Cognition 167 (2017) 38–45

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Emotion-based learning systems and the development of morality

R.J.R. Blair

Section on Affective Cognitive Neuroscience, National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA

ARTICLE INFO

Article history:
Received 7 March 2016
Revised 7 March 2017
Accepted 24 March 2017
Available online 7 April 2017

Keywords:
Psychopathy
Amygdala
Moral judgment
Emotion-based learning

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.

1. 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., Blair, 1995; 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 (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

E-mail address: jamesblair@mail.nih.gov

http://dx.doi.org/10.1016/j.cognition.2017.03.013
0010-0277/

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).

2. Psychopathy

The classification *psychopathy* characterizes an individual who shows an increased risk for antisocial behavior that is coupled with pronounced emotional deficits (Frick, 1995; Hare, 2003). It is this emotional component reflecting reduced guilt, remorse and empathy that is critical (Blair, 2007). 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 (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; 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 (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; 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) (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, 2012, 2014). 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., 2012, 2014).

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, 2014b) though not in adult forensic samples (Harenski, Edwards, Harenski, & Kiehl, 2014a; 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 (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 (Blair, 1995, 2007). It is this aversive expected value that contributes to the individual's rating of the badness of the transgression (Blair, 1995, 2007). 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