HHS Public Access

Author manuscript

Cognition. Author manuscript; available in PMC 2018 October 01.

Published in final edited form as:

Cognition. 2017 October ; 167: 38–45. doi:10.1016/j.cognition.2017.03.013.

Emotion-based learning systems and the development of morality

R. J. R. Blair, Ph.D.[1,*]

[1]Section on Affective Cognitive Neuroscience, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA

Abstract

In this paper it is proposed that important components of moral development and moral judgment rely on two forms of emotional learning: stimulus-reinforcement and response-outcome learning. Data in support of this position will be primarily drawn from work with individuals with the developmental condition of psychopathy as well as fMRI studies with healthy individuals. Individuals with psychopathy show impairment on moral judgment tasks and a pronounced increased risk for instrumental antisocial behavior. It will be argued that these impairments are developmental consequences of impaired stimulus-aversive conditioning on the basis of distress cue reinforcers and response-outcome learning in individuals with this disorder.

Introduction

The field of moral development and moral reasoning has been transformed over the past twenty years. In the early 90s, there was one dominant view: moral reasoning involved decisions based on accessing conceptual domains (e.g., Colby & Kohlberg, 1987). Moral development involved the construction of these conceptual domains through some form of rational thought processes. With very few exceptions (e.g., Kagan & Lamb, 1987), there was little consideration given to any role of emotion in moral development/reasoning. Then, in the mid 1990s, the first studies to use results from psychopathy to infer the core role of emotion in moral development were conducted (e.g., R. J. R. Blair, 1995; R.J.R. Blair, Jones, Clark, & Smith, 1995). These were followed by seminal fMRI studies indicating that moral reasoning recruits brain regions implicated in emotion processing (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Moll, De Oliveira-Souza, Bramati, & Grafman, 2002).

The aim of this paper is to consider the role of emotional learning in the development of morality. In particular, this paper will consider two forms of emotional learning: stimulus-reinforcement and response-outcome learning. It will be argued that these forms of learning are critical for the valence-based valuations on which much moral judgment (e.g., murder is bad and charitable giving is good) and decisions to commit (im)moral acts are based. But this is not to say that these two forms of emotional learning process give rise to the development of the system(s) that mediate all forms of moral reasoning. As previously noted

*Corresponding author.

(Nichols, 2002), emotion-based mechanisms can generate judgments of "badness" and undesirability. However, they cannot generate judgments of *immorality*. For example, an individual killing 10 people and a hurricane killing 10 people are both "bad" events but only the first is usually considered to be immoral. Nichols has proposed that judgments of immorality require the participant accessing semantic knowledge. The individual has to "recognize" the act as immoral based on their semantic concept of acts that are immoral. This may be the case. However, the learning/formation of conceptual structures will not be covered here.

Moreover, relatively recent theoretical work has stressed the importance of model-based reasoning/learning (Crockett, 2013; Cushman, 2013). As noted by Crockett (2013):

"The model-based system generates a forward-looking decision tree representing the contingencies between actions and outcomes, and the values of those outcomes. It evaluates actions by searching through the tree and determining which action sequences are likely to produce the best outcomes" (p. 363).

Importantly, the addition of a reference to a model-based system allows an explanation of various data regarding participants' reasoning on complex moral reasoning tasks such as the Trolley problem (in particular, side-effect vs. means variants of the Trolley problem; Crockett, 2013; Cushman, 2013). These data cannot be explained through reference to stimulus-reinforcement and response-outcome learning that will be considered here. Notably though, the valence information that the model-based system likely relies on is determined by prior stimulus-reinforcement and response-outcome learning. Indeed, Crockett (2013) refers to a "simple model-free system" that roughly corresponds to the response-outcome learning discussed here and a "Pavlovian system" that corresponds to the stimulus-reinforcement learning discussed here.

The paper will also stress data from clinical populations – particularly individuals with psychopathic traits. As such, the paper will first consider the disorder of psychopathy. Data from work with this population and fMRI work from healthy participants and also individuals with psychopathy will then be considered with respect to two forms of emotional learning, stimulus-reinforcement learning and response-outcome learning, with respect to moral judgments and behavior. Note though, given this clinical population, the literature considered here will almost exclusively involve consideration of judgments of transgressions and antisocial behavior. Discussion of prosocial behavior can be found elsewhere (Gesiarz & Crockett, 2015).

Psychopathy

The classification *psychopathy* characterizes an individual who shows an increased risk for antisocial behavior that is coupled with pronounced emotional deficits (P. J. Frick, 1995; Hare, 2003). It is this emotional component reflecting reduced guilt, remorse and empathy that is critical (R. J. R. Blair, 2007a). For children with the disorder, this emotional component is typically referred to as callous-unemotional (CU) traits (Frick, Stickle, Dandreaux, Farrell, & Kimonis, 2005). These CU traits are at the core of developmental

trajectory associated with psychopathy (Paul J. Frick & White, 2008). Youth with CU traits are at notably increased risk for meeting criteria for psychopathy as adults (Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007; L. C. Munoz & Frick, 2007).

It is important to note that the classification of psychopathy is not the same as the diagnosis of antisocial personality disorder (ASPD). ASPD and the corresponding diagnosis of conduct disorder (CD) for children are DSM diagnoses. They characterize individuals showing elevated levels of antisocial behavior and aggression. The core emotional deficits of psychopathy are not necessary to receive a diagnosis of ASPD or CD. Instead, it is only necessary to present with elevated levels of antisocial behavior.

Individuals with psychopathy are at significantly increased risk for the commission of acts of violence. In particular, criminals with psychopathy are at risk for committing instrumental (goal-directed) aggression relative to non-psychopathic criminals (though they also commit higher levels of reactive aggression; Williamson, Hare, & Wong, 1987). The groups show notably less difference for levels of (threat- or frustration-based) reactive aggression. This is especially true for the more serious offenses such as serious sexual assault or homicide. Notably, individuals with psychopathy are about twice as likely to have committed primarily instrumental homicides as non-psychopathic offenders (Woodworth & Porter, 2002). In fact, 93.3% of homicides committed by psychopathic offenders were instrumental in nature, compared with 48% of those by non-psychopathic offenders (Woodworth & Porter, 2002). In addition, individuals with psychopathy are significantly more likely to re-offend following release from prison than criminals without psychopathy (Olver & Wong, 2015).

In addition to their elevated risk of committing acts that can be considered immoral behaviors, individuals with psychopathy are compromised in their moral judgments. One of the earliest indices of moral development is the emergence of what has been termed the “moral/conventional distinction” (Smetana & Braeges, 1990; Turiel, Killen, & Helwig, 1987). This is the distinction between care-based moral transgressions (e.g., a child hitting another) and social disorder-based conventional transgressions (e.g., a child talking to another child in class). Typically developing children from the age of 3–4 years judge moral transgressions as less permissible, more serious and, critically, less rule contingent (i.e., moral transgressions remain non-permissible even in the absence of rules prohibiting them) than conventional transgressions (Smetana & Braeges, 1990; Turiel et al., 1987). Adults and youth with high psychopathic traits make significantly less of a moral/conventional distinction particularly with respect to the rule contingency judgments relative to even antisocial controls with low psychopathic traits (R. J. R. Blair, 1995, 1997). This is also seen for antisocial youth relative to comparison youth (Arsenio & Fleiss, 1996; Nucci & Herman, 1982; Smetana, 1990). Individuals with psychopathy/antisocial youth are also significantly less likely than comparison individuals to make reference to other individuals’ harm when justifying why care-based transgressions are wrong to commit (Arsenio & Fleiss, 1996; R. J. R. Blair, 1995, 1997). This does not mean that individuals with psychopathy make no distinction between moral and conventional transgressions. They judge moral transgressions as more serious than conventional transgressions (just to a lesser extent than comparison individuals) (R. J. R. Blair, 1995, 1997). Moreover, level of psychopathic traits has no predictive power for the ability to respond correctly when asked if an antisocial act causes

harm (Aharoni, Sinnott-Armstrong, & Kiehl, 2014; Aharoni, Sinnott-Armstrong, & Kiehl, 2012). Similarly, level of psychopathic traits has no predictive power for the ability to select four of eight (four moral and four conventional) transgressions that are "morally wrong" (Aharoni et al., 2014; Aharoni et al., 2012).

Using other paradigms, studies have reported abnormally utilitarian moral judgments in individuals with high levels of psychopathy personality traits (Gao & Tang, 2013; Glenn, Koleva, Iyer, Graham, & Ditto, 2010), including incarcerated individuals with high psychopathy levels (Koenigs, Kruepke, Zeier, & Newman, 2011) relative to comparison individuals. In addition, higher psychopathy scores are associated with reduced severity ratings of transgressions – at least in youth in forensic institutions (Harenski, Harenski, & Kiehl, 2014) though not in adult forensic samples (Harenski, Edwards, Harenski, & Kiehl, 2014; Harenski, Harenski, Shane, & Kiehl, 2010). Work in both subclinical and clinical populations has shown that individuals with psychopathy show reduced endorsement of care-based norms (Aharoni, Antonenko, & Kiehl, 2011; Glenn, Iyer, Graham, Koleva, & Haidt, 2009). They also show an increased likelihood to allow actions that indirectly harm another (Koenigs et al., 2011) and regard accidents that harm others as more permissible than comparison individuals (Young, Koenigs, Kruepke, & Newman, 2012). Victim salience is an important determinant of permissibility for healthy individuals that has significantly less impact on the permissibility judgments of individuals with higher psychopathic traits. Specifically, Marsh and colleagues in a series of studies have reported that individuals with higher psychopathic traits judge actions that cause fear in others as significantly more acceptable than individuals with lower psychopathic traits (Cardinale & Marsh, 2015; Marsh & Cardinale, 2013).

The empirical literature on moral judgment in individuals with psychopathy began from that claim that if emotional responses, specifically aversive emotional responses to the distress of others, were necessary for the development of morality then a population with reduced responsiveness to the distress of others, individuals with psychopathy, should show disrupted moral development (R. J. R. Blair, 1995). The above literature largely supports this position. The basic idea is that emotional learning processes engender an expected (aversive) value of the transgression (R. J. R. Blair, 1995, 2007b). It is this aversive expected value that contributes to the individual's rating of the badness of the transgression (R. J. R. Blair, 1995, 2007b). The idea is that emotional learning processes are disrupted in psychopathy and this results in an individual who lacks the emotional response underpinning the sense of badness of the transgression. The literature behind these claims will be explored below; specifically, the roles of two specific forms of emotional learning, stimulus-reinforcement and response-outcome in determining moral judgments and behavior.

Stimulus-reinforcement learning

Stimulus-reinforcement learning allows the individual to associate a valence with a stimulus; i.e., to learn whether something is good (positive) or bad (negative). The neural circuitry underlying fear and appetitive conditioning, two core forms of stimulus-reinforcement learning, is conserved across species (Johansen, Cain, Ostroff, & LeDoux, 2011). A core region is the amygdala. Within this region, information about the conditioned stimulus (CS)

and unconditioned stimulus (US) converge. Additional regions that are importantly involved include the hippocampus, ventromedial prefrontal cortex (vmPFC) and the dorsal anterior cingulate cortex (dACC).

It has been argued that appropriate stimulus-reinforcement learning is a core requirement for appropriate moral judgments (R. J. R. Blair, 2007b). The individual must form an association between a representation of the transgression and an aversive unconditioned stimulus. A failure in stimulus-reinforcement learning is thought to underpin the impairments in moral judgment shown by individuals with psychopathy described above (R. J. R. Blair, 2007b). In line with this position, a body of studies has indicated that individuals with elevated psychopathic traits show impaired fear conditioning. Studies have shown that compared to healthy comparison adults, adults with psychopathy fail to acquire conditioned responses to stimuli associated with shock or unpleasant odors (Flor, Birbaumer, Hermann, Ziegler, & Patrick, 2002; Lopez, Poy, Patrick, & Molto, 2013; Rothemund et al., 2012). Consistent with the suggestion of impaired amygdala functioning in individuals with psychopathy, they also show reduced amygdala responses to CS+ relative to comparison individuals (Birbaumer et al., 2005). Perhaps, most interestingly, several studies have reported that a relative failure in the capacity to show aversive conditioning is a risk for antisocial behavior in adulthood. For example, Raine and colleagues reported that a group of 127 individuals with a criminal record at age 23 showed poorer electrodermal fear conditioning data at age 3 than a gender, ethnicity, and social adversity noncriminal comparison group (N=274) (Gao, Raine, Venables, Dawson, & Mednick, 2010)

An important thing to note about stimulus-reinforcement learning is that the stimulus with which the reinforcement is associated is the stimulus that best predicts the emergence of the reinforcement (Rescorla & Wagner, 1972). It is this fact that probably explains the relative lack of efficacy for socialization of punishment child-rearing strategies (Gershoff et al., 2010). Indeed, punishment-based child-rearing strategies are associated with negative outcomes. In their cross-cultural study, Gershoff and colleagues showed that mothers' use of corporal punishment, expressing disappointment and yelling were significantly positively related to (likely threat-based, reactive) aggressive behavior. Giving a time out, using corporal punishment, expressing disappointment, and shaming were significantly positively related to greater child anxiety symptoms (Gershoff et al., 2010). Importantly, the best predictor of the emergence of the reinforcement (punishment) is not the transgression (which may have been conducted without anyone noticing and therefore gone unpunished) but the person delivering the punishment (and possibly an association with a representation of the punisher witnessing the transgression). The punisher may become a stimulus associated with fear rather than the transgression becoming associated with guilt (cf. relationship of child anxiety with use of punishment socialization strategies in Gershoff et al., 2010).

The victim's distress is a form of aversive reinforcement present, at least in early childhood, whenever an individual commits a transgression that harms another individual. A series of studies have documented that the social consequences of care-based (moral) transgressions are different from those of social order-based conventional transgressions (Nucci & Nucci, 1982a, 1982b). Care-based moral transgressions are associated with the victim's distress and are more likely to engender care-givers focusing attention on the victim's distress to the

perpetrator (Nucci & Nucci, 1982a, 1982b). These social consequences are very different from those incurred by conventional transgressions. Conventional transgressions engender references to social order and rules and are more likely to be associated with the care-giver's anger (Nucci & Nucci, 1982a, 1982b).

The importance of care-based moral transgressions being associated with the victim's displays of distress is: First that these displays are aversive and induce vicarious conditioning (Aniskiewicz, 1979; House & Milligan, 1976) and avoidance responses of objects associated with them (Meffert, Brislin, White, & Blair, 2015); and Second that the best predictor for the victim's distress is the action that caused it. The individual should rapidly learn the "badness" of care-based transgressions because of the pairing of the victim's distress with the commission (or observation of someone else committing) the care-based transgression (R. J. R. Blair, 1995, 2007b).

Valence-based learning on the basis of the emotional expressions of others is well established in the context of observational fear learning (Mineka & Zinbarg, 2006) or social referencing (Klinnert, Emde, Butterfield, & Campos, 1987) paradigms. In these paradigms, an infant observes the reaction of a caregiver to a novel object. If the caregiver reacts positively, the infant is significantly more likely to approach the object than if the caregiver shows a negative reaction/distress. Animal work has shown that social referencing is disrupted by amygdala damage (Jeon et al., 2010). Recent fMRI work has also implicated the amygdala in social referencing in humans (Meffert, Brislin, White, & Blair, 2014). In the context of social referencing, the amygdala is sensitive to expression prediction errors (i.e., the degree to which the expression induced by an object deviates from the expected emotional reaction). Learning occurs as a function of the prediction error such that greater learning (a greater change in the value associated with the object) occurs in response to greater prediction errors (Rescorla & Wagner, 1972). In addition, Meffert et al. (2014) demonstrated that the amygdala showed a differential response to objects associated with other individuals' happiness relative to objects associated with other individuals' fear.

Of course, this learning will only occur if the individual finds the distress of other individuals' aversive. An assumption that individuals with psychopathy were less responsive to the distress of others drove the original tests of moral judgment in individuals with psychopathy (R. J. R. Blair, 1995; R.J.R. Blair et al., 1995). This assumption was based on early psychophysiological data indicating reduced autonomic responses to the distress of others (specifically, the sight of other individuals' in pain) in individuals with psychopathy (Aniskiewicz, 1979; House & Milligan, 1976).

Following this, a series of studies have shown that youth with conduct problems, particularly those with psychopathic or CU traits, show deficits in expression recognition (e.g., R. J. R. Blair, Colledge, Murray, & Mitchell, 2001; Stevens, Charman, & Blair, 2001). Interestingly, this impairment in expression recognition is relatively selective. Recognition of fearful, sad and happy expressions is reduced while the recognition of disgusted and angry expressions remains intact (Dawel, O'Kearney, McKone, & Palermo, 2012; Marsh & Blair, 2007). Indeed, the impairment in the recognition of fearful expressions is seen even if the expression is presented too rapidly for eye gaze to have an influence on recognition accuracy

(Jusyte, Mayer, Kunzel, Hautzinger, & Schonenberg, 2014; Sylvers, Brennan, & Lilienfeld, 2011). As such the impairment cannot be considered to be a secondary consequence of impairment in attention to particular regions of the face (cf. Dadds, Jambrak, Pasalich, Hawes, & Brennan, 2011). The impaired recognition of fearfulness and sadness is pervasive, applying also to vocal tones (R. J. R. Blair, Budhani, Colledge, & Scott, 2005; Stevens et al., 2001) and body postures (L. Munoz, 2009).

Consistent with the finding of unimpaired anger and disgusted expression recognition, individuals with psychopathy show amygdala responses to angry expressions similar to those shown by comparison individuals (the response to disgust expressions has not been tested; Carre, Hyde, Neumann, Viding, & Hariri, in press; Marsh et al., 2008; White, Williams, et al., 2012). Again consistent with the findings of impaired sad and particularly fearful expression recognition, patients with CD, particularly those with elevated psychopathic or CU traits, show reduced amygdala responses to sad (Passamonti et al., 2010) and fearful relative to neutral (Jones, Laurens, Herba, Barker, & Viding, 2009; Lozier, Cardinale, VanMeter, & Marsh, 2014; Marsh et al., 2008; Viding et al., 2012; White, Marsh, et al., 2012) expressions. Individuals with psychopathy also show reduced amygdala (and associated regions such as anterior insula and rostral medial frontal cortex) responses to pain cues (Lockwood, Bird, Bridge, & Viding, 2013; Lockwood, Sebastian, et al., 2013; Marsh et al., 2013; Michalska, Zeffiro, & Decety, 2015). Notably, the reduced amygdala response is seen even if the expression is presented too rapidly for attention to the eye region to have an influence on BOLD response (Viding et al., 2012).

Interestingly, studies have reported that elevated CU traits are associated with an increased risk for instrumental aggression (e.g., Thornton, Frick, Crapanzano, & Terranova, 2013). A recent study has shown that this association is mediated by the amygdala response to distress cues (Lozier et al., 2014). In short, reduced responsiveness to the distress of others is associated with symptoms of reduced guilt, lower empathy and increased levels of instrumental aggression.

If the amygdala is important for stimulus-reinforcement learning and stimulus-reinforcement learning is important for moral judgment, it can be predicted that the amygdala should be importantly involved when participants are making moral judgments. In line with this, fMRI research has shown that the amygdala is one of a core set of regions (that also include orbitofrontal cortex, insula, anterior cingulate cortex, precuneus and posterior cingulate cortex) consistently activated during moral judgment tasks (see, for a meta-analytic review, Boccia et al., 2016). Notably, this role of the amygdala in moral judgment appears to be disrupted in individuals with psychopathic traits (Harenski et al., 2010; Marsh & Cardinale, 2014; Marsh et al., 2011; Yoder, Harenski, Kiehl, & Decety, 2015). Indeed, while healthy individuals show a positive relationship between ratings of severity of transgressions and amygdala responsiveness, this is significantly less in individuals with psychopathy (Harenski, Harenski, et al., 2014; Harenski et al., 2010).

In short, stimulus-reinforcement learning allows the individual to learn the valence of a stimulus. It is thought that the sense of badness of care-based moral transgressions is because representations of these transgressions have been associated with an aversive

unconditioned stimulus. It is argued here that this unconditioned stimulus is not pain induced by a caregiver's physical punishments. Such punishments induce fear of the punisher rather than moral socialization (cf. relationship of child anxiety with use of punishment socialization strategies in Gershoff et al., 2010). The aversive unconditioned stimulus is the distress of the other individual. Healthy individuals learn to avoid actions/objects associated with another individual's distress. Individuals with psychopathy show impairment in stimulus-reinforcement learning generally (at least aversive conditioning) and in their responsiveness to the distress of other individuals, functions reliant on the integrity of the amygdala. It is argued that these impairments disrupt their ability to learn the emotion-based sense of badness of care-based moral transgressions.

Response-outcome learning

There is a considerable animal and human literature on response-outcome learning and the use of response-outcome associations in reinforcement-based decision-making. An adequate review of this literature is beyond the scope of the current paper. However, useful recent reviews are available (O'Doherty, 2012; Rangel & Clithero, 2012; Schoenbaum, Takahashi, Liu, & McDannald, 2011). Regions of the brain critical for representing value in reinforcement-based decision-making include vmPFC, posterior cingulate cortex and striatum (Clithero & Rangel, 2014). Regions that appear to be particularly important in organizing avoidance responses on the basis of expected value information include anterior insula cortex and dorsomedial frontal cortex [dmFC] (Kuhnen & Knutson, 2005; White, Pope, et al., 2013). FMRI studies have implicated all of these regions (except striatum) in moral reasoning (see Boccia et al., 2016)

This paper concentrates on the regions most relevant to moral learning; i.e., the regions implicated in response-outcome learning. Both the striatum and vmPFC are thought to be critical for this type of learning (even if (ventral) striatum and vmPFC also play a role in stimulus-reinforcement learning; O'Doherty, 2012; Rangel & Clithero, 2012; Schoenbaum et al., 2011). An individual choosing to make a response relies on reinforcement expectancy information provided by these regions on the basis of prior learning. The individual will approach objects/initiate actions that are associated with positive reinforcement expectancies. The striatum is also critical for prediction error signaling (signaling the difference between the amount of reward or punishment received and the amount expected; Dayan & Balleine, 2002; O'Doherty, 2012). Prediction error signals are thought to spur reinforcement learning. The greater the prediction error, the greater the alteration in the reinforcement associated with the stimulus (Rescorla & Wagner, 1972).

Considerable data supports the suggestion of dysfunction in reinforcement-based decision-making in individuals with CD and ASPD. Patients with CD and ASPD show impairment on passive avoidance learning, reversal learning and risky choice tasks (Budhani & Blair, 2005; Budhani, Richell, & Blair, 2006; De Brito, Viding, Kumari, Blackwood, & Hodgins, 2013; Fairchild et al., 2009; Fisher & Blair, 1998; Newman & Kosson, 1986). These impairments are seen in both patients with CD with high and low levels of CU traits (White, Pope, et al., 2013); i.e., they are a risk factor for antisocial behavior rather than CU symptomatology. FMRI work suggests that these behavioral impairments relate to (i) reduced neural

responsiveness to reward within both striatum and vmPFC (Cohn et al., 2015; Crowley et al., 2010; Finger et al., 2011; Rubia et al., 2009; White, Pope, et al., 2013); (ii) atypical responsiveness to punishment information within striatum and vmPFC in youth with CD (Crowley et al., 2010; Finger et al., 2011; Finger et al., 2008b; White, Pope, et al., 2013); and (iii) reduced ability to organize avoidance responses on the basis of expected value information include anterior insula cortex, dmFC and striatum in avoidance behavior (White, Fowler, et al., 2014; White, Pope, et al., 2013). With respect to emotional learning, patients with CD may show particular problems in prediction error signaling (Finger et al., 2011; White, Pope, et al., 2013).

Impairment in these regions is thought to underpin the problems shown by individuals with psychopathy on many moral judgment tasks. As noted above, moral judgment tasks are associated with activity within all these regions except striatum (Boccia et al., 2016). It is suggested that this is because a core component of moral judgment is representing the expected value of the (im)moral action to be considered. As the salience of the victim increases (i.e., the expected value of the action diminished because of representation of the cost of the victim), the action is more likely to be judged as “bad”.

Impairment in these regions is thought to underpin the increased risk for antisocial behavior. Disrupted representation of expected value, particularly when coupled with a relative failure to learn value on the basis of distress cues of others, is likely to mean that the individual is less likely to avoid actions that harm others. In addition, more generally, impaired learning about the value of actions will increase the probability of making poorer decisions leading to frustration and frustration-based aggression.

An example of how dysfunction in the representation of expected value might result in increased “immoral” (antisocial) behavior is provided by recent work with a modified version of the Ultimatum game (White, Brislin, Meffert, Sinclair, & Blair, 2013; White, Brislin, Sinclair, & Blair, 2014; White et al., 2016). In this task, the participant plays against named opponents (in reality a computer). At the beginning of each trial, one of these opponents has $20 and has to share some of this with the participant. The sharing might be fair ($10 each) or progressively unfair (up to $18 for the opponent and only $2 for the participant). In the classic ultimatum game, the participant simply chooses whether to accept or reject the offer. However, in this modified version, the participant can accept the offer *or* spend between $1 and $3 to retaliate. For every retaliatory $1 spent, $7 is withdrawn from the opponent. Unsurprisingly, healthy individuals increase their retaliation the more unfair the offer (White, Brislin, et al., 2013; White, Brislin, et al., 2014). But this effect of provocation is significantly greater on individuals with CD; individuals with CD punish more than healthy youth (White et al., 2016). Interestingly, retaliatory activity on this task is positively associated with reported reactive aggression in youth with CD. More interesting still, the impairment in youth with CD relates to reduced representing of the costs of retaliation within vmPFC (and reduced integrated activity of this region with the amygdala) (White et al., 2016). Indeed, this neuro-computational impairment underpins the association between retaliation on task and reported reactive aggression.

In short, response-outcome learning allows the individual to represent the value of committing a particular action. Moral judgments (and of course (im)moral behaviors) are actions. As such they rely on the individual's ability to represent the value of the specific choices available. Regions implicated in response-outcome learning, and the representation of outcome, include vmPFC and striatum. Individuals with psychopathy show impairment in response-outcome learning. This impairment, together with the disruptions in learning the aversive value associated with actions that cause harm to others is thought to underpin their impairment on moral judgment tasks and increase their risk for instrumental antisocial behavior.

Caveats and Conclusions

The goal of this paper is to develop an argument regarding the role of emotion-based learning mechanisms in the development of moral judgment. However, before concluding it is worth considering a couple of caveats. First, as mentioned above, these emotion-based learning mechanisms can allow the individual to represent the "badness" of actions (i.e., their negative expected value based on their costs for the individual and particularly for others). However, these mechanisms *alone* would not allow the individual to represent the actions as *immoral* (cf. Nichols, 2002). As noted above, judgments of immorality require more than an aversive emotional response to the transgression. In this regard, it is worth very briefly considering the role of rostral medial frontal cortex. This region was implicated in moral judgment in the very earliest fMRI studies (Greene et al., 2001) and has continued to be implicated in more recent work (Boccia et al., 2016). This region is not typically implicated in emotional learning or the representation of expected value. However, it has been implicated in the underspecified process of self-referential processing (cf. Amodio & Frith, 2006). At least within the limitations of that underspecified process, one might consider that it is involved in accessing some of the self-relevant explicit norm information that is important for judgments of immorality (cf. Nichols, 2002). The learning of explicit norm information will involve forms of learning not covered by this paper.

Second, the current paper focused on stimulus-reinforcement learning on the basis of punishment, particularly if the "punishment" is the distress of a victim. However, there are other forms of stimulus-reinforcement learning, for example taste aversion learning that have implications for moral development. Taste aversion learning involves the individual associating the taste/sight of a certain food with symptoms caused by a toxic, spoiled, or poisonous substance (e.g., nausea). While the amygdala may play some role, the insula is particularly important for taste aversion learning; lesions of the insula block the acquisition and expression of taste aversion learning (Cubero, Thiele, & Bernstein, 1999). It is hypothesized that disgusted expressions may trigger taste aversion learning (R. J. R. Blair, 2007b). The individual will associate the aversive US of the expression with the stimulus that induced the expression. It has been argued that there are norms, often concerning matters of sexuality, that when transgressed are associated with disgust in those who hold the norm (Haidt, 2007). It appears plausible that the emotive force behind these norms may be acquired by experience of caregivers displaying disgust expressions when considering violations of these norms. This is interesting as while individuals with psychopathy show significantly reduced endorsement of care-based moral norms relative to comparison

individuals, they show comparable endorsement of disgust-based norms (Aharoni et al., 2011; Glenn et al., 2009). This would suggest that the form of emotional learning underpinning the acquisition of disgust-based norms is intact in individuals with psychopathy.

Currently, the status of the functional integrity of the insula in individuals with psychopathy is unclear. Individuals with psychopathy have shown appropriate recruitment of the anterior insula and inferior frontal cortices in the context of response control tasks (Finger et al., 2008a; Marsh et al., 2011). However, youth with conduct problems generally (i.e., not youth with conduct problems and psychopathy in particular) reveal atypical responses in this region when generating avoidance responses relative to controls (White, Pope, et al., 2013). Importantly, for the position above, individuals with psychopathy generally do not show impairment in the recognition of disgusted expressions (Dawel et al., 2012; Marsh & Blair, 2008). This might mean that emotional learning on the basis of these expressions is intact in individuals with psychopathy even when emotional learning on the basis of distress cues is not. This would explain the preservation of disgust-based norms in the population in the presence of significant impairment in processing care-based norms.

Third, an aim of this paper was not to conclude that non-victim-based punishment – physical or non-physical (e.g., time outs or separation from valued resources) – was irrelevant for socialization. Such techniques may be particularly associated with the socialization of conventional (social order based; e.g., talking in class, tidying toys away after use) transgressions even if they are also used in response to care-based transgressions (Nucci & Nucci, 1982a, 1982b). Indeed, relatively intact sensitivity to angry expressions in individuals with psychopathy (Carre et al., in press; Marsh et al., 2008; White, Williams, et al., 2012) may mean that these norms are encoded in individuals with psychopathy relatively similarly to conventional transgressions (anonymous reviewer's suggestion). This may give rise to findings that individuals with psychopathy treat care-based and conventional transgressions more similarly that comparison individuals (Arsenio & Fleiss, 1996; R. J. R. Blair, 1995, 1997; Nucci & Herman, 1982; Smetana, 1990).

Relatedly, and fourth, the paper did not consider fairness. This will be very briefly addressed here. Humans appear to show an expectation that resources will be equally distributed within the second year of life (Sloane, Baillargeon, & Premack, 2012) that at least partially reflects their experience with equitable sharing behaviors (Ziv & Sommerville, 2016). Humans show aversion to unequal resource allocations and may be will to sacrifice material payoffs to increase equality (Fehr & Schmidt, 1999). Individuals may show disadvantageous inequity aversion (DI), a desire to avoid receiving less than a peer, and advantageous inequity aversion (AI), a desire to avoid receiving more than a peer. DI emerges earlier in childhood and is seen in more cultures than AI (Blake et al., 2015). DI and AI choices are associated with comparable activations within dorsomedial frontal cortex, anterior insula cortex and dorsolateral prefrontal cortex (Guroglu, Will, & Crone, 2014). This has been argued to reflect a role of dorsomedial frontal cortex in responding to unexpected outcomes (including norm violations; cf. Alexander & Brown, 2011) and, anterior insula cortex in orchestrating potentially necessary changes in behavioral response (R. J. R. Blair & Cipolotti, 2000; Budhani, Marsh, Pine, & Blair, 2007); (see White, Brislin, et al., 2013). Interestingly, AI

choices were additionally associated with activity in the ventral striatum and ventromedial prefrontal cortex regions, as noted above, linked to the computation of subjective value and reward (Guroglu et al., 2014). This potentially relates to the reward value that can be received for helping others (cf. Moll et al., 2006).

Retaliation to DI situations (even at cost to the self) is seen from the age of 4 years (McAuliffe, Blake, & Warneken, 2014). This likely reflects a desire to preserve one's status relative to potential competitors (Crockett, Ozdemir, & Fehr, 2014; McAuliffe et al., 2014). Unfair resource allocations can be seen as disrespectful and, as such, may trigger retaliation (White, Brislin, et al., 2013). Interestingly, in healthy individuals magnitude of costly retaliatory behavior is inversely related to vmPFC activity, potentially reflecting representation of the costs of retaliation (White, Brislin, et al., 2013; White, Brislin, et al., 2014; White et al., 2016). As noted above, youth with CD (irrespective of level of psychopathy/callous-unemotional traits) show impaired representation of value within vmPFC (Crowley et al., 2010; Finger et al., 2011; Finger et al., 2008b; White, Pope, et al., 2013). In line with this, youth with CD show impaired representation of the costs of retaliation within vmPFC during retaliation to DI (White et al., 2016). Extent of this impaired representation is associated with their increased level of retaliation to DI (White et al., 2016). In short, according to the index of DI, individuals with CD (irrespective of level of psychopathy/callous-unemotional traits) show heightened levels of "fairness".

However, at least in this culture, "fairness" relates to both DI and AI (Blake et al., 2015). Given this it is notable that in norm endorsement paradigms, individuals with *psychopathy* show reduced endorsement of fairness/justice-based norms (Aharoni et al., 2011; Glenn et al., 2009). I would argue that this reflects reduced sensitivity to the social reinforcement of others' pleasure at receiving allocations and distress when these are removed. However this remains to be empirically determined.

In short, the current paper is not a complete view on the learning mechanisms necessary for moral judgment. It simply concentrates on the putative importance of stimulus-aversive conditioning on the basis of distress cue reinforcers and response-outcome learning. Impairments in these forms of emotional learning are thought to lead to the disruption in moral judgment particularly with respect to care-based transgressions and increased risk for antisocial behavior in individuals with psychopathy.

Acknowledgments

Dr. Blair is with the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health. This work was supported by the Intramural Research Program at the National Institute of Mental Health, National Institutes of Health under grant number 1-ZIA-MH002860-08 to Dr. Blair.

References

Aharoni E, Antonenko O, Kiehl KA. Disparities in the moral intuitions of criminal offenders: The role of psychopathy. J Res Pers. 2011; 45(3):322–327. DOI: 10.1016/j.jrp.2011.02.005 [PubMed: 21647247]

Aharoni E, Sinnott-Armstrong W, Kiehl K. What's wrong? Moral understanding in psychopathic offenders. Journal of Research on Personality. 2014; 53:175–181.

Aharoni E, Sinnott-Armstrong W, Kiehl KA. Can psychopathic offenders discern moral wrongs? A new look at the moral/conventional distinction. J Abnorm Psychol. 2012; 121(2):484–497. DOI: 10.1037/a0024796 [PubMed: 21842959]

Alexander WH, Brown JW. Medial prefrontal cortex as an action-outcome predictor. Nat Neurosci. 2011; 14(10):1338–1344. DOI: 10.1038/nn.2921 [PubMed: 21926982]

Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci. 2006; 7(4):268–277. DOI: 10.1038/nrn1884 [PubMed: 16552413]

Aniskiewicz AS. Autonomic components of vicarious conditioning and psychopathy. J Clin Psychol. 1979; 35:60–67. [PubMed: 422732]

Arsenio WF, Fleiss K. Typical and behaviourally disruptive children's understanding of the emotion consequences of socio-moral events. British Journal of Developmental Psychology. 1996; 14:173–186.

Birbaumer N, Veit R, Lotze M, Erb M, Hermann C, Grodd W, Flor H. Deficient Fear Conditioning in Psychopathy: A Functional Magnetic Resonance Imaging Study. Arch Gen Psychiatry. 2005; 62(7): 799–805. [PubMed: 15997022]

Blair RJR. A cognitive developmental approach to morality: Investigating the psychopath. Cognition. 1995; 57:1–29. [PubMed: 7587017]

Blair RJR. Moral reasoning in the child with psychopathic tendencies. Personality and Individual Differences. 1997; 22:731–739.

Blair RJR. The amygdala and ventromedial prefrontal cortex in morality and psychopathy. Trends in Cognitive Sciences. 2007a; 11(9):387–392. DOI: 10.1016/j.tics.2007.07.003 [PubMed: 17707682]

Blair RJR. The amygdala and ventromedial prefrontal cortex in morality and psychopathy. Trends Cogn Sci. 2007b; 11(9):387–392. [PubMed: 17707682]

Blair RJR, Budhani S, Colledge E, Scott S. Deafness to fear in boys with psychopathic tendencies. Journal of Child Psychology and Psychiatry. 2005; 46(3):327–336. [PubMed: 15755308]

Blair RJR, Cipolotti L. Impaired social response reversal: A case of "acquired sociopathy". Brain. 2000; 123:1122–1141. [PubMed: 10825352]

Blair RJR, Colledge E, Murray L, Mitchell DG. A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. J Abnorm Child Psychol. 2001; 29(6):491–498. [PubMed: 11761283]

Blair RJR, Jones L, Clark F, Smith M. Is the psychopath "morally insane"? Personality and Individual Differences. 1995; 19:741–752.

Blake PR, McAuliffe K, Corbit J, Callaghan TC, Barry O, Bowie A, Warneken F. The ontogeny of fairness in seven societies. Nature. 2015; 528(7581):258–261. DOI: 10.1038/nature15703 [PubMed: 26580018]

Boccia M, Dacquino C, Piccardi L, Cordellieri P, Guariglia C, Ferlazzo F, Giannini AM. Neural foundation of human moral reasoning: an ALE meta-analysis about the role of personal perspective. Brain Imaging Behav. 2016; doi: 10.1007/s11682-016-9505-x

Budhani S, Blair RJR. Response reversal and children with psychopathic tendencies: success is a function of salience of contingency change. J Child Psychol Psychiatry. 2005; 46(9):972–981. [PubMed: 16109000]

Budhani S, Marsh AA, Pine DS, Blair RJ. Neural correlates of response reversal: considering acquisition. Neuroimage. 2007; 34(4):1754–1765. DOI: 10.1016/j.neuroimage.2006.08.060 [PubMed: 17188518]

Budhani S, Richell RA, Blair RJ. Impaired reversal but intact acquisition: probabilistic response reversal deficits in adult individuals with psychopathy. J Abnorm Psychol. 2006; 115(3):552–558. [PubMed: 16866595]

Cardinale EM, Marsh AA. Impact of Psychopathy on Moral Judgments about Causing Fear and Physical Harm. PLoS One. 2015; 10(5):e0125708.doi: 10.1371/journal.pone.0125708 [PubMed: 25992566]

Carre JM, Hyde LW, Neumann CS, Viding E, Hariri AR. The neural signatures of distinct psychopathic traits. Social Neuroscience. in press.

Clithero JA, Rangel A. Informatic parcellation of the network involved in the computation of subjective value. Soc Cogn Affect Neurosci. 2014; 9(9):1289–1302. DOI: 10.1093/scan/nst106 [PubMed: 23887811]

Cohn MD, Veltman DJ, Pape LE, van Lith K, Vermeiren RR, van den Brink W, Popma A. Incentive Processing in Persistent Disruptive Behavior and Psychopathic Traits: A Functional Magnetic Resonance Imaging Study in Adolescents. Biol Psychiatry. 2015; 78(9):615–624. DOI: 10.1016/j.biopsych.2014.08.017 [PubMed: 25497690]

Colby, A., Kohlberg, L. The measurement of moral judgement. New York: Cambridge University Press; 1987.

Crockett MJ. Models of morality. Trends Cogn Sci. 2013; 17(8):363–366. DOI: 10.1016/j.tics.2013.06.005 [PubMed: 23845564]

Crockett MJ, Ozdemir Y, Fehr E. The value of vengeance and the demand for deterrence. J Exp Psychol Gen. 2014; 143(6):2279–2286. DOI: 10.1037/xge0000018 [PubMed: 25285429]

Crowley TJ, Dalwani MS, Mikulich-Gilbertson SK, Du YP, Lejuez CW, Raymond KM, Banich MT. Risky decisions and their consequences: neural processing by boys with Antisocial Substance Disorder. PLoS One. 2010; 5(9):e12835.doi: 10.1371/journal.pone.0012835 [PubMed: 20877644]

Cubero I, Thiele TE, Bernstein IL. Insular cortex lesions and taste aversion learning: effects of conditioning method and timing of lesion. Brain Res. 1999; 839(2):323–330. [PubMed: 10519056]

Cushman F. Action, outcome, and value: a dual-system framework for morality. Pers Soc Psychol Rev. 2013; 17(3):273–292. DOI: 10.1177/1088868313495594 [PubMed: 23861355]

Dadds MR, Jambrak J, Pasalich D, Hawes DJ, Brennan J. Impaired attention to the eyes of attachment figures and the developmental origins of psychopathy. J Child Psychol Psychiatry. 2011; 52(3):238–245. DOI: 10.1111/j.1469-7610.2010.02323.x [PubMed: 20883520]

Dawel A, O'Kearney R, McKone E, Palermo R. Not just fear and sadness: meta-analytic evidence of pervasive emotion recognition deficits for facial and vocal expressions in psychopathy. Neurosci Biobehav Rev. 2012; 36(10):2288–2304. DOI: 10.1016/j.neubiorev.2012.08.006 [PubMed: 22944264]

Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. Neuron. 2002; 36(2):285–298. [PubMed: 12383782]

De Brito SA, Viding E, Kumari V, Blackwood N, Hodgins S. Cool and hot executive function impairments in violent offenders with antisocial personality disorder with and without psychopathy. PLoS One. 2013; 8(6):e65566.doi: 10.1371/journal.pone.0065566 [PubMed: 23840340]

Fairchild G, van Goozen SH, Stollery SJ, Aitken MR, Savage J, Moore SC, Goodyer IM. Decision making and executive function in male adolescents with early-onset or adolescence-onset conduct disorder and control subjects. Biol Psychiatry. 2009; 66(2):162–168. DOI: 10.1016/j.biopsych.2009.02.024 [PubMed: 19362293]

Fehr E, Schmidt KMA. A theory of fairness, competition, and cooperation. Quarterly Journal of Economics. 1999; 114:817–868.

Finger EC, Marsh AA, Blair KS, Reid ME, Sims C, Ng P, Blair RJR. Disrupted reinforcement signaling in the orbital frontal cortex and caudate in youths with conduct disorder or oppositional defiant disorder and a high level of psychopathic traits. American Journal of Psychiatry. 2011; 168(2):834–841.

Finger EC, Marsh AA, Mitchell DG, Reid ME, Sims C, Budhani S, Blair JR. Abnormal ventromedial prefrontal cortex function in children with psychopathic traits during reversal learning. Arch Gen Psychiatry. 2008a; 65(5):586–594. DOI: 10.1001/archpsyc.65.5.586 [PubMed: 18458210]

Finger EC, Marsh AA, Mitchell DGV, Reid ME, Sims C, Budhani S, Blair RJR. Abnormal ventromedial prefrontal cortex function in children with psychopathic traits during reversal learning. Archives of General Psychiatry. 2008b; 65(5):586–594. [PubMed: 18458210]

Fisher L, Blair RJR. Cognitive impairment and its relationship to psychopathic tendencies in children with emotional and behavioural difficulties. Journal of Abnormal Child Psychology. 1998; 26:511–519. [PubMed: 9915656]

Flor H, Birbaumer N, Hermann C, Ziegler S, Patrick CJ. Aversive Pavlovian conditioning in psychopaths: Peripheral and central correlates. Psychophysiology. 2002; 39:505–518. [PubMed: 12212643]

Frick, Stickle, Dandreaux, Farrell, Kimonis. Callous-unemotional traits in predicting the severity and stability of conduct problems and delinquency. Journal of abnormal child psychology. 2005; 33(4): 471–487. [PubMed: 16118993]

Frick PJ. Callous-unemotional traits and conduct problems: a two-factor model of psychopathy in children. Issues in Criminological and Legal Psychology. 1995; 24:47–51.

Frick PJ, White SF. Research review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. Journal of Child Psychology and Psychiatry. 2008; 49(4):359–375. DOI: 10.1111/j.1469-7610.2007.01862.x [PubMed: 18221345]

Gao Y, Raine A, Venables PH, Dawson ME, Mednick SA. Association of poor childhood fear conditioning and adult crime. Am J Psychiatry. 2010; 167(1):56–60. DOI: 10.1176/appi.ajp.2009.09040499 [PubMed: 19917592]

Gao Y, Tang S. Psychopathic personality and utilitarian moral judgment in college students. J Crim Justice. 2013; 41:342–349.

Gershoff ET, Grogan-Kaylor A, Lansford JE, Chang L, Zelli A, Deater-Deckard K, Dodge KA. Parent discipline practices in an international sample: associations with child behaviors and moderation by perceived normativeness. Child Development. 2010; 81(2):487–502. [PubMed: 20438455]

Gesiarz F, Crockett MJ. Goal-directed, habitual and Pavlovian prosocial behavior. FrontBehav Neurosci. 2015; 9:135.doi: 10.3389/fnbeh.2015.00135

Glenn AL, Iyer R, Graham J, Koleva S, Haidt J. Are all types of morality compromised in psychopathy. Journal of Personality Disorders. 2009; 23:384–398. [PubMed: 19663658]

Glenn AL, Koleva S, Iyer R, Graham J, Ditto PH. Moral identity in psychopathy. Judgm Decis Mak. 2010; 5:497–505.

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. Science. 2001; 293:1971–1972. [PubMed: 11557854]

Guroglu B, Will GJ, Crone EA. Neural correlates of advantageous and disadvantageous inequity in sharing decisions. PLoS One. 2014; 9(9):e107996.doi: 10.1371/journal.pone.0107996 [PubMed: 25238541]

Haidt J. The new synthesis in moral psychology. Science. 2007; 316(5827):998–1002. [PubMed: 17510357]

Hare, RD. Hare Psychopathy Checklist-Revised (PCL-R; 2nd Ed). Toronto: Multi Health Systems; 2003.

Harenski CL, Edwards BG, Harenski KA, Kiehl KA. Neural correlates of moral and non-moral emotion in female psychopathy. Frontiers of Human Neuroscience. 2014; 25(8):741.

Harenski CL, Harenski KA, Kiehl KA. Neural processing of moral violations among incarcerated adolescents with psychopathic traits. Dev Cogn Neurosci. 2014; 10:181–189. DOI: 10.1016/j.dcn.2014.09.002 [PubMed: 25279855]

Harenski CL, Harenski KA, Shane MS, Kiehl KA. Aberrant neural processing of moral violations in criminal psychopaths. J Abnorm Psychol. 2010; 119(4):863–874. DOI: 10.1037/a0020979 [PubMed: 21090881]

House TH, Milligan WL. Autonomic responses to modeled distress in prison psychopaths. Journal of Personality and Social Psychology. 1976; 34:556–560. [PubMed: 993975]

Jeon D, Kim S, Chetana M, Jo D, Ruley HE, Lin SY, Shin HS. Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. Nat Neurosci. 2010; 13(4):482–488. DOI: 10.1038/nn.2504 [PubMed: 20190743]

Johansen JP, Cain CK, Ostroff LE, LeDoux JE. Molecular mechanisms of fear learning and memory. Cell. 2011; 147(3):509–524. DOI: 10.1016/j.cell.2011.10.009 [PubMed: 22036561]

Jones AP, Laurens KR, Herba CM, Barker GJ, Viding E. Amygdala hypoactivity to fearful faces in boys with conduct problems and callous-unemotional traits. American Journal of Psychiatry. 2009; 166:95–102. [PubMed: 18923070]

Jusyte A, Mayer SV, Kunzel E, Hautzinger M, Schonenberg M. Unemotional traits predict early processing deficit for fearful expressions in young violent offenders: an investigation using continuous flash suppression. Psychol Med. 2014; :1–13. DOI: 10.1017/s0033291714001287

Kagan, J., Lamb, S. The emergence of morality in young children. Chicago: University of Chicago Press; 1987.

Klinnert MD, Emde RN, Butterfield P, Campos JJ. Social referencing: The infant's use of emotional signals from a friendly adult with mother present. Annual Progress in Child Psychiatry and Child Development. 1987; 22:427–432.

Koenigs M, Kruepke M, Zeier J, Newman JP. Utilitarian moral judgment in psychopathy. Soc Cogn Affect Neurosci. 2011; doi: 10.1093/scan/nsr048

Kuhnen CM, Knutson B. The neural basis of financial risk-taking. Neuron. 2005; 47:763–770. [PubMed: 16129404]

Lockwood PL, Bird G, Bridge M, Viding E. Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. Front Hum Neurosci. 2013; 7:760.doi: 10.3389/fnhum.2013.00760 [PubMed: 24294197]

Lockwood PL, Sebastian CL, McCrory EJ, Hyde ZH, Gu X, De Brito SA, Viding E. Association of callous traits with reduced neural response to others' pain in children with conduct problems. Curr Biol. 2013; 23(10):901–905. DOI: 10.1016/j.cub.2013.04.018 [PubMed: 23643836]

Lopez R, Poy R, Patrick CJ, Molto J. Deficient fear conditioning and self-reported psychopathy: the role of fearless dominance. Psychophysiology. 2013; 50(2):210–218. DOI: 10.1111/j.1469-8986.2012.01493.x [PubMed: 23240559]

Lozier LM, Cardinale EM, VanMeter JW, Marsh AA. Mediation of the relationship between callous-unemotional traits and proactive aggression by amygdala response to fear among children with conduct problems. JAMA Psychiatry. 2014; 71(6):627–636. DOI: 10.1001/jamapsychiatry.2013.4540 [PubMed: 24671141]

Lynam DR, Caspi A, Moffitt TE, Loeber R, Stouthamer-Loeber M. Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. J Abnorm Psychol. 2007; 116(1):155–165. [PubMed: 17324026]

Marsh AA, Blair RJ. Deficits in facial affect recognition among antisocial populations: A meta-analysis. Neurosci Biobehav Rev. 2007

Marsh AA, Blair RJR. Deficits in facial affect recognition among antisocial populations: a meta-analysis. Neuroscience and Biobehavioral Reviews. 2008; 32(3):454–465. [PubMed: 17915324]

Marsh AA, Cardinale EM. When psychopathy impairs moral judgments: neural responses during judgments about causing fear. Social Cognitive and Affective Neuroscience. 2013; 9(1):3–11.

Marsh AA, Cardinale EM. When psychopathy impairs moral judgments: neural responses during judgments about causing fear. Soc Cogn Affect Neurosci. 2014; 9(1):3–11. DOI: 10.1093/scan/nss097 [PubMed: 22956667]

Marsh AA, Finger EC, Fowler KA, Adalio CJ, Jurkowitz IT, Schechter JC, Blair RJ. Empathic responsiveness in amygdala and anterior cingulate cortex in youths with psychopathic traits. J Child Psychol Psychiatry. 2013; 54(8):900–910. DOI: 10.1111/jcpp.12063 [PubMed: 23488588]

Marsh AA, Finger EC, Fowler KA, Jurkowitz IT, Schechter JC, Yu HH, Blair RJ. Reduced amygdala-orbitofrontal connectivity during moral judgments in youths with disruptive behavior disorders and psychopathic traits. Psychiatry Research. 2011; 194(3):279–286. DOI: 10.1016/j.pscychresns.2011.07.008 [PubMed: 22047730]

Marsh AA, Finger EC, Mitchell DGV, Reid ME, Sims C, Kosson DS, Blair RJR. Reduced amygdala response to fearful expressions in children and adolescents with callous-unemotional traits and disruptive behavior disorders. American Journal of Psychiatry. 2008; 165(6):712–720. [PubMed: 18281412]

McAuliffe K, Blake PR, Warneken F. Children reject inequity out of spite. Biol Lett. 2014; 10(12): 20140743.doi: 10.1098/rsbl.2014.0743 [PubMed: 25540156]

Meffert H, Brislin SJ, White SF, Blair JR. Prediction errors to emotional expressions: The roles of the amygdala in social referencing. Soc Cogn Affect Neurosci. 2014; doi: 10.1093/scan/nsu085

Meffert H, Brislin SJ, White SF, Blair JR. Prediction errors to emotional expressions: the roles of the amygdala in social referencing. Soc Cogn Affect Neurosci. 2015; 10(4):537–544. DOI: 10.1093/scan/nsu085 [PubMed: 24939872]

Michalska KJ, Zeffiro TA, Decety J. Brain response to viewing others being harmed in children with conduct disorder symptoms. J Child Psychol Psychiatry. 2015; doi: 10.1111/jcpp.12474

Mineka S, Zinbarg R. A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. Am Psychol. 2006; 61(1):10–26. [PubMed: 16435973]

Moll J, De Oliveira-Souza R, Bramati IE, Grafman J. Functional networks in emotional moral and nonmoral social judgments. Neuroimage. 2002; 16:696–703. [PubMed: 12169253]

Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J. Human fronto-mesolimbic networks guide decisions about charitable donation. Proceedings of the National Academy of Sciences. 2006; 103(42):15623–15628.

Munoz L. Callous-unemotional traits are related to combined deficits in recognizing afraid faces and body poses. Journal of American Academy of Child and Adolescent Psychaitry. 2009; 48(5):554–562.

Munoz LC, Frick PJ. The reliability, stability, and predictive utility of the self-report version of the Antisocial Process Screening Device. Scandinavian Journal of Psychology. 2007; 48:299–312. [PubMed: 17669220]

Newman JP, Kosson DS. Passive avoidance learning in psychopathic and nonpsychopathic offenders. Journal of Abnormal Psychology. 1986; 95:252–256. [PubMed: 3745647]

Nichols S. Norms with feeling: towards a psychological account of moral judgment. Cognition. 2002; 84(2):221–236. [PubMed: 12175573]

Nucci LP, Herman S. Behavioral disordered children's conceptions of moral, conventional, and personal issues. Journal of Abnormal Child Psychology. 1982; 10:411–425. [PubMed: 7175046]

Nucci LP, Nucci M. Children's responses to moral and conventional transgressions in free-play settings. Child Development. 1982a; 52:1337–1342.

Nucci LP, Nucci M. Children's social interactions in the context of moral and conventional transgressions. Child Development. 1982b; 53:403–412.

O'Doherty JP. Beyond simple reinforcement learning: the computational neurobiology of reward-learning and valuation. Eur J Neurosci. 2012; 35(7):987–990. DOI: 10.1111/j.1460-9568.2012.08074.x [PubMed: 22487029]

Olver ME, Wong SC. Short- and long-term recidivism prediction of the PCL-R and the effects of age: a 24-year follow-up. Personality Disorders. 2015; 6(1):97–105. [PubMed: 25364945]

Passamonti L, Fairchild G, Goodyer IM, Hurford G, Hagan CC, Rowe JB, Calder AJ. Neural abnormalities in early-onset and adolescence-onset conduct disorder. Arch Gen Psychiatry. 2010; 67(7):729738.doi: 10.1001/archgenpsychiatry.2010.75

Rangel A, Clithero JA. Value normalization in decision making: theory and evidence. Curr Opin Neurobiol. 2012; 22(6):970–981. DOI: 10.1016/j.conb.2012.07.011 [PubMed: 22939568]

Rescorla, RA., Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black, AH., Prokasy, WF., editors. Classical Conditioning II. Appleton: Century-Crofts; 1972. p. 64-99.

Rothemund Y, Ziegler S, Hermann C, Gruesser SM, Foell J, Patrick CJ, Flor H. Fear conditioning in psychopaths: event-related potentials and peripheral measures. Biol Psychol. 2012; 90(1):50–59. DOI: 10.1016/j.biopsycho.2012.02.011 [PubMed: 22387928]

Rubia K, Smith AB, Halari R, Matsukura F, Mohammad M, Taylor E, Brammer MJ. Disorder-specific dissociation of orbitofrontal dysfunction in boys with pure conduct disorder during reward and ventrolateral prefrontal dysfunction in boys with pure ADHD during sustained attention. American Journal of Psychiatry. 2009; 166:83–94. [PubMed: 18829871]

Schoenbaum G, Takahashi Y, Liu TL, McDannald MA. Does the orbitofrontal cortex signal value? Ann N YAcad Sci. 2011; 1239:87–99. DOI: 10.1111/j.1749-6632.2011.06210.x

Sloane S, Baillargeon R, Premack D. Do infants have a sense of fairness? Psychol Sci. 2012; 23(2):196–204. DOI: 10.1177/0956797611422072 [PubMed: 22258431]

Smetana, JG. Morality and conduct disorders. In: Lewis, M., Miller, SM., editors. Handbook of developmental psychopathology: Perspectives in developmental psychology. 157. New York: Plenum Press; 1990.

Smetana JG, Braeges JL. The development of toddlers' moral and conventional judgments. MPQ. 1990; 36:329–346.

Stevens D, Charman T, Blair RJR. Recognition of emotion in facial expressions and vocal tones in children with psychopathic tendencies. Journal of Genetic Psychology. 2001; 162(2):201–211. [PubMed: 11432605]

Sylvers PD, Brennan PA, Lilienfeld SO. Psychopathic traits and preattentive threat processing in children: a novel test of the fearlessness hypothesis. Psychol Sci. 2011; 22(10):1280–1287. DOI: 10.1177/0956797611420730 [PubMed: 21881061]

Thornton LC, Frick PJ, Crapanzano AM, Terranova AM. The incremental utility of callous-unemotional traits and conduct problems in predicting aggression and bullying in a community sample of boys and girls. Psychol Assess. 2013; 25(2):366–378. DOI: 10.1037/a0031153 [PubMed: 23244642]

Turiel, E., Killen, M., Helwig, CC. Morality: Its structure, functions, and vagaries. In: Kagan, J., Lamb, S., editors. The emergence of morality in young children. Chicago: University of Chicago Press; 1987. p. 155-245.

Viding E, Sebastian CL, Dadds MR, Lockwood PL, Cecil CA, De Brito SA, McCrory EJ. Amygdala response to preattentive masked fear in children with conduct problems: the role of callous-unemotional traits. Am J Psychiatry. 2012; 169(10):1109–1116. DOI: 10.1176/appi.ajp.2012.12020191 [PubMed: 23032389]

White SF, Brislin SJ, Meffert H, Sinclair S, Blair RJ. Callous-unemotional traits modulate the neural response associated with punishing another individual during social exchange: a preliminary investigation. J Pers Disord. 2013; 27(1):99–112. DOI: 10.1521/pedi.2013.27.1.99 [PubMed: 23342960]

White SF, Brislin SJ, Sinclair S, Blair JR. Punishing unfairness: rewarding or the organization of a reactively aggressive response? Hum Brain Mapp. 2014; 25(5):2137–2147. DOI: 10.1002/hbm.22316

White SF, Fowler KA, Sinclair S, Schechter JC, Majestic CM, Pine DS, Blair RJ. Disrupted expected value signaling in youth with disruptive behavior disorders to environmental reinforcers. J Am Acad Child AdolescPsychiatry. 2014; 52(5):579–588.e579. DOI: 10.1016/j.jaac.2013.12.023

White SF, Marsh AA, Fowler KA, Schechter JC, Adalio C, Pope K, Blair RJR. Reduced amygdala responding in youth with Disruptive Behavior Disorder and Psychopathic Traits reflects a reduced emotional response not increased top- down attention to non-emotional features. American Journal of Psychiatry. 2012; 169(7):750–758. [PubMed: 22456823]

White SF, Pope K, Sinclair S, Fowler KA, Brislin SJ, Williams WC, Blair RJ. Disrupted expected value and prediction error signaling in youths with disruptive behavior disorders during a passive avoidance task. Am J Psychiatry. 2013; 170(3):315–323. DOI: 10.1176/appi.ajp.2012.12060840 [PubMed: 23450288]

White SF, VanTieghem M, Brislin SJ, Sypher I, Sinclair S, Pine DS, Blair RJ. Neural Correlates of the Propensity for Retaliatory Behavior in Youths With Disruptive Behavior Disorders. Am J Psychiatry. 2016; 172(3):282–290. DOI: 10.1176/appi.ajp.2015.15020250

White SF, Williams WC, Brislin SJ, Sinclair S, Blair KS, Fowler KA, Blair RJ. Reduced activity within the dorsal endogenous orienting of attention network to fearful expressions in youth with disruptive behavior disorders and psychopathic traits. Dev Psychopathol. 2012; 24(3):1105–1116. DOI: 10.1017/s0954579412000569 [PubMed: 22781874]

Williamson S, Hare RD, Wong S. Violence: Criminal psychopaths and their victims. Canadian Journal of Behavioral Science. 1987; 19:454–462.

Woodworth M, Porter S. In cold blood: Characteristics of criminal homicides as a function of psychopathy. Journal of Abnormal Psychology. 2002; 111(3):436–445. [PubMed: 12150419]

Yoder KJ, Harenski C, Kiehl KA, Decety J. Neural networks underlying implicit and explicit moral evaluations in psychopathy. Transl Psychiatry. 2015; 5:e625.doi: 10.1038/tp.2015.117 [PubMed: 26305476]

Young L, Koenigs M, Kruepke M, Newman JP. Psychopathy increases perceived moral permissibility of accidents. J Abnorm Psychol. 2012; 121(3):659–667. DOI: 10.1037/a0027489 [PubMed: 22390288]

Ziv T, Sommerville JA. Developmental Differences in Infants' Fairness Expectations From 6 to 15 Months of Age. Child Dev. 2016; doi: 10.1111/cdev.12674