Hendin, H. M., & Cheek, J. M. (1997). Assessing hypersensitive narcissism: A reexamination of Murray's narcism scale. *Journal of Research in Personality*, *31*(4), 588–599. https://doi.org/10.1006/jrpe.1997.2204

Hornsveld, R. H. J., Nijman, H. L. I., Hollin, C. R., & Kraaimaat, F. W. (2007). An adapted version of the Rosenzweig Picture-Frustration Study (PFS-AV) for the measurement of hostility in violent forensic psychiatric patients. *Criminal Behaviour and Mental Health*, *17*(1), 45–56. https://doi.org/10.1002/cbm.638

Huesmann, L. R. (1998). The role of social information processing and cognitive schema in the acquisition and maintenance of habitual aggressive behavior. In *Human Aggression* (pp. 73–109). Elsevier.

Hurezan, L., Turi, A., Ion, A., & Visu-Petra, L. (2024). Dark and bright personality dimensions as predictors of criminal behavior and recidivism. *Scientific Reports*, *14*(1), 18565. https://doi.org/10.1038/s41598-024-69288-5

Hutchings, J. N., Gannon, T. A., & Gilchrist, E. (2010). A preliminary investigation of a new pictorial method of measuring aggression-supportive cognition among young aggressive males. *International Journal of Offender Therapy and Comparative Criminology*, *54*(2), 236–249. https://doi.org/10.1177/0306624X08325350

Huq, S. F., Garety, P. A., & Hemsley, D. R. (1988). Probabilistic judgements in deluded and non-deluded subjects. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *40*(4), 801–812. https://doi.org/10.1080/14640748808402300

Jiang, Y., Tong, L., Cao, W., & Wang, H. (2024). Dark Triad and relational aggression: the mediating role of relative deprivation and hostile attribution bias. *Frontiers in Psychology*, *15*, 1487970. https://doi.org/10.3389/fpsyg.2024.1487970

Jones, S. E., Miller, J. D., & Lynam, D. R. (2011). Personality, antisocial behavior, and aggression: A meta-analytic review. *Journal of Criminal Justice*, *39*(4), 329–337. https://doi.org/10.1016/j.jcrimjus.2011.03.004

Jusyte, A., & Schönenberg, M. (2016). Impaired social cognition in violent offenders: perceptual deficit or cognitive bias? *European Archives of Psychiatry and Clinical Neuroscience*, *267*(3), 257–266. https://doi.org/10.1007/s00406-016-0727-0

Jusyte, A., Stein, T., & Schönenberg, M. (2019). Fear processing deficit in violent offenders: Intact attentional guidance but impaired explicit categorization. *Psychology of Violence*, *9*(3), 308–318. https://doi.org/10.1037/vio0000109

Kersten, R., & Greitemeyer, T. (2024). Human aggression in everyday life: An empirical test of the general aggression model. *The British Journal of Social Psychology*, *63*(3), 1091–1111. https://doi.org/10.1111/bjso.12718

Kessler H, Bayerl P, Deighton RM, Traue HC (2002) Facially expressed emotion labeling (FEEL): PC-gestützter Test zur Emotionserkennung.Verhaltenstherapie Verhaltensmedizin. 23:297–306

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

Köbach, A., Schaal, S., & Elbert, T. (2015). Combat high or traumatic stress: violent offending is associated with appetitive aggression but not with symptoms of traumatic stress. *Frontiers in Psychology*, *5*, 1518. https://doi.org/10.3389/fpsyg.2014.01518

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. https://doi.org/10.1016/j.appdev.2017.07.007

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

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud faces database. *Cognition & Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., Janavs, J., & Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *12*(5), 224–231. https://doi.org/10.1016/s0924-9338(97)83296-8

Lehrl, S., Triebig, G., & Fischer, B. (1995). Multiple choice vocabulary test MWT as a valid and short test to estimate premorbid intelligence. *Acta Neurologica Scandinavica*, *91*(5), 335–345. https://doi.org/10.1111/j.1600-0404.1995.tb07018.x

Levenson, M. R., Kiehl, K. A., & Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutionalized population. *Journal of Personality and Social Psychology*, *68*(1), 151–158. https://doi.org/10.1037/0022-3514.68.1.151

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: State Narcissism and Aggression. *Aggressive Behavior*, *42*(4), 333–345. https://doi.org/10.1002/ab.21629

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology*, *62*(10), e1-34. https://doi.org/10.1016/j.jclinepi.2009.06.006

Lilienfeld SO, Widows MR, Staff PAR. Psychopathic Personality Inventory TM Revised. Soc Influ SOI. 2005; 61:97

Lim, L., Day, A., & Casey, S. (2011). Social cognitive processing in violent male offenders. *Psychiatry, Psychology, and Law: An Interdisciplinary Journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, *18*(2), 177–189. https://doi.org/10.1080/13218711003739490

Lin, S., Wang, Y., Cheng, G., & Bai, X. (2023). Relationship between harsh parenting and aggressive behaviors in male juvenile delinquents: Potential mediating roles of peer victimization and hostile attribution bias. *Behavioral Sciences*, *13*(7), 610. https://doi.org/10.3390/bs13070610

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

Lukoff, D., Nuechterlien, K., & Ventura, J. (1986). Manual for the Expanded Brief Psychiatric Rating Scale. *Schizophrenia Bulletin*, *13*, 261–276.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *Karolinska Directed Emotional Faces (KDEF)* [Database record]. APA PsycTests. https://doi.org/10.1037/t27732-000

MacBrayer, E. K., Milich, R., & Hundley, M. (2003). Attributional biases in aggressive children and their mothers. *Journal of Abnormal Psychology*, *112*(4), 698–708. https://doi.org/10.1037/0021-843X.112.4.598

Maccoon, D. G., & Newman, J. P. (2006). Content meets process: Using attributions and standards to inform cognitive vulnerability in psychopathy, antisocial personality disorder, and depression. *Journal of Social and Clinical Psychology*, *25*(7), 802–824. https://doi.org/10.1521/jscp.2006.25.7.802

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment*, *27*(2), 101–114. https://doi.org/10.1007/s10862-005-5384-y

Marsh, A. A., & Blair, R. J. R. (2008). Deficits in facial affect recognition among antisocial populations: a meta-analysis. *Neuroscience and Biobehavioral Reviews*, *32*(3), 454–465. https://doi.org/10.1016/j.neubiorev.2007.08.003

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. https://doi.org/10.1111/j.1559-1816.2002.tb01418.x

McNiel, D. E., Eisner, J. P., & Binder, R. L. (2003). The relationship between aggressive attributional style and violence by psychiatric patients. *Journal of Consulting and Clinical Psychology*, *71*(2), 399–403. https://doi.org/10.1037/0022-006x.71.2.399

Milich, R., & Dodge, K. A. (1984). Social information processing in child psychiatric populations. *Journal of Abnormal Child Psychology*, *12*(3), 471–489. https://doi.org/10.1007/bf00910660

Miller, J. D., & Lynam, D. R. (2006). Reactive and proactive aggression: Similarities and differences. *Personality and Individual Differences*, *41*(8), 1469–1480. https://doi.org/10.1016/j.paid.2006.06.004

Moshagen, M., Hilbig, B. E., & Zettler, I. (2025). Reconceptualizing ethically and socially aversive ("dark") personality traits. *Current Opinion in Psychology*, *66*(102111), 102111. https://doi.org/10.1016/j.copsyc.2025.102111

Müller, S. A. (2014). *ProRea-rating scale of proactive and reactive violence [ProRea-ein Instrument zur Klassifizierung gewalttätigen Verhaltens]*. Dissertation.

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. *International Journal of Law and Psychiatry*, *38*, 92–99. https://doi.org/10.1016/j.ijlp.2015.01.012

Neumann, C. S., Schmitt, D. S., Carter, R., Embley, I., & Hare, R. D. (2012). Psychopathic traits in females and males across the globe: Psychopathic traits. *Behavioral Sciences & the Law*, *30*(5), 557–574. https://doi.org/10.1002/bsl.2038

Orobio de Castro, B., 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. https://doi.org/10.1111/1467-8624.00447

Partington, J. E., & Leiter, R. G. (1949). Partington's Pathway Test. *The Psychological Service Center Bulletin*

Patrick, C. J., Fowles, D. C., & Krueger, R. F. (2009). Triarchic conceptualization of psychopathy: developmental origins of disinhibition, boldness, and meanness. *Development and Psychopathology*, *21*(3), 913–938. https://doi.org/10.1017/S0954579409000492

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*(3), 598–609. https://doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L., Curtis, S. R., & Jones, D. N. (2018). Aggression as a trait: the Dark Tetrad alternative. *Current Opinion in Psychology*, *19*, 88–92. https://doi.org/10.1016/j.copsyc.2017.04.007

Penton-Voak, I. S., Thomas, J., Gage, S. H., McMurran, M., McDonald, S., & Munafò, M. R. (2013). Increasing recognition of happiness in ambiguous facial expressions reduces anger and aggressive behavior. *Psychological Science*, *24*(5), 688–697. https://doi.org/10.1177/0956797612459657

Philipp-Wiegmann, F., Rösler, M., Retz-Junginger, P., & Retz, W. (2017). Emotional facial recognition in proactive and reactive violent offenders. *European Archives of Psychiatry and Clinical Neuroscience*, *267*(7), 687–695. https://doi.org/10.1007/s00406-017-0776-z

Pilch, I., Sanecka, E., Hyla, M., & Atłas, K. (2015). Polska adaptacja skali TRIPM do badania psychopatii w ujęciu triarchicznym. *Psychologia Społeczna, 4*, 435–454. https://doi.org/10.7366/1896180020153506

Potter, K. (2024). Seeing Red: Hostile Attribution Bias, Aggression, and Antisocial Personality Disorder Traits. Graduate Theses, Dissertations, and Problem Reports. 12310. https://researchrepository.wvu.edu/etd/12310

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., Stouthamer-Loeber, M., & Liu, J. (2006). The reactive-Proactive Aggression Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, *32*(2), 159–171. https://doi.org/10.1002/ab.20115

Raskin, R., & Terry, H. (1988). A principal-components analysis of the narcissistic personality inventory and further evidence of its construct validation. *Journal of Personality and Social Psychology*, *54*, 890–902

Raven, J. C., Court, J. H., & Raven, J. (2000). *Raven Manual: section 3. Standard Progressive Matrices*. Pearson Assessment and Information

Ren, Z., Zhao, Z., Yu, X., Zhang, L., & Li, X. (2021). Effects of cognitive bias modification for interpretation on hostile interpretation bias and self-reported aggression in juvenile delinquents. *International Journal of Clinical and Health Psychology*, *21*(2), 100226. https://doi.org/10.1016/j.ijchp.2021.100226

Reynolds, W. M. (1982). Development of reliable and valid short forms of the marlowe-crowne social desirability scale. *Journal of Clinical Psychology*, *38*(1), 119–125. https://doi.org/10.1002/1097-4679(198201)38:1<119::aid-jclp2270380118>3.0.co;2-i

Rose, D. T., Abramson, L. Y., Hodulik, C. J., Halberstadt, L., & Leff, G. (1994). Heterogeneity of cognitive style among depressed inpatients. *Journal of Abnormal Psychology*, *103*(3), 419–429. https://doi.org/10.1037//0021-843x.103.3.419

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*(1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102

Sato, W., Uono, S., Matsuura, N., & Toichi, M. (2009). Misrecognition of facial expressions in delinquents. *Child and Adolescent Psychiatry and Mental Health*, *3*(1), 27. https://doi.org/10.1186/1753-2000-3-27

Schienle, A., Wabnegger, A., Leitner, M., & Leutgeb, V. (2017). Neuronal correlates of personal space intrusion in violent offenders. *Brain Imaging and Behavior*, *11*(2), 454–460. https://doi.org/10.1007/s11682-016-9526-5

Schönenberg, M., & Jusyte, A. (2014). Investigation of the hostile attribution bias toward ambiguous facial cues in antisocial violent offenders. *European Archives of Psychiatry and Clinical Neuroscience*, *264*(1), 61–69. https://doi.org/10.1007/s00406-013-0440-1

Schönenberg, M., Louis, K., Mayer, S., & Jusyte, A. (2013). Impaired identification of threat-related social information in male delinquents with antisocial personality disorder. *Journal of Personality Disorders*, *27*(4), 496–505. https://doi.org/10.1521/pedi_2013_27_100

Schönenberg, M., Mayer, S. V., Christian, S., Louis, K., & Jusyte, A. (2016). Facial Affect Recognition in Violent and Nonviolent Antisocial Behavior Subtypes. *Journal of Personality Disorders*, *30*(5), 708–719. https://doi.org/10.1521/pedi_2015_29_217

Selner-O'Hagan, M. B., Kindlon, D. J., Buka, S. L., Raudenbush, S. W., & Earls, F. J. (1998). Assessing exposure to violence in urban youth. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *39*(2), 215–224. https://doi.org/10.1111/1469-7610.00315

Serin, R. C. (1991). Psychopathy and violence in criminals. *Journal of Interpersonal Violence*, *6*(4), 423–431. https://doi.org/10.1177/088626091006004002

Shields, I. W., & Simourd, D. J. (1991). Predicting predatory behavior in a population of incarcerated young offenders. *Criminal Justice and Behavior, 18*(2), 180–194. https://doi.org/10.1177/0093854891018002006

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). Shipley-2. *Los Angeles, CA: Western Psychological Services*. doi: https://doi.org/10.1037/t48948-000

Slaby, R. G., & Guerra, N. G. (1988). Cognitive mediators of aggression in adolescent offenders: I. Assessment. *Developmental Psychology*, *24*(4), 580–588. https://doi.org/10.1037/0012-1649.24.4.580

Smeijers, D. (2022). Hostility Bias: A key-characteristic of aggressive behavior. In *Handbook of Anger, Aggression, and Violence* (pp. 1–20). Springer International Publishing.

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

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: Hostile Inerpretation Bias and Pathological Aggression. *Aggressive Behavior*, *43*(4), 386–397. https://doi.org/10.1002/ab.21697

Spielberger, C. D. (1999). Manual for the State-Trait Anger Expression Inventory-2. Odessa, FL: Psychological Assessment Resources

Stalenheim, E. G. (2004). Long-term validity of biological markers of psychopathy and criminal recidivism: follow-up 6-8 years after forensic psychiatric investigation. *Psychiatry Research*, *121*(3), 281–291. https://doi.org/10.1016/j.psychres.2003.07.002

Stanford, M. S., Houston, R. J., Mathias, C. W., Villemarette-Pittman, N. R., Helfritz, L. E., & Conklin, S. M. (2003). Characterizing aggressive behavior. *Assessment*, *10*(2), 183–190. https://doi.org/10.1177/1073191103010002009

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1), 137–149. https://doi.org/10.3758/BF03207704

Stein, T., Gehrer, N., Jusyte, A., Scheeff, J., & Schönenberg, M. (2024). Perception of emotional facial expressions in aggression and psychopathy. *Psychological Medicine*, *54*(12), 1–9. https://doi.org/10.1017/S0033291724001417

Subra, B. (2023). Why narcissists are more likely to be aggressive? The role of hostile attribution bias. *International Journal of Psychology: Journal International de Psychologie*, *58*(6), 518–525. https://doi.org/10.1002/ijop.12924

Tharshini, N. K., Ibrahim, F., Kamaluddin, M. R., Rathakrishnan, B., & Che Mohd Nasir, N. (2021). The link between individual personality traits and criminality: A systematic review. *International Journal of Environmental Research and Public Health*, *18*(16), 8663. https://doi.org/10.3390/ijerph18168663

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. https://doi.org/10.1016/j.psychres.2008.05.006

Ursulet, A., de Repentigny, É., E. Quansah, J., Bessette, M., & Gagnon, J. (2022). The mediating role of hostile attribution bias in the relationship between cluster B personality traits and reactive aggression: An event-related potentials study. In *Aggression and Violent Behaviour [Working Title]*. IntechOpen

Van Bockstaele, B., van der Molen, M. J., van Nieuwenhuijzen, M., & Salemink, E. (2020). Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents. *Journal of Experimental Child Psychology*, *194*(104811), 104811. https://doi.org/10.1016/j.jecp.2020.104811

VanOostrum, N., & Horvath, P. (1997). The effects of hostile attribution on adolescents' aggressive responses to social situations. *Canadian Journal of School Psychology*, *13*(1), 48–59. https://doi.org/10.1177/082957359701300105

Verhoef, R. E. J., Alsem, S. C., Verhulp, E. E., & De Castro, B. O. (2019). Hostile intent attribution and aggressive behavior in children revisited: A meta-analysis. *Child Development*, *90*(5), e525–e547. https://doi.org/10.1111/cdev.13255

Vitale, J. E., Newman, J. P., Serin, R. C., & Bolt, D. M. (2005). Hostile attributions in incarcerated adult male offenders: An exploration of diverse pathways. *Aggressive Behavior*, *31*(2), 99–115. https://doi.org/10.1002/ab.20050

Walters, G. D. (2007). Measuring proactive and reactive criminal thinking with the PICTS: correlations with outcome expectancies and hostile attribution biases: Correlations with outcome expectancies and hostile attribution biases. *Journal of Interpersonal Violence*, *22*(4), 371–385. https://doi.org/10.1177/0886260506296988

Walters, G. D., & DeLisi, M. (2013). Antisocial cognition and crime continuity: Cognitive mediation of the past crime-future crime relationship. *Journal of Criminal Justice*, *41*(2), 135–140. https://doi.org/10.1016/j.jcrimjus.2012.12.004

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. https://doi.org/10.1186/s40359-017-0186-z

Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT 4: Wide Range Achievement Test*. Lutz, FL: Psychological Assessment Resources. doi: https://doi.org/10.1037/e672132007-004

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. https://doi.org/10.1016/j. jrp.2006.07.003

Wilson, K., Juodis, M., & Porter, S. (2011). Fear and loathing in psychopaths: A meta-analytic investigation of the facial affect recognition deficit. *Criminal Justice and Behavior*, *38*(7), 659–668. https://doi.org/10.1177/0093854811404120

Wistedt, B., Rasmussen, A., Pedersen, L., Malm, U., Träskman-Bendz, L., Wakelin, J., & Bech, P. (1990). The development of an observer-scale for measuring social dysfunction and aggression. *Pharmacopsychiatry*, *23*(6), 249–252. https://doi.org/10.1055/s-2007-1014514

Zajenkowska, A., Jakubowska, A., Rajchert, J., Koszałkowska, K., Bodecka-Zych, M., & Gehrer, N. (2025). Psychopathy and hostile attributions in impulsive and premeditated homicide offenders. *Personality and Individual Differences*, *247*(113400), 113400. https://doi.org/10.1016/j.paid.2025.113400

Zajenkowska, A., Prusik, M., Jasielska, D., & Szulawski, M. (2021). Hostile attribution bias among offenders and non-offenders: Making social information processing more adequate. *Journal of Community & Applied Social Psychology*, *31*(2), 241–256. https://doi.org/10.1002/casp.2493

Zajenkowska, A., & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology (Hove, England)*, *32*(2), 180–198. https://doi.org/10.1080/20445911.2020.1717498