Cognition 121 (2011) 101–114

Contents lists available at ScienceDirect

Cognition

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

When sentimental rules collide: "Norms with feelings" in the dilemmatic context

Edward B. Royzman [a,*], Geoffrey P. Goodwin [a], Robert F. Leeman [b]

[a] Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104, United States
[b] Department of Psychiatry, Yale University School of Medicine, CMHC, Room S200, New Haven, CT 06519, United States

ARTICLE INFO

Article history:
Received 5 January 2011
Revised 8 June 2011
Accepted 13 June 2011
Available online 12 July 2011

Keywords:
Sentimentalism
Social domain theory
Moral
Dilemmas
Moral development
Moral reasoning
Emotions

ABSTRACT

According to a recently prominent account of moral judgment, genuine moral disapprobation is a product of two convergent vectors of normative influence: a strong negative affect that arises from the mere consideration of a given piece of human conduct and a (socially acquired) belief that this conduct is wrong (Nichols, 2002). The existing evidence in favor of this "norms with feelings" proposal is rather mixed, with no obvious route to an empirical resolution. To help shed further light on the situation, we test a previously unexamined prediction that this account logically yields in a novel dilemmatic context: when individuals are faced with a moral dilemma that pits two or more "affectively-charged" moral norms against each other, the norm underwritten by the strongest feeling ought to determine the content of dilemmatic resolution. Across three studies, we find evidence that directly challenges this prediction, offering support for a Kolhberg-style "rationalist" alternative instead. More specifically, we find that it is not the participants' degree of norm-congruent emotion (whether situationally or dispositionally assessed) or its correlates, but rather their appraisal of the relative costs associated with various alternative courses of action that appears to be most predictive of how they resolve the experimentally induced moral conflict. We conclude by situating our studies within an overarching typology of moral encounters, which, we believe, can help guide future research as well as shed light on some current controversies within this literature.

> Textbook psychology preaches the cliché that moral decisions are a product of the algebraic resolution of conflicting quantitative affective forces. Though efforts to predict moral decisions by this model have yielded slim results, the metaphor continues to have currency. We are claiming instead that the moral force in personality is cognitive. (Kohlberg, 1971, p. 230).

1. Introduction

Sentimentalism, the idea that emotions or sentiments are crucially implicated in the etiology of moral judgment, has dominated moral psychological research for close to a decade. One particularly sophisticated variant of this approach was recently set forth by Shaun Nichols (Nichols, 2002, 2004, 2008). Following in the trail of the highly influential social domain theory (Nucci, 2001; Smetana, 1983; Turiel, 2002), Nichols' proposal postulates two qualitatively distinct ways in which an act (or a set of acts) may be deemed impermissible or wrong. On the one hand, there are acts that are judged to be merely *conventionally* wrong and that represent "a violation of an implicit uniformity or an explicit regulation within the social system (e.g., the school)" (Turiel, 1983, p. 44) (see also Smetana, 1983), e.g., addressing a teacher by her first name. On the other hand, acts judged to be *morally* wrong (such as malicious lying, stealing, or physical harm) likewise represent a violation of an implicit or explicit code of conduct within a

* Corresponding author. Address: Department of Psychology, University of Pennsylvania, 3720 Walnut St., Solomon Lab Building, Philadelphia, PA 19104, USA.
E-mail address: royzman@psych.upenn.edu (E.B. Royzman).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.06.006

given social system, but in addition, carry intrinsic negative consequences for others, making them *worse off*. According to the social domain theory, our condemnation of moral and conventional wrongs is said to differ in content as well as in form.

With regards to judgment form, the significant and interesting distinction is that the wrongness of prototypically immoral acts (stealing, physically hurting others) is seen as *socially transcendent* or largely independent of existing social standards or norms. Thus, a prototypically immoral act will generally be judged to be wrong even when individuals are instructed to envisage that it is no longer "against the rules", that it has been allowed by a recognized authority, or that it takes place within a cultural milieu where its performance is normatively sanctioned. Moreover, all else being equal, moral transgressions are generally seen as more serious or severely counter-normative than their conventional counterparts.

According to Nichols (2002), these differences in judgment form have a common psychological source. At its most essential, Nichols's account holds that the basis of a genuinely moral judgment of wrong is to be sought in a certain co-mingling of strong (negative) affect and some kind of socially transmitted prohibition. Thus, a certain category of *rules* (e.g., "Hitting is wrong") and, by extension, *case-specific judgments* ("It was wrong for Paul to hit Bill") originating from these rules, will take on a genuinely moral status insofar as the behavior they proscribe is a source of strong negative affect (e.g., primordial sympathy caused by the victim's distress) independent of the rule itself. "Thou shall not kill" would be a prime example of a "sentimental rule" (underwritten, presumably, by our basic sympathy for the victim and the bereft). On Nichols's view, however, sentimental rules are by no means restricted to norms regulating commission (or passive acceptance of) interpersonal harm, but also encompass "norms prohibiting disgusting behavior" (Nichols, 2004, p. 29), such as disgust-underwritten violations of dining etiquette or acts of sexual impropriety.

Alternative accounts of the origins of the moral–conventional distinction also exist, however. One classic alternative (Nucci, 2001; Turiel, 1983, Turiel, 2002) is that it is harm considerations, *broadly construed* (see Royzman, Leeman, & Baron, 2009, p. 166), that give judgments against transgressions as diverse as murder, rape, stealing, schoolyard teasing, tax evasion, disrespecting one's elders, and preventing one's elders' souls from reaching salvation their life as moral entities (Turiel, Killen, & Helwig, 1987; Turiel & Wainryb, 1994). In fact, Turiel suggests a specific "test" that a child may employ to establish whether a given normative breach is a matter of convention or morality proper. By mentally undoing the act (while taking into account the reason for the offender's conduct) the child asks whether the interpersonal consequences are *worse with the act or without it*. As a consequence of these steps, the child will come to represent acts that are inherently *detrimental* to others as morally wrong (Turiel, 1983, pp.42–44).

Consistent with this proposal, which we will call the "reason-based" alternative, there is a wide range of cross-cultural (e.g., Hollos, Leis, & Turiel, 1986; Song, Smetana, & Kim, 1987) and cross-generational (e.g., Smetana & Braeges, 1990; Smetana, Campione-Barr, & Yell, 2003) evidence to suggest that people of various ages and in various places regard transgressions intrinsically harmful to others as having a special moral status and weight that is largely non-existent for transgressions deemed intrinsically harm-free (Nucci, 2001; Turiel, 1983, 2002 for review).[1] Interestingly, the emphasis on assessment of relative harm as a guide to moral judgment has its psychological origins at least with Lawrence Kohlberg, who explored this idea largely in the dilemmatic context in which two duties (values) and their associated costs were pitted against each other via a series of hypothetical vignettes. For Kohlberg, the orientation towards "utilitarian justice" was the definitive principle of post-conventional thinking at Stage 5, the ultimate, realistically achievable stage of moral development for young adults (Colby & Kohlberg, 1987). The judicious balancing of competing utilities (and disutilities) was part and parcel of what he called the "Prior-to-society perspective" ("Perspective of a rational individual aware of values and rights prior to social attachments and contracts") and expressed itself in a "concern that laws and duties be based on rational calculations of overall utility, 'the greatest good for the greatest number" (Kohlberg, 1976, p.34).

In sum, these two factors, *abstract assessment of relative costs* on the one hand, and the presence of *prepotent affective reactions* on the other, are distinguishable both in principle and in practice. Yet, the currently available data remain indecisive as to which of them best accounts for individuals' moral judgment capacities, including the

[1] Some authors have challenged this thesis (e.g., Haidt, Koller, & Dias, 1993; Shweder, Mahapatra, & Miller, 1987; see also Westermark, 1906). For example, Shweder and colleague's research (1987) among the Brahmins of Orissa (India) revealed that violations pertaining to diet (a son eating chicken shortly after his father's death) could be judged as socially transcendent/genuinely immoral as an act of interpersonal harm. These findings have been used to argue that morality is hypocognized in the West, especially among the so-called "social liberals" (Graham, Haidt, & Nosek, 2009). While Shweder et al.'s claims are intriguing, it has been pointed out that they largely fail to take into account substantial differences in the factual (cosmological) assumptions made by his Indian and American subjects. Taking these differences into account (e.g., a belief that a man who eats chicken following his father's death prevents his father's soul from reaching salvation, a belief that a woman wearing bright colors shortly following her husband's death devalues his and her family's reputation) makes it reasonably likely that these purportedly harm-free acts are not perceived as being harm-free after all (Turiel & Wainryb, 1994; Turiel et al., 1987). This re-interpretation has been supported by the field work of Madden (1992), who found that those Hindu priests who were willing to entertain harm-negating counterfactual beliefs (e.g., the idea that deeds of the living do not in fact affect the souls of the deceased) tended to judge the target behaviors as no longer morally inappropriate (see also Turiel & Wainryb, 1994; Wainryb, 1991; Wainryb & Turiel, 1993 for similar demonstrations). Similarly, it appears that a number of culture war – defining issues (Graham et al., 2009), such as abortion and gay marriage, may ultimately be traced to differing existential assumptions about the inception of human life (Smetana, 1981) and differing beliefs about various long-term effects of sex same unions on society in general and mental/sexual health of young children in particular (see Corley, 2009). Consistent with this thesis, Royzman et al. (2009) found that harm considerations, broadly construed, not negative affect or its correlates, were the best predictors of subjects' tendency to moralize two prototypically "harm-free" infractions (spitting at a dinner table, sibling incest) within the domains of food and sex, respectively.