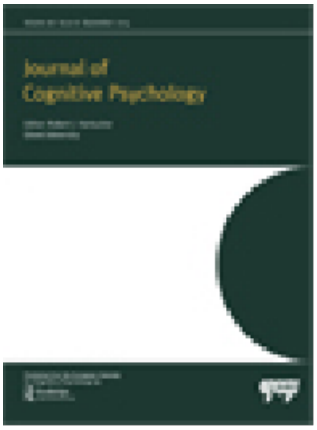
Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion



Journal of Cognitive Psychology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/pecp21

Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion

Indrajeet Patil[a]
[a] Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, SISSA-ISAS, Via Bonomea 265, 34136 Trieste, Italy
Published online: 03 Feb 2015.

To cite this article: Indrajeet Patil (2015): Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion, Journal of Cognitive Psychology, DOI: 10.1080/20445911.2015.1004334

To link to this article: http://dx.doi.org/10.1080/20445911.2015.1004334

Journal of Cognitive Psychology, 2015
http://dx.doi.org/10.1080/20445911.2015.1004334

Routledge
Taylor & Francis Group

Trait psychopathy and utilitarian moral judgement: The mediating role of action aversion

Indrajeet Patil

Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, SISSA-ISAS, Via Bonomea 265, 34136 Trieste, Italy

(*Received 5 May 2014; accepted 23 December 2014*)

Although past research has established that the utilitarian bias (increased willingness to agree to personally kill someone for the greater good) in psychopathy on moral dilemmas stems from weaker negative affect at the prospect of harming others due to reduced harm aversion, it remains to be seen if this is owing to reduced aversion to witnessing harmful outcomes (outcome aversion) or performing harmful actions (action aversion). In this study, we show that trait psychopathy is associated with both reduced outcome and action aversion and that only action aversion negatively mediates the influence of trait psychopathy on utilitarian moral judgement. Thus, the increased tendency in psychopathy to make utilitarian moral judgements is in part due to reduced aversion to carrying out harmful actions.

> Moments like this require someone who will act. Who will do the unpleasant thing, the necessary thing. (Frank Underwood, *House of Cards*, 2013)

Imagine the so-called footbridge dilemma (Thomson, 1985) where a trolley speeding down the track threatens to kill five people if nothing is done. You can save these people by pushing a large person standing next to you off of a footbridge to his death so that his weight would stop the trolley on collision. In such a situation, agreeing to sacrifice one to save many is said to be a utilitarian choice because utilitarianism (Mill, 1863/1998) entails that only consequences of moral actions matter and consequences which lead to maximisation of well-being for maximum number of agents ought to be preferred. On the other hand, the school of deontology evaluates actions based not just on consequences but also the actions themselves and forbids those actions that transgress some universal rights, duties, and obligations (Kant, 1785/2005). Thus, deontology forbids using individuals as a means to achieve the greater good. From a psychological point of view, the interesting question is what kind of personality traits would predispose an individual to accomplish this necessary evil for the greater good? We will explore this question by studying a personality trait that is well known to be associated with utilitarian bias, namely psychopathy.

Correspondence should be addressed to Indrajeet Patil, Neuroscience Sector, International School for Advanced Studies, SISSA-ISAS, Via Bonomea 265, 34136 Trieste, Italy. E-mail: ipatil@sissa.it

The author gratefully acknowledges Emanuela Liaci for providing Italian translation of the action/outcome aversion questionnaire. Thanks are also due to Paul Conway, Doris McIlwain, Giorgia Silani, and three anonymous reviewers for their invaluable comments on the earlier version of the manuscript. The author also wishes to thank Prof. Linden Ball for providing feedback on grammatical mistakes in the manuscript.

No potential conflict of interest was reported by the author.

1.1. Dual-process model of moral judgement

Recent research with moral dilemmas like the footbridge dilemma has led Joshua Greene and colleagues (Greene, 2007; Greene et al., 2009; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Shenhav & Greene, 2014) to formulate the dual-process model of moral judgements which posits two sets of computational processes that underlie moral decision-making: (1) emotional intuitions that arise in response to the aversive nature of personal harm that such dilemmas feature and subserve deontological responses; (2) conscious, deliberative processes that engage in cost–benefit analysis and support the utilitarian solution. This model reasons that because dilemmas like the footbridge dilemma feature emotionally aversive harm carried out in a personal manner (referred to as "personal moral dilemma"), most people would face a stronger cognitive conflict and decline to endorse utilitarian solutions on these dilemmas. This is exactly what is observed (Cushman, Young, & Hauser, 2006; Gleichgerrcht & Young, 2013; Greene et al., 2001, 2004, 2008, 2009; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Mikhail, 2007). Additionally, it also maintains that dilemmas with less emotionally salient harm carried out in an impersonal manner (called as impersonal moral dilemma) would lead to a relatively weaker cognitive conflict and most people would find utilitarian solution acceptable for such dilemmas. One example of such dilemma would be the Standard Fumes dilemma (Greene et al., 2004) where toxic fumes in a hospital threaten to kill five patients in one room and can be diverted, by pressing a switch, to another room where there is just one patient. Indeed, most people agree that the best course of action in such impersonal moral dilemmas is sacrificing one to save many (Cushman et al., 2006; Gleichgerrcht & Young, 2013; Greene et al., 2001, 2004, 2008, 2009; Hauser et al., 2007; Mikhail, 2007). Additionally, it also maintains that dilemmas with less emotionally salient harm carried out in an impersonal manner (called as impersonal moral dilemma) would lead to a relatively weaker cognitive conflict and most people would find utilitarian solution acceptable for such dilemmas. One example of such dilemma would be the Standard Fumes dilemma (Greene et al., 2004) where toxic fumes in a hospital threaten to kill five patients in one room and can be diverted, by pressing a switch, to another room where there is just one patient. Indeed, most people agree that the best course of action in such impersonal moral dilemmas is sacrificing one to save many (Cushman et al., 2006; Gleichgerrcht & Young, 2013; Greene et al., 2001, 2004, 2008, 2009; Hauser et al., 2007; Mikhail, or because she has reduced negative emotional reaction to the prospect of harming someone personally. Although previous research shows that both of these routes are taken when people make utilitarian moral judgements, in this study, we will exclusively focus on the path of reduced negative affect since it is more relevant for studying utilitarian bias in psychopathy. We will return to the enhanced deliberation path in the Section 4 when we explore alternative interpretations.

1.2. Reduced negative affect and utilitarian moral judgements

There is converging evidence to support the claim that when people contemplate personal moral dilemmas like the footbridge dilemma, they experience reflexive, prepotent negative emotional responses (for a review, see Miller & Cushman, 2013) stemming from a deep-seated human aversion to harming others (Haidt & Joseph, 2004). Neuroimaging studies show that brain regions involved in emotional processing are more active when people face personal moral dilemmas (Greene et al., 2001, 2004; Shenhav & Greene, 2014) and increased skin conductance activity (Moretto, Làdavas, Mattioli, & di Pellegrino, 2010; also see Patil, Cogoni, Zangrando, Chittaro, & Silani, 2014), which indexes arousal of the autonomic nervous system, also attests to this fact. Additionally, people are consciously aware of this emotional arousal when they contemplate personal moral dilemmas and report being more aroused on self-report measures (Szekely & Miu, 2014). Although this evidence corroborates the role of reduced negative affect in utilitarian moral judgements, a more nuanced analysis is in order because negative affect can derive from two different psychological mechanisms, namely outcome and action aversion which respectively stem from victim and agent perspective-taking (Cushman, Gray, Gaffey, & Mendes, 2012; Hannikainen, Miller, & Cushman, 2015; Miller, Hannikainen, & Cushman, 2014). Therefore, it needs to be determined which of these sources are responsible for reduced negative

2007). Note that the dual-process model posits two independent processes that contribute to the final moral judgement and they are not inversely proportional to each other (Conway & Gawronski, 2013). What this means is that someone can endorse utilitarian solution on personal moral dilemma either because she has predilection for enhanced cognitive deliberation (for a nuanced analysis, see Royzman, Landy, & Leeman, 2014) affect that underlies increased utilitarian tendencies. We will discuss the origin of and evidence for each of these sources in turn.

1.2.1. Outcome aversion

Moral condemnation of harmful behaviour can arise from the consideration of the harmful outcomes because harmful actions clearly target a victim. Empathising with the victim can lead to personal

distress in the observer, which motivates the observer not to carry out such actions and also to condemn such actions when carried out by others (Pizarro, 2000). In other words, when observers engage in mental simulation of situations featuring harmful actions, taking the perspective of the *victim* can lead them to be apprehensive of victim distress (and vicarious experience of that distress) and to their condemning the behaviour of those responsible for this negative outcome. Thus, according to the outcome aversion model, people refuse to endorse utilitarian moral judgements in personal moral dilemmas because they take into account the suffering and pain that such action would elicit in the proximal victim, e.g., death of the man that needs to be sacrificed in the footbridge dilemma. There is plenty of evidence to show that reduction in empathic response can increase the likelihood of utilitarian response for personal moral dilemmas.

Meta-analysis of brain imaging studies show that moral cognition recruits a relatively small subset of the brain areas (as compared to the theory of mind network) involved in empathy (Bzdok et al., 2012) and damage to these areas results in aberrant empathic skills and moral judgements. Those patient populations with developmental- and adult-onset ventromedial prefrontal cortex (vmPFC, a brain region essential for proper emotional processing) lesion and frontotemporal dementia (bvFTD, which also results in deterioration of prefrontal cortex) are well known for their emotional and empathic dysfunctions (e.g., see Mendez, 2009). Both of these populations show elevated levels of utilitarian moral judgements on personal moral dilemmas (Chiong et al., 2013; Ciaramelli, Muccioli, Làdavas, & di Pellegrino, 2007; Gleichgerrcht, Torralva, Roca, Pose, & Manes, 2011; Koenigs et al., 2007; Mendez, Anderson, & Shapira, 2005; Schroeter, Bzdok, Eickhoff, & Neumann, 2014; Taber-Thomas et al., 2014) as compared to brain-damaged and neurotypical control populations. This is arguably due to reduced prosocial concern for welfare of others and callous unemotionality, as shown by reduced autonomic arousal in vmPFC lesion patients (Moretto et al., concern (which gauges individual's propensity to experience feelings of warmth, compassion, and concern for other people) predict higher proportion of utilitarian moral judgements (Gleichgerrcht & Young, 2013; Jack, Robbins, Friedman, & Meyers, 2014; McIlwain et al., 2012; Miller et al., 2014; Patil & Silani, 2014). Making people emotionally more averse to perceived harmful acts by pharmacologically enhancing serotonin levels in the brain lessens the frequency of decisions that endorse utilitarian ends and, more interestingly, this effect is more prominent for people scoring higher on empathy (Crockett, Clark, Hauser, & Robbins, 2010).

Thus, one source of negative affect is the extent to which you find it personally upsetting to think about the suffering in victim who needs to be sacrificed in order to achieve the greater good. Reduction in this source of negative affect can predispose people to approve utilitarian moral judgements.

1.2.2. Action aversion

Another source of negative affect that drives moral condemnation of harmful behaviour stems from aversion to harmful actions themselves without further considering outcomes (Crockett, 2013; Cushman, 2013; Miller & Cushman, 2013; Miller et al., 2014). Blair and colleagues offer a developmental framework of morality to explain how we acquire aversion to performing harmful actions (Blair, 1995, 2007, 2013; Blair, White, Meffert, & Hwang, 2013). According to this model, humans (like other social animals) innately find distress cues (facial expressions, body posture, prosody denoting pain, sadness, fear, etc.) in conspecifics aversive because by virtue of empathy (a cognitive ability to correctly identify distress cues in others and experience an affective state which is not necessarily isomorphic, e.g., one can feel pain when perceiving pain in others or feel concern for them), this leads to aversive arousal in the observer. With aversive conditioning involving distress cues as "social punishment" (unconditioned stimulus), mental representations of actions

arousal in vmPFC lesion patients (Moretto et al., 2010) and reduced emotional empathy on self-report measures in bvFTD patients (Gleichgerrcht et al., 2011) when endorsing utilitarian moral judgement on personal moral dilemmas. People who score highly on trait emotional empathy also show a reduced tendency to endorse personal harms as morally acceptable and resort to deontological responses (Choe & Min, 2011), whereas self-reported low scores on dispositional empathic

that lead to such distress cues come to be associated with aversive reinforcement in the form of negative emotional arousal even in the absence of any distress cues. Consequently, the mere thought of performing harmful actions (conditioned stimulus) induces negative affect and leads to behavioural suppression. Thus, according to the action aversion model, a proper recognition of distress cues (e.g., correct identification of sad

and fearful facial expressions) and empathic response to these cues are necessary to developmentally acquire harm norms (Blair et al., 2013) but once these norms are acquired, the sensorimotor and perceptual properties of the actions or categorical descriptions of harmful actions (Miller et al., 2014) are sufficient to trigger a negative emotional response. It needs to be noted that actions are associated with aversive arousal only in certain contexts or states depending on the learned values from instrumental conditioning associated with those actions in the past, e.g., pushing a person is aversive but pushing a door is not (Crockett, 2013). It also needs to be mentioned that individuals need not engage in such harmful behaviours first-hand for stimulus-reinforcement learning to associate negative values to canonically harmful actions; these can also be learned via observation and simulation (Miller & Cushman, 2013).

Thus, on the action aversion model, when people think about personal moral dilemmas where they have to imagine carrying out harmful actions, they put themselves in the shoes of the *agent* and mentally simulate actions that have aversive reinforcement history associated with them which leads to increased negative arousal. This motivates them to condemn such actions and resort to non-utilitarian moral judgements. Indeed, people who show increased psychophysiological aversive reactivity to carrying out simulated harmful actions (e.g., thrashing a realistic-looking baby doll) find utilitarian moral actions to be less acceptable (Cushman et al., 2012). People with greater scores on trait harm avoidance show suppressed corticospinal excitability selectively when they simulate an agent's immoral actions (e.g., stealing a wallet) versus non-moral actions (e.g., picking up a notepaper; Liuzza, Candidi, Sforza, & Aglioti, 2014). Greene and colleagues (2009) also show that people find utilitarian trade-offs more aversive not because of spatial proximity or physical contact with the victim, but because the agent intentionally uses muscular force to push the

actions and are supposed to produce equally harmful "imagined" outcomes, one is found to be more aversive than the other due to sensorimotor properties of typical actions which have commonly been associated with aversive reinforcement in the past (Cushman & Dillon, 2015). An elevated level of action aversion on self-report measures also predicts a reduced tendency to endorse utilitarian moral judgements on both impersonal and personal moral dilemmas (Miller et al., 2014). Also, people who report themselves as more focused on actions than outcomes when deciding on moral appropriateness of harmful behaviour and who prefer taking perspective of the agent (over that of the victim) when judging third-party harmful behaviours condemn utilitarian course of action more (Hannikainen et al., 2015). Finally, people exhibit more autonomic arousal when they *act* to sacrifice one person for the greater good of saving five lives than when they *omit* to act and let the same outcome materialise (Navarrete, McDonald, Mott, & Asher, 2012). Taken together, these findings demonstrate that certain canonically violent actions are imbued with an aversive arousal independent of any consideration of harm.

Thus, action aversion arising from taking the perspective of the agent who needs to sacrifice one individual for the benefit of many is another source of negative affect. Diminution in the capacity to experience this negative affect (due to improper moral development) can prompt people to approve utilitarian moral judgements.

1.3. Psychopathy and utilitarian moral judgements

After discussing sources of negative affect that portray utilitarian options on personal moral dilemmas in a negative light, we now turn to a personality trait which is associated with increased utilitarian tendencies, namely psychopathy. Both incarcerated, clinical psychopaths (Koenigs, Kruepke, Zeier, & Newman, 2012; but see Rosas & Koenigs, 2014) and

victim. This underscores the importance of simulation of agent's motor behaviour and goals (action aversion) over victim suffering (outcome aversion) when condemning harmful actions. People show increased arousal when performing pretend harmful actions in a "typical" way (e.g., pulling the trigger of the gun with a finger) as compared to when the same actions are performed in an "atypical" manner (e.g., pulling the trigger of the gun using a string). Although both are harmful

non-incarcerated, subclinical individuals with psychopathic tendencies show a preference for utilitarian solutions on emotionally aversive moral dilemmas (Arvan, 2013; Bartels & Pizarro, 2011; Djeriouat & Trémolière, 2014; Gao & Tang, 2013; Glenn, Koleva, Iyer, Graham, & Ditto, 2010; Kahane, Everett, Earp, Farias, & Savulescu, 2015; Langdon & Delmas, 2012; McIlwain et al., 2012; Seara-Cardoso, Dolberg, Neumann, Roiser, & Viding, 2013; Tassy, Deruelle, Mancini, Leistedt, &