



ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: [www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)

## When sentimental rules collide: “Norms with feelings” in the dilemmatic context

Edward B. Royzman<sup>a,\*</sup>, Geoffrey P. Goodwin<sup>a</sup>, Robert F. Leeman<sup>b</sup>

<sup>a</sup> Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104, United States

<sup>b</sup> 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 xxxx

#### 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 Kohlberg-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.

© 2011 Elsevier B.V. All rights reserved.

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,

\* 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](mailto:royzman@psych.upenn.edu) (E.B. Royzman).

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 uniform 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

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).<sup>1</sup> 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

<sup>1</sup> Some authors have challenged this thesis (e.g., Haidt, Koller, & Dias, 1993; Shweder, Mahapatra, & Miller, 1987; see also Westermarck, 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.

capacity crucially implicated in moral–conventional distinction. The main point of the present paper is to offer a new angle on this issue. We next turn to reviewing the existing evidence relevant to each of these general proposals before outlining the present empirical strategy which aims to adjudicate between the two.

### 1.1. Existing evidence

In addition to being theoretically attractive (see Royzman, Baron, & Goodwin, in preparation; Royzman et al., 2009 for discussion), Nichols's sentimentalist proposal enjoys some apparent evidentiary support, including the results emanating from Nichols's own elegantly simple study (Nichols, 2002, Study 2), in which high-disgust individuals appeared significantly more likely than low-disgust individuals to assign a moral-like (authority-independent) status to a disgust-tinged breach of etiquette (a person spitting into a napkin whilst at a dinner party). Along similar lines, Blair (1995) found, that, compared to IQ-matched criminal controls, incarcerated psychopaths, also known to be relatively unresponsive to signs of human distress (Blair, Jones, Clark, & Smith, 1997), were substantially less adept at making the moral–conventional distinction, and virtually never appealed to harm considerations as part of their account of what made prototypically moral transgressions wrong.

On the other hand, there are a number of recent (and not so recent) findings that make Nichols's creative and thoughtful account of norm moralization rather difficult to accept. Blair (1999), for example, also found that children with psychopathic tendencies, though severely impaired in their emotional responses to the distress of others (much more so than adult criminal psychopaths (Blair, 1995)), still differentiated moral and conventional events on all the classic dimensions, including permissibility, seriousness, and authority jurisdiction. They also appealed to *harm considerations* in justifying their responses at a level that was statistically indistinguishable from that of non-psychopathic controls. More recently, Dolan and Fullam (2010) reported a similar (and even stronger) pattern of results based on a sample of juvenile offenders, finding *no* significant relationship between affective components of psychopathy and the ability to make the moral–conventional distinction. And, as noted earlier, in our own previous look at Nichols's theory (Royzman et al., 2009), we found that harm considerations, broadly construed, not affect or its proxies, were the significant predictors of our student subjects' tendency to moralize some (seemingly harmless) violations of dining etiquette and sexual morality.

In sum, when it comes to assessing how it is that individuals make uniquely moral judgments of wrong, we appear to be left with something of a standoff. Using the classic moral–conventional distinction task has generated a body of evidence pointing in two opposite directions, offering, it seems, little hope for constructive synthesis or reconciliation.

One useful exercise in the face of such evidentiary impasse is to examine the competing theories in a novel context, one in which two or more (putative) sentimental rules

are in competition for the determination of an individual's moral response. This is a context that Nichols's account does not directly address, but about which relevant predictions can be logically derived from the account itself combined with the body of evidence and prior theorizing on which it rests. Should the predictions pan out, this will significantly increase our confidence in the theory and give it a new lease on life (while redounding to the merit of sentimentalist tradition as a whole). Should the predictions not be confirmed, it will help us further delineate the boundary conditions for the theory's application. Conveniently, the dilemmatic context also provides a clear way to test Nichols's theory against the reason-based alternative outlined above, which dictates that an individual's dilemmatic moral response should be determined not as “a product of the algebraic resolution of conflicting . . . affective forces” (Kohlberg, 1971, p. 230), but rather by the perceived balance of costs and benefits (i.e., the ‘utilities’) arising from the various alternative courses of action under consideration.

## 2. Overview of the studies: norms with feelings in the dilemmatic context

According to Nichols, rules backed or underwritten by strong negative emotion should be encoded as far more serious and more genuinely moral than rules without such a backing (yielding more serious and moral-like case-specific judgments downstream). What would this proposal mean in a dilemmatic context? All things being equal, in a contest between two sentimental rules, the more serious and morally weighty one – which, on Nichols's view, is also the one that enjoys greater affective backing – should triumph over its less serious or more conventional rival, thus becoming the ultimate arbiter of judgment in a variety of morally charged situations, including those made famous by Kohlberg's classic developmental model (Colby & Kohlberg, 1987).

Kohlberg's method, it will be recalled, relies on a series of thought-provoking moral dilemmas defined by conflicts of duties. In his best-known exemplar, one is asked to judge whether it is morally appropriate for a man (Heinz) to burglarize an upstanding member of his community (in possession of an overpriced, cancer-beating drug) if this is the only way to prevent the untimely death of his spouse. In another (less well-known) vignette, the protagonist has to determine whether or not to “tell on” a sibling who recently misappropriated some family funds, knowing that the sibling is likely to be harshly punished if found out. Both vignettes (and we think this point has not been stressed nearly enough by Kohlberg's detractors and adherents alike) are rife with emotion. One problem, however, with using Kohlberg's own vignettes to test the sentimentalist theory's predictions is that the relevant competing emotions are far too homogenous a set. In the Heinz dilemma, for instance, both courses of action (concerned with values of life and property, respectively) may be construed as motivated by sympathy or some emotional concern for others. The alternative would be to consider a situation in which the actions targeted by the conflicting

sentimental rules are (a) rich sources of *pre-normative affect*, i.e., “made to evoke (strong) negative feelings independent of the normative/evaluative response” (Royzman et al., 2009, p. 162) and lead to emotional reactions that are (b) clearly distinct and (c) lend themselves to effective measurement at the trait level.

Disgust and sympathy fit the bill on all counts. They are clearly phenomenologically distinct. Both are identified by Nichols as being independent of norm theory, and virtually all of his arguments hinge on these two. Also, reliable and temporally stable measures exist for both: for disgust – a revised trait disgust measure (DS-R) by Olatunji et al. (2007), a psychometrically refined version of Haidt, McCauley, and Rozin’s (1994) Disgust sensitivity scale, used by Nichols (2002) and ourselves (Royzman et al., 2009) and, for sympathy – the Empathic concern scale (EC), a sub-component of Davis’s Interpersonal Reactivity Index (Davis, 1994).<sup>2</sup>

Our aim was to investigate a scenario in which a sympathy-based rule having to do with the special moral obligation to one’s kin (such as in the “Heinz” and other Kohlberg-style vignettes) collides with a disgust-tinged rule of comparable authority. Thus we required a disgusting rule that is sufficiently universal (and entrenched) and that has a proven record of being treated as genuinely moral by a substantial number of (but not all) participants such as ours (college students). Prohibition on sibling incest seemed to be a prime candidate for the part (see Royzman, Leeman, & Sabini, 2008; Royzman et al., 2009 for further arguments on this point).

Consider, for example, the following pair of vignettes:

#### The hike

Meg has a very rare, and unusual, medical condition. Because of a hormonal imbalance, Meg periodically suffers from a fit, which, if allowed to run its course, has a 99% chance of resulting in a stroke, causing her to become paralyzed from the waist down. The fit can be quickly countered by a very strong dose of oxytocin, a hormone that spikes when nursing or just after having penetrative sex. One day, Meg and her twin brother Dave are hiking in the mountains, and she has one of her fits. To her chagrin, she discovers that her vial of oxytocin has cracked, and is unusable. The only chance she has to produce a sufficiently high spike of oxytocin quickly is to have a full vaginal intercourse with her brother Dave. So Meg (who is on the pill) asks Dave to have sex with her then and there. Dave fully grasps the gravity of the situation, but feels that he simply cannot be a part of what he views as a deeply immoral and even heinous act. . . As a result, Meg suffers a stroke and becomes paralyzed from the waist down.

#### Gregor’s choice

Gregor (17) and Katriana (17) are twin siblings and the only surviving members of a recently deposed ruling-class family in the democratic republic of Ylandia, a country in the throes of an all-out civil war. They live

in a province currently under the thumb of the ruthless UPAY (United Patriots for better Ylandia), a religious fundamentalist organization whose draconian prosecution of the members of the former regime seems to know no bounds. Gregor and Katriana’s parents and grandparents have already been executed or disappeared, while Gregor and Katriana remain under “protective custody” in their roomy but desolate suburban mansion. Luckily for Gregor, according to UPAY’s own rules of engagement, 17 is too young to be harmed for a Yalandian male. . . Katriana is not so lucky. Because 17 is “the age of reason” for a Yalandian female, she can and will be executed in short order. The date of the execution has been set for three weeks from now. Escape is impossible. Katriana’s only chance to keep alive long enough to benefit from the impending international rescue mission is to exploit a loophole in UPAY’s fundamentalist code: absolutely no pregnant woman may be killed, mutilated or otherwise harmed prior to giving birth. Unfortunately, the only male available to impregnate Katriana is her twin brother Gregor. Knowing this, Katriana asks Gregor to have sex with her. . . Gregor understands what is at stake and wants to help, but simply cannot see himself as being a party to what he regards as a deeply immoral and even heinous act; so he says “no”. . . Katriana is executed three weeks later.

Imagine now someone dispositionally lacking in sympathetic concern (a psychopath). Assuming such a person has normal levels of disgust, the rule relevant to the incest avoidance option (the incest taboo) would be expected to have been encoded in this person’s mind as far more serious and more socially transcendent (“genuinely moral”) than its putatively sympathy-tinged counterpart concerned with the duty to assist one’s kin. Thus, according to the sentimentalist account, the person in question should be more likely to prescribe sacrificing a sibling so as to avert giving into the incestuous demands.

The opposite prediction would be made for someone who is “all sympathy, no disgust” (or could be assessed to be such, using the appropriate trait measures). Obviously most people fall at neither extreme, but they would be expected to vary in their level of disgust dominance (computed, say, as the difference between their trait disgust and trait sympathy). The variance on this dimension should presumably predict which sentimental rule will “call the shots” on what the individual regards to be the morally right choice for the protagonist to take.

This design also allows us to test the two possible variants of Nichols’s theory (Nichols, 2004; Royzman et al., 2009). According to the first of these – the “on-line” variant – the trait measures of disgust and sympathy will strongly correlate with individuals’ moral judgment, because of the trait measures’ relation to activation of corresponding emotions (sympathy and disgust) at the time of judgment, with this proximal emotional activation being the true causal driver of judgment. According to the second, dispositional, or “off-line” variant (which appears to be Nichols’s own designated default), the trait measures of the relevant emotions should predict the kinds of higher

<sup>2</sup> For evidence of temporal stability for EC, see Davis and Franzoi (1991); for evidence of temporal stability for DS-R, see Rozin (2008); Rozin, Haidt, McCauley, Dunlop, and Ashmore (1999).

order prohibitions that individuals adopt with respect to harm and purity violations. From these higher-order prohibitions, more specific prohibitions on incest or on leaving a family member in the lurch could be (perhaps unemotionally) derived. In turn, when applied to particular cases, these more specific prohibitions should give rise to case-specific judgments corresponding to the higher-order rules.

In each case, the main sentimentalist prediction is the same: a person's perception of the morally right choice will be dictated to a considerable degree by the balance of disgust and sympathy within that individual's "affective portfolio".<sup>3</sup>

This design also allows us to test the classic reason-based theory described above, according to which moral judgments should be determined not by affective responses, but instead by abstract assessments of the relative degree of harm (costs) caused by each alternative course of action. On this Turielian/Kohlbergian account, when two moral rules (the incest taboo on the one hand, the obligation to protect and tend to next of kin on the other) collide, it is this assessment of relative harms, embodied in something like the principle of "doing least harm", that will be appealed to in order to "break the tie".

### 3. Study 1: testing the dispositional account

#### 3.1. Method

##### 3.1.1. Participants

Our analyses focus on a sample of 84 participants (59 female) who completed both sessions 1 and 2 portions of the study as part of an in-class exercise in a large introductory psychology course at the University of Pennsylvania.

##### 3.1.2. Design, procedure, and materials

The two sessions of the study were separated by a period of three weeks. Both sessions were described to the

participants as being relevant to course material, but no indication was given that the two sessions were related in any way.

Session 1 was designed to assess participants' dispositional levels of disgust, and sympathy, and to gauge their prioritization of different higher-order ethics, as measured by the moral foundations questionnaire (Graham et al., 2009). On the opening page, participants were provided with the ruse that the study was a "test-retest reliability project" and that its overall point was to assess the reliability of the various measures being used. They were instructed that they would be filling out the same measures at some point later in the semester. In line with this ruse, they were asked to provide a unique personal ID code so as to enable us to identify their later responses without compromising their anonymity.

Two separate versions of the survey were constructed, which counter-balanced the order of the disgust sensitivity and sympathy scales. The disgust-sensitivity scale, DS-R (Olatunji et al., 2007), featured a total of 25 items based, in part, on the original Disgust Scale or DS (Haidt et al., 1994). The Cronbach's alpha for these items in our study was .80. The sympathy scale consisted of the seven items comprising Davis's (1994) empathic concern sub-scale (Cronbach's alpha = .79), tapping a construct most proximate to the lay meaning of compassion or sympathy (Royzman & Rozin, 2006). Following these two measures, participants were given a filler task followed by the moral foundations questionnaire (Graham et al., 2009), which is a self-report measure of the extent to which individuals accord importance to different sets of moral principles (harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, purity/sanctity). The most relevant sub-sections of this scale for our purposes were the harm/care sub-scale, which could be seen as a measure of the extent to which sympathy concerns are translated into corresponding higher order moral prohibitions, and the purity/sanctity subscale, which could be seen as a measure of the extent to which disgust concerns are translated into corresponding higher order moral prohibitions. Cronbach's alphas for the harm/care and purity/sanctity sub-sections of this measure were .464 and .644, respectively. Due to the low internal reliability for this measure, a summary score for harm/care was not included in the analyses for this study. Given that we nonetheless wanted the concept of a higher order ethic of harm/care to be represented in these analyses, we relied on the score from a single item whose content appears to be most relevant to the present study's concerns. On this item, participants rated the importance of "whether or not someone has suffered emotionally" to their considerations of whether something is right or wrong (referred to as "suffered emotionally" subsequently).

Session 2 consisted of a questionnaire designed to measure participants' moral responses to four separate moral dilemmas. The questionnaire was presented to participants under the ruse of being concerned with "morality, gender, and culture". Participants were warned that they might find some of the descriptions unpleasant and/or offensive and that they were free to withdraw from the study at any point.

<sup>3</sup> Nichols's theory is not the only one predicting that one's reaction to either *The hike* or *Gregor's choice* will be largely determined by the balance of affective forces underpinning the competing deontological commitments. The sentimentalist component of Haidt's (2001) SIM, Prinz's (2008) neo-Humean account of affect-based morals, and Greene's dual-process theory of moral judgment seem to offer a similar prediction. Greene's view (e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), for example, appears to be that in any situation where aversive action is pitted against utilitarian gains (as in the case where one has to decide whether it is appropriate to shove *one* large man off a bridge to deter a runaway trolley from slaughtering *five* of his peers) the underlying dynamic involves a conflict between the pre-potent emotional response at the thought of murderously assaulting a fellow human being, (what Greene, 2011, construes to be "the impassioned voice" of our ancestral instincts), and a more reason-based prudential calculus that counsels (coldly and serenely) "5 is more than 1 (and this is all that matters now)." On this view, the moral outcome is thought to be determined by the relative strength of these two responses. This theory naturally extends to a case where the "dispute" is not between an affectively-tinged deontological commitment and utilitarian calculus, but rather, between two distinct affectively tinged commitments (two "impassioned" voices speaking at cross purposes). On Greene's account, it seems that, in this sort of case, the moral judgment should be determined substantially by the balance of the two affective responses underpinning the conflicting deontological commitments; in a word, the shriller voice should win.

Each of the four dilemmas described a protagonist who was faced with an unavoidable decision about whether to perform incestuous sexual intercourse in order to prevent a serious, sometimes life-threatening harm to another person. In each case, the protagonist decided not to perform incest, thereby causing serious harm (in an informal pre-study, this narrative development was found to be most conducive to producing the requisite normative tension and thus requisite response variability). The *Gregor's choice* scenario and *The hike* scenario described above were among the four scenarios. The remaining scenarios were as follows. The *Art film* scenario described the actions of a film actor who refuses to have sex with her half-brother, thereby denying him much needed funds that the film would have provided to pay for his son's (her nephew's) medical treatment. The *Kinky druggist* scenario described the actions of a woman who refuses to have sexual intercourse with her brother, thereby failing to accede to the demands of a kinky druggist, who if satisfied, would have paid for a treatment that might have saved their father from becoming blind. The *Art film* and *Kinky druggist* scenarios are included in the [Appendix A](#).

These scenarios were presented in one of four different random orders. After each one, participants were asked to indicate whether they thought the protagonist had made the right choice (yes/no); what they thought of the morality of the protagonist's choice, on a 9-point scale ranging from  $-4$  (extremely morally wrong) to  $+4$  (extremely morally right); and their overall impression of the protagonist on a scale ranging from  $-4$  (deeply immoral) to  $+4$  (deeply moral). Following these measures, participants were asked to indicate which of the two courses of action (having sex vs. not having sex with the sibling) facing the protagonist carried "greater overall costs for all concerned". Thus, for the *Gregor's choice* scenario, the participants were instructed to indicate their opinion "as to which of the two courses of action open to Gregor by the scenario's end carried greater overall costs for all concerned" where "costs" were defined to include physical harm, but also negative effects on "emotional and spiritual health, family relations, self-respect, reputation, and the like."

Following presentation of the four scenarios, participants were asked to indicate their sibling status, by indicating the number of opposite sex siblings they had, and the number of years they co-resided with each of them between the ages of 0 and 18. These items, added for exploratory purposes, were taken from [Lieberman, Tooby, and Cosmides \(2003\)](#), who previously found them to predict moral opposition to third-party sibling incest and construed them as a proxy for hard-wired affect-laden intuitions mediating incest avoidance. Finally, participants indicated their general political orientation (conservative vs. liberal), and completed a suspicion check on whether they thought the study had been related to the one they had performed three weeks earlier. Once all participants had completed the questionnaire, they were asked to write down the same personal ID code that they had used in session one.

If the sentimental theory is correct, moral judgments should be best predicted by the trait measures of disgust and empathy, and perhaps by the relative balance of the

**Table 1**

Descriptives for one-time measures and scenario-specific items in Study 1.

Variable	Scenario		
	"Gregor's choice"	"Art film"	"The hike"
<i>One-time measures</i>			
Disgust score (0–100)	48.10 (5.51)		
Empathy score (0–28)	19.19 (4.82)		
Disgust dominance score (standardized)	−0.03 (1.13)		
Purity score	14.44 (3.70)		
Importance of emotional suffering rating (0–5)	3.89 (1.01)		
Purity dominance (standardized)	0.00 (1.26)		
Total time of co-residence	18.15 (14.02)		
<i>Scenario-based items</i>			
Percent endorsing that target made right choice morally to refuse an incestuous sexual encounter	54.3%	68.7%	56.6%
Morality rating of target ( $-4$ to $4$ )	0.16 (1.80)	0.72 (1.75)	0.34 (1.58)
Overall impression of target ( $-4$ to $4$ )	0.69 (1.70)	1.10 (1.33)	0.80 (1.57)
Percent endorsing that target having sex carried greater costs	42.9%	35.7%	32.1%

*Notes.* Disgust dominance score is the empathy score standardized, subtracted from the disgust score standardized. Purity dominance score is the importance of emotional suffering rating standardized, subtracted from the purity score standardized. Numbers in parentheses next to the name of the variables are the range of possible scores.

purity/sanctity and harm/care subscales of the moral foundations questionnaire (as well as, on a certain view, a participant's sibling status). On the other hand, if the reason-based theory is correct, moral judgments should be determined solely by proximal assessments of the degree of harm caused by each course of action.<sup>4</sup>

## 4. Results

### 4.1. Preliminary analyses

Excluding those who completed only one of the two study questionnaires left us with a sample of 84 ( $n = 59$ , 70.2% female). Sample descriptive statistics for each variable included in the analyses can be found in [Table 1](#). Distributions were examined for all continuous variables. All distributions approximated normality. Just over half of the sample indicated that the target made the right moral choice by not engaging in incestuous sex in the *Gregor's choice* and *The hike* scenarios. Just over two-thirds of the sample indicated that avoiding incestuous sex was morally right in the *Art film* scenario. Unlike the other three

<sup>4</sup> One might think that the reason-based theory would predict that scores on the harm/care sub-scale of the moral foundations questionnaire should also predict moral judgment. This idea has some merit, although the chief prediction of the reason-based theory is that proximal assessments of harm should be of greatest importance, rather than a more general orientation towards concern with harm. This is so because the reason-based theory assumes that the large majority of people are already generally oriented towards a concern with harm. Thus, what should matter most in predicting moral judgment is how much harm each individual discerns as having been caused in any particular situation.

**Table 2**

Odds ratios from final models of logistic regressions predicting endorsement of the agent's refusal to engage in incestuous sex, in Study 1.

Variable	Scenario					
	"Gregor's choice"		"Art film"		"The hike"	
	Odds ratio	95% CI for O.R.	Odds ratio	95% CI for O.R.	Odds ratio	95% CI for O.R.
Disgust dominance (standardized)	0.47	0.24–0.94	0.98	0.55–1.72	0.95	0.60–1.53
Purity dominance (standardized)	1.16	0.67–2.04	1.18	0.74–1.88	1.10	0.72–1.68
Total time of co-residence	0.98	0.93–1.03	1.01	0.97–1.05	0.97	0.93–1.01
Endorsement that target acting carried greater costs <sup>a</sup>	173.25**	18.90–1587.87	28.05*	3.18–247.57	10.36**	2.53–42.52
Gender <sup>b</sup>	0.60	0.10–3.69	0.34	0.09–1.27	0.37	0.10–1.38
Order of scenario presentation	W = 3.98	ns	W = 2.90	ns	W = 3.69	ns

Notes. Variables were entered hierarchically with the order item in step 1, gender in step 2 and all other variables in step 3. Disgust dominance score is the empathy score standardized, subtracted from the disgust score standardized. Purity dominance score is the importance of emotional suffering rating standardized, subtracted from the purity score standardized.

<sup>a</sup> Endorsement of target's actions: not having incestuous sex carries greater risk = 0; having incestuous sex carries greater risk = 1.

<sup>b</sup> Gender: male = 1, female = 0.

\*  $p \leq .01$ .

\*\*  $p \leq .01$ .

scenarios used in Study 1 (*Gregor's choice*, *The hike*, *Art film*), which elicited a substantial level of moral disagreement among our participants, there was little disagreement with respect to the Kinky Druggist scenario where only a minority (6 out of 82 subjects) of subjects thought the targets' opting to have incestuous sex for the sake of procuring a sight-saving medicine for their father would be a morally superior choice. Given this lack of variability, the responses to the Kinky Druggist scenario offered little opportunity to test our main hypothesis (predictably, when the relevant analyses were performed, neither costs nor affective variables were significant correlates of moral choice) and are not reported further.<sup>5</sup>

#### 4.1.1. Primary analyses

The disgust and empathy scores were combined to yield a single "disgust dominance" score. We did this by subtracting a standardized version of the empathy score from the standardized disgust score. For the moral foundations questionnaire, a "purity dominance" score was created by subtracting a standardized version of the "suffered emotionally" item from the standardized purity score<sup>6</sup>. Due to the use of multiple tests of significance, we adopted an alpha value of .01 in all regression analyses conducted for Study 1.

A series of hierarchical logistic regressions was carried out to predict participants' endorsement of the binary item that the target made the right choice morally to refuse to engage in incestuous intercourse (yes/no). A separate regression was conducted for each of the three scenarios.

<sup>5</sup> This decision is consistent with prior practices in this area. For example, in a study that represents a major theoretical antecedent to the present work, Nichols (2002) found that one of the two scenarios he used to instantiate a breach of tableside manners failed to generate sufficient response variability. Pointing out that "this left little opportunity to explore individual differences" (Nichols, 2002, p. 231), Nichols opted to base his analysis on the single remaining scenario. Our decision followed a similar rationale. It should be noted however, that in our case, the number of scenarios remaining was three rather than one, with all three yielding a very similar, theoretically meaningful pattern of results.

<sup>6</sup> Essentially the same results hold for this variable when the other items from the harm/care sub-scale are used in these analyses in place of the "suffered emotionally" item.

The order in which scenarios were presented was entered in the first step of the model,<sup>7</sup> followed by gender in the second step, and finally the emotion (purity dominance) and norm endorsement (disgust dominance) items, co-residence and the item asking which alternative (i.e., sex or no sex) carried the greater costs to all involved (heretofore referred to as the "costs" variable), which were all entered in the third step. As predicted by the reason-based theory, the "costs" item was a statistically significant predictor for all three scenarios. In all cases, endorsement of the notion that having incestuous intercourse carried greater costs for all involved was predictive of endorsement that the target's (incest-avoidant) actions were morally right. None of the other variables were statistically significant predictors in any model (Table 2).

#### 4.1.2. Additional analyses

We also conducted multiple linear regressions to predict endorsement of the morality of the target's choices (i.e., "morality") and their impression of the target's moral character (i.e., "impression") on the aforementioned 9-point scales. Results were similar to the logistic regressions, with "costs" being a significant predictor for both morality and impression, and gender being a significant predictor of "morality" ratings in the *Art film* and *The hike*.

An alternate series of logistic and multiple regressions was carried out substituting separate disgust and empathy variables, along with separate purity and emotional suffering variables instead of the combined variables utilized in the prior analyses. Just as the combined emotional response variables were not significant predictors in any models, the separate emotional response variables were not significant predictors in any model either. In all of these alternate models, the significant predictors reported above were again statistically significant (i.e., "costs" being a significant predictor in all models and gender a significant predictor of the 9-point morality variable in the *Art film* and *The hike* scenarios).

<sup>7</sup> The results were the same with the order variable omitted from the regression models.

## 5. Study 2: testing the on-line account

The results of Study 1 lend support to our conjecture that individuals' assessments of the relative costs of competing courses of action, rather than their affective reactions, are likely to be the primary determinant of their moral judgment in the dilemmatic context. As such, these results do not support either the dispositional, or the on-line variants of Nichols' account, while lending some validity to the reason-based account instead. However, Study 1, while constituting an adequate test of the dispositional variant of Nichols's theory, is not an ideal test of the on-line variant. Our measures of disgust and sympathy in Study 1 were general, dispositional measures, and were less than precise gauges of participants' affective responses with respect to the specific dilemmas described. In contrast, participants' views about the costs and benefits of the different available courses of action were, by definition, assessed in relation to each specific dilemma. Moreover, our measure of costs and benefits was not only more specific, but also more proximate in time to participants' moral responses, which may have given it a further predictive edge. Accordingly, to better test the on-line variant of Nichols's sentimentalist theory against the reason-based cost-benefit alternative, Study 2 measured subjects' specific affective responses to each moral dilemma, doing so immediately after they provided their moral judgment to that dilemma.

### 5.1. Method

#### 5.1.1. Participants

Twenty-eight participants (16 female) completed the questionnaire as part of a class exercise in a small undergraduate lecture course at the University of Pennsylvania.

#### 5.1.2. Design, procedure, and materials

The study was conducted in a single session. Participants were presented with two of the moral dilemmas that were used in Study 1, the *Gregor's choice* and *The hike* scenarios, with the salient difference being that, in an attempt to further enhance response variability, Katriana's punishment was now "commuted" from a speedy execution to a health-wrecking ordeal in a labor camp. Also, while the scenarios used in Study 1 ended invariably with the protagonist taking a moral stance on the issue (all in the direction of incest avoidance and concomitant sibling sacrifice, the mode of dilemmatic resolution that was found, during an informal pre-test, to generate greatest moral tension and, thus, the promise of sufficient response variability), the Study 2 scenarios concluded at the earlier juncture in the decision-making process, with the protagonist still being torn between two alternate courses of action as the narrative came to a halt. This change was introduced for a number of reasons, including as a means to further test the robustness of the original effect as well as to set the stage for follow-up questions regarding the participants' emotional reactions, which were elicited by asking participants to contemplate, *in turn*, one of the two possible counterfactual endings (incest vs. labor camp interment; see the questions below).

As before, the scenarios were presented in counter-balanced order, followed by four questions. Participants were first asked which course of action would be the morally right one to take. Thus, for the *Gregor's choice* scenario described above, the participants were asked: "In your opinion, what would be the morally right thing for Gregor to do in this situation?" with the participants having to check either "Having sex with his twin sister Katriana", or "Not having sex with his twin sister Katriana" as their preferred option. Participants were then asked to imagine two different courses of action a protagonist could take (along with their various consequences) and rate their affective responses to each. They rated, for example, how much sympathy they were feeling, on a scale from 0 to 10, with respect to the harmful consequences of the course of action in which incest was not performed and, separately, how "grossed out" they were feeling with respect to the consequences of performing the incestuous act. Thus, in the case of the *Gregor's choice* scenario, they were instructed: "Please imagine now Katriana being sent to a labor camp. How much sympathy are you feeling right now?" Next, they were asked to consider the alternative course of action, in which incest was performed, and to rate how "grossed out" they were feeling right now. Thus, for the *Gregor* scenario, they were instructed: "Please imagine now Gregor and Katriana having sex. How grossed out are you feeling right now?" Finally, as in Study 1 (and using similar wording), the participants were asked to check which of the two courses of action available to Gregor (having sex with Katriana vs. refusing to have sex with her) (or Dave) carries greater overall costs for all concerned.

## 6. Results

### 6.1. Preliminary analyses

Sample descriptives for each variable included in analyses reported subsequently can be found in Table 3. Distributions were examined for all continuous variables. The sympathy variables tied to both scenarios were found to have a negative skew, which was corrected by squaring each sympathy variable.

Just over half of the sample endorsed incest avoidance as the morally appropriate response in *Gregor's choice* scenario, with only five participants making a similar endorsement in *The hike* scenario. However, mean ratings of disgust and sympathy were similar for the two vignettes as were evaluations of cost.

### 6.2. Primary analyses

Given the small *N* of Study 2, rather than including order of scenario presentation and gender in the primary models, they were tested separately for their relationship to judgments of the morality of the target's decision. Neither gender nor order had a significant relationship with evaluation of whether the target was morally right in either scenario and thus were excluded from further analyses.

Similar to Study 1, the disgust and sympathy scores for each scenario were combined to yield a single "disgust



**Table 3**

Descriptives for one-time measures for scenario-specific items, in Study 2.

Variable	Scenario	
	"Gregor's choice"	"The hike"
Percent endorsing that target made right choice morally to refuse an incestuous sexual encounter	50%	17.9%
Disgust rating of targets having sex (0–10)	6.35 (2.91)	6.21 (2.69)
Sympathy rating regarding negative events that would result from targets not having sex (0–10)	7.52 (2.21)	7.77 (2.65)
Disgust dominance score	−0.03 (1.42)	0.00 (1.38)
Percent endorsing that target having sex carried greater costs	42.3%	32.1%

Notes. Sympathy score was square root transformed for the purpose of analysis but the raw score is presented for ease of interpretation. Disgust dominance score is the sympathy score standardized, subtracted from the disgust score standardized. Numbers in parentheses next to the name of the variables indicate the range of possible scores.

dominance" score. We did this by subtracting a standardized version of the empathy score from the standardized disgust score.

Again, in keeping with Study 1, logistic regression was used as the primary analytic approach. However, due to concerns about the small *N*, the more conservative chi-square analysis was used as the supplementary test.

Given the distribution of moral judgments in *The hike* scenario (and the small *N* of the sample), use of logistic regression was not appropriate in this case, leading us to pursue logistic regression analysis with *Gregor's choice* only. In the initial logistic regression, only the "costs" item and the disgust dominance variables were entered. As in Study 1, the "costs" item was a significant predictor of judgments of moral rightness, such that participants who thought that the incestuous act resulted in greater overall costs were more inclined to morally oppose it,  $B = 3.54$ ,  $SE B = 1.28$ , odds ratio = 34.62 (95%  $CI = 2.84–422.80$ ),  $p = 0.006$ , whereas disgust dominance was not,  $B = 0.16$ ,  $SE B = 0.44$ , odds ratio = 1.17 (95%  $CI = 0.49–2.78$ ),  $p = 0.724$ . A subsequent logistic regression was carried out substituting the individual affect items for the disgust dominance score. The individual affect items were not significant predictors of moral judgment either. "Costs" was again a significant predictor in these analyses.

The chi-square analysis revealed a similar pattern. There was a significant relationship between moral judgment and evaluation of costs for the *Gregor's choice* scenario,  $\chi^2(1, N = 26) = 12.76$ ,  $p < .001$ . Those who endorsed incest avoidance as the morally right course of action also tended to view incestuous sex as carrying greater costs than the alternative, with the reverse tending to be true for those who endorsed incestuous sex as the morally appropriate choice (given the facts at hand). Considering the low number of participants who endorsed incest avoidance as the morally appropriate response in *The hike* scenario, a Fisher's exact test was utilized in lieu of the chi-square. The test was statistically significant at  $p < .001$  with a similar relationship between judgments of moral rightness and costs as in the *Gregor's choice* scenario.

### 6.3. Discussion

The results of Study 2 are consistent with the main results of Study 1, showing that participants' assessments of the relative costs associated with alternative courses of action under consideration was a significant determinant of the moral judgment in the dilemmatic context. Again, no

significant relation was found between participants' reported emotions (disgust and sympathy) and their moral judgments. In contrast to Study 1, participants' emotions were measured immediately after their moral judgments, allowing us to rule out differential proximity as an alternative explanation for the stronger relation between the costs measure and the moral judgment measure in Study 1. Moreover, emotions were assessed with respect to the specific incidents described in the scenarios, which rules out greater specificity as an alternative explanation for the stronger relation between the costs measure and moral judgment in Study 1.

However, a further deflationary possibility remains which if true would compromise these results. The costs question – requiring a direct comparison between the alternative courses of action, incest or its refusal, was a dichotomous 'yes' or 'no' answer – thus having a form that is more similar to the moral judgment question than is the form of the emotion probes, which asked about each course of action separately. It remains possible, therefore, that the greater predictive power of the costs question merely reflects the participants' relying on the similarity of form as a cue for their responses. This reliance on formal similarity could be construed either as a kind of demand characteristic, such that participants felt an implicit pressure to align their answers to the costs question with their previous answers to the moral judgment question, or more benignly, as simply reflecting a kind of mental shortcut used by participants in the interests of cognitive economy – the first response to the moral judgment question acting as a kind of implicit prime for the second response to the costs question.

Neither of these possibilities explains why the costs question was still the best predictor of a number of continuously measured dependent variables in Study 1 (degree of wrongness of the action, and overall morality of the protagonist). Nor do they explain why no connection was found between temporally stable measures of disgust and sympathy and the target moral choice variable in Study 1. Nevertheless, to rule these possibilities out more conclusively, we decided to conduct an additional study, in which the form of the costs question and emotion question were equalized.

## 7. Study 3: replication controlling for question form

### 7.1. Method

Twenty six undergraduates in a small introductory psychology course (18 female, two not identified), were asked

to volunteer 5 min of their time to complete a version of the survey used in Study 2, featuring the later (Study 2) version of the *Gregor's choice* scenario (one scenario that consistently produced high-variability responses on the moral choice variable). Following the scenario was an emotional response probe that read as follows:

Now, given the facts of the case, which of the following do you find more emotionally distressing to think about?

Katriana being shipped off to a labor camp

Gregor and Katriana having sex

This was followed by the costs probe that was the same as in Study 2. The order of the options within each probe was counterbalanced.

## 7.2. Results

Given the small  $N$  and the dichotomous nature of both independent and dependent variables, chi-square (followed by Fisher's exact test) was used as the primary means of analysis. Neither order effects, nor gender effects were significant, so the data were collapsed across them. The results corroborate those of the previous studies. First, the emotion probe (phrased dichotomously and being more proximate to the moral choice question than the costs question) was, as in Studies 1 and 2, not significantly predictive of moral choice.  $\chi^2(1, N = 26), p < .114$ , Fisher's exact test,  $p = 0.190$ . However, as in Studies 1 and 2, the costs question was significantly predictive of moral choice, with those participants who thought greater costs resulted from the incestuous act being more inclined to morally oppose it,  $\chi^2(1, N = 26), p < .019$ , Fisher's exact test,  $p = 0.038$ . Thus we believe our findings cannot be explained away on the basis of the formal relation between the relevant questions. The negative findings, i.e., the lack of any discernible relationship between relevant pre-normative affect and moral judgment (witnessed across all three studies), are particularly hard to dismiss on this basis.

## 8. General discussion

In the current studies, two variants of Nichols's sentimental account (most generally, the idea that judgments of wrong underwritten by sufficiently strong negative emotion will come to be imbued with special status and weight) were examined in a novel dilemmatic context. This sentimental account was set against the more "rationalist" proposal, according to which it is the consideration of relative costs associated with alternate courses of action, not the relative strengths of conflicting affective reactions, that is the principle determinant of dilemmatic resolution.

More specifically, we constructed a variety of scenarios that pitted two alternative courses of action against each other, in such a manner as to give rise to two strong and conflicting emotions. Each scenario featured a protagonist faced with the tragic decision of whether to perform an incestuous act in order to prevent a serious harm befalling a loved one (either their incestuous partner or a third party). We anticipated that, while the incestuous action

would rouse disgust, the actual or anticipated harm brought on by the refusal to commit the incestuous act, would be conducive to compassion or sympathetic distress. The primary sentimentalist prediction (common to either a "dispositional" or an "on-line" variant of Nichols's theory) is that the relative balance of these two reactions within an individual's "affective portfolio" should determine (or at the very least, significantly contribute to) their ultimate moral verdict. In contrast, on the reason-based alternative, affect is seen as largely irrelevant, with the individual's assessment of the overall costs and benefits being the key factor in determining judgment. Of course, it could have turned out that both affect and cost-benefit assessments mattered to some extent and our studies were capable of detecting this.

However, all three studies reported here provide in fact clear evidence *against* the sentimentalist account of moral judgment, while offering support to the reason-based alternative instead. In Study 1, neither dispositional disgust-sensitivity nor dispositional sympathy nor the balance between the two predicted *any* aspect of the participants' moral response. Nor was any aspect of the moral response predicted by participants' endorsement of higher-order prohibitions against harmful/unkind and impure actions (supposedly rooted in sympathy and disgust, respectively), as measured by the recently developed moral foundations survey (Graham et al., 2009). Rather, participants' moral judgments were robustly predicted by their overall assessments of the relative costs of the different courses of action. These data thus count against the dispositional variant of Nichols's sentimental account. However, it was less clear to us that they would count against the on-line variant of Nichols's theory, since participants' general tendency to experience disgust or sympathy may not have actually translated into greater actual disgust or sympathy at the moment of the judgment.

This concern was addressed in, and largely neutralized by, Studies 2 and 3, which replicated the overall pattern of results seen in Study 1. In both later studies the affective measures did not predict moral judgments in spite of being both specific and proximal to these judgments. Instead, participants' moral judgments were reliably associated with their assessments of the relevant costs associated with competing courses of action. Study 3 further showed that the predictive advantage of our "cost-analysis" probe cannot be accounted for by its greater formal similarity with the moral judgment question. Taken together, the data are quite inconsistent with either the "on-line" or the "dispositional" variants of Nichols's account, while favoring a reason-based, Kohlberg-style alternative instead.

One possible objection to this line of thought is that we did not measure all of the relevant emotions in play. Anger, in particular, stands out as a highly relevant moral emotion that is evoked by a variety of transgressive behaviors and thus might have been predictive of the participants' ultimate judgment even though disgust was not. While this is possible, we see no strong reasons for thinking it to be likely. More importantly, anger is not a case of "pre-normative" affect (Royzman et al., 2009), in the sense that disgust and sympathy are. That is, anger is generally not

independent of a previously encoded moral prohibition, meaning that which makes people angry already presupposes and signals antecedent normative commitments and expectations. As such, anger does not meet Nichols's own criteria for being an independent vector of normative influence. In fact, it seems difficult to conceive of how anger might be experienced without some normative content already being a part of it. Indeed, one of the virtues of Nichols' (2002) sentimental account is that it has been quite careful to distinguish moral emotions, such as anger (say, in response to a partner breaking a promise or a co-worker invading one's privacy), from pre-moral (in our language: pre-normative) affective influences, such as physical disgust or distress-induced sympathy. Thus, even if it were to turn out that anger, but not disgust or sympathy, predicted moral judgment in dilemmatic cases such as ours, this would not constitute, in view of Nichols's own reasoning on this matter, clear evidence for a purely affective influence being the independent driver of moral judgment in cases of this kind.

In conclusion, we note that, though the joint effect of the aforementioned studies is to lend some support to one *particular* aspect of Kohlberg's theoretical scheme, that concerned with the relative importance of affective vs. "rational" (harm-based) influences on moral judgment, as assayed in the context of a *particular* type of moral dilemma, it would be a mistake to see the above results as offering a general validation of Kohlberg's developmental model. Indeed, there are a number of features of this model (its presupposed invariant multi-stage sequence, its emphasis on justice as mature morality's ultimate concern) that the current studies simply do not and could not reasonably address.

### 8.1. Relation to broader literature

Some recent scholarship within moral psychology has argued that researchers ought to move beyond the current debate between emotion and reason because it leads to an unnecessary and ultimately unproductive theoretical stalemate (Leeman, 2003; Monin, Pizarro, & Beer, 2007). In accord with some of these arguments, we believe that the current findings offer guidance to future research in yet another interesting and relatively unexplored way. Just as people are capable of a variety of "religious experiences" (to co-opt the title of James' inimitable book), there is also a variety of moral experiences or "moral encounters" to be had in the course of one's daily existence. One of the peculiar features of moral psychology since its inception in the early 20th century has been that virtually all of its practitioners of notice have concerned themselves with *one* particular type of moral encounter while holding (explicitly or otherwise) their results to have import for morality as a whole. Yet, as Table 4's Moral Encounters Typology (MET) indicates, there is a remarkable, even dazzling heterogeneity within the moral domain. Among other things, the table illustrates the possible intersections between three different drivers of, or inputs to, moral thought. One of these inputs is what we have been referring to consistently as *sentimental rules*, by which we mean any moral rule (value, precept, belief) linked to a strong emotion. Importantly,

such rules could concern issues of "symbolic harm" (e.g., desecration of a national monument) or seemingly harmless acts of sexual impropriety (consensual incest between adults). But they can equally concern issues that do involve harm proper, as long as they are linked to strong emotion – such as the emotion inherent in rules against torturing children or puppies. A second discreet input to moral judgment specified by MET is *unsentimental rules* – moral rules that are not linked to strong emotion. Examples of such rules might include rules prohibiting tax evasion or document forgery (see Royzman et al. (2009), for a demonstration that people do treat business tax evasion as a genuinely moral offense even though little emotion appears to be involved). Finally, the third basic input specified by MET is represented by the rigors of "consequentialist logic", which may be regarded as the master-rule of prudential reasoning (and which may operate in intrapersonal and interpersonal arenas alike). According to this rule, the magnitude of the consequences of different possible courses of action are weighed against each other in order to make a moral or prudential judgment.



As Table 4 illustrates, each of these inputs has an "unfettered" application in which they are enacted in an undiluted (i.e., non-combinatorial) way. For instance, a simple case-specific judgment that it is wrong to eat a puppy or torture a child represents an unfettered application of "sentimental" considerations about these matters (ditto for simple case-specific judgments concerning tax evasion or document forgery as "unsentimental" rules in action), just as a decision about which of two public works projects, A or B, is more appropriate to fund, based on their consequences, represents an unfettered application of consequentialist logic.

However, in addition to their unfettered applications, each of these inputs can come into collision with each other. Such collisions manifest themselves as moral dilemmas and we offer several representative cases of these in Table 4. Two sentimental rules can clash, as in the examples we have investigated in this paper, when it has to be decided whether it is permissible to commit incest in order to save another person's life. This pits two rules that evoke strong emotions, disgust and sympathy, respectively, against each other. Sentimental rules can also clash with unsentimental rules (Cell 4). Similarly, sentimental rules can collide with consequentialist logic, as when a decision needs to be made whether it is permissible to torture one person in order to prevent a terrorist attack which will kill and injure many people (Cell 7). Unsentimental rules can also collide with consequentialist logic (Cell 8). Finally, there are possible intersections between all three inputs (Cell 9).

It is also worth considering that the types of moral encounters depicted in Cells 1–9 are themselves predicated on at least two more general sorts of moral "decisions", both of which would presumably have to arise at some point in an individual's developmental history (see the bottom of Table 4). We have in mind here, first and foremost, decisions concerning whether a given normative proposition (i.e., "Thou shalt not kill") should be accorded a genuinely moral status at all (see Blair, 1995; Nichols,

**Table 4**

A tentative typology of “simple” (“unfettered”), “dilemmatic” and “foundational” moral encounters.

	Unfettered applications	Sentimental rule	Unsentimental rule	Consequentialist logic
Sentimental rule	1. (a) Is it wrong for these brother and sister to have sex?  (b) Is it wrong to torture this child?	4. (a) Is it wrong for these brother and a sister to have sex, if this is the only way to save the woman's child from an incurable disease? (b) Is it wrong for Heinz to steal a drug in order to save his wife's life?	5. (a) Is it wrong to evade paying taxes if this is the only way to save someone's life? (b) Is it wrong to forge this document if this is the only way to prevent a brother and sister having sex?	7. (a) Is it wrong to torture this child if this is the only way to prevent a terrorist attack that will kill many people? (b) Is it wrong to have sex with your sibling if this is the only way to prevent five other cases of incest?
Unsentimental rule	2. (a) It is wrong to evade paying taxes.  (b) It is wrong to forge this document to obtain a job.		6. (a) Is it wrong to forge this document if this is the only way to prevent yourself from evading your taxes?	8. 8.(a) Is it wrong to forge one document if this is the only way to stop five other documents being forged?
Consequentialist logic	3. (a) Which of two public works projects, A or B, should receive funding, based on their projected overall benefits and costs?			
Collision between all three inputs, sentimental rules, unsentimental rules, and consequentialist logic	9.  (a) Is it wrong to forge five documents if this is the only way to prevent a brother and sister having sex?			
				
	Is this a genuine moral rule?	To whom does it apply (what is the circle of moral concern)?		

2002; the above discussion) and, second, what range of entities or beings fall within its scope (Bloom, 2010; Singer, 1981). These queries too may be regarded as moral encounters, albeit of relatively unique and “foundational” variety, affecting much of the case-specific evaluative work going on “above” them.

While incomplete and unvarnished, MET, we think, provides some perspective on current debates within the field. Consistent with Leeman (2003) and Monin et al.'s (2007) astute observations, it strikes us that much of present-day moral psychology is, and has been, narrowly focused on one particular type of moral encounter (or, in some cases, a particular subset of a particular type of encounter), while purporting to make claims about morality as such. For instance, Haidt's research into the affective antecedents of case-specific moral judgments has had as its main focus the unfettered application of one particular type of sentimental rule – sex- and food-related sentimental rules backed by disgust. Yet, it is often presented (e.g., Haidt, 2001) as offering a serious challenge to Kohlberg's earlier work even though Kohlberg's research was concerned with quite different sorts of moral encounter – broadly speak-

ing, clashes between two or more sentimental rules (or, on occasion, a sentimental rule and its unsentimental counterpart) whose content was neither sex- nor food-related. Similarly, Joshua Greene has claimed (e.g., Greene et al., 2001) that his own dual-process theory of moral judgment, commonly illustrated via his neuroimaging work on trolley/footbridge dilemmas, may furnish a superior, more balanced view of morality than either Kohlberg's pure “rationalism” or Haidt's (2001) “emotivism” can provide alone, even though, again, Greene's work is concerned, by and large, with a rather specific moral situation (a particular kind of moral dilemma illustrated as MET's Cell 7) and deploys a rather specific type of methodological approach. *In principle*, it is conceivable then that *all* available theoretical perspectives offer claims that are relatively well-suited to their respective explanandums but fail to generalize beyond them. Of course, it is also possible that a given perspective is false as applied even to its explanandum of choice.

Be that as it may, as the present studies indicate, Kohlberg's pure rationalist account of moral judgment, “unbalanced” and unfashionable though it may be, appears

to be largely correct when applied to the particular type of moral encounter that he was historically concerned with (and that the sentimentalist research of the past decade or so has presumed to have successfully challenged). This general point, we think, deserves more recognition in future research. Since there is not yet any obvious way to estimate which sort of moral encounter is most “representative” of (or of greatest significance for) the moral universe we inhabit, the best approach (and one that, we think, should precede any “comprehensive” model of moral evaluation) is a thorough and balanced investigation of the inner workings of each of the possible varieties of moral encounters contained therein.

## Acknowledgements

Edward B. Royzman, Department of Psychology, University of Pennsylvania; Geoffrey P. Goodwin, Department of Psychology, University of Pennsylvania; Robert F. Leeman, School of Medicine, Yale University.

We thank Rob Kurzban and Jon Baron for their valuable advice. And we thank Ghebremariam and Sarah Patrick for research assistance.

## Appendix A. The additional scenarios used in Study 1, not reported in the main text

### A.1. Art film

Lisa and Michael are both struggling actors in southern California. They are also half-siblings. Michael has a son with a rare medical condition which, if not treated properly very soon, will cause him serious mental disability within a year. Unfortunately, the treatment is enormously expensive, and will cost him close to a million dollars to finance. Michael is not rich, and is desperate to raise money for his son’s treatment. He has already raised what money he can by soliciting friends and family, but he is still several hundred thousand dollars short. Recently, Michael has been offered a lucrative job in an “art film” which requires several graphic sex scenes with a female partner who, according to the script, is his character’s long lost cousin. This part is far beyond Michael’s normal pay-grade, and, factoring in the points from the DVD sales, would earn him enough money to pay for the rest of his son’s treatment. The director of the film, despite (or, in part, as a result of) knowing that Lisa and Michael are half-siblings demands that Lisa (also an actor) play Michael’s sexual partner. He believes that Lisa is made for the part and that the two of them would “fit” perfectly together. He demands that either they both play their parts in the film, or else neither of them will. What’s more, he demands that at least one of their sex scenes, all taking place at a costumed ball, be real, and not simulated. Although very well-paid, Lisa and Michael’s parts in the film do not go much beyond the sex scenes, and the director has assured them that he and he alone will be in the room when shooting these scenes, that their true identities will never be divulged, and that they will be wearing masks through it all making them unidentifiable by people who know them. He

also promised to supply Lisa with a complimentary morning-after kit that will effectively reduce the risk of conception to a nil. But the director will not budge on their having sex with each other. Michael, seeing no alternative, asks Lisa to sign up for the role. But Lisa says “no”. Though eager to help, she feels that they simply cannot give into the director’s demands and participate in what she regards as a deeply immoral and even heinous act. . . As a result, Michael is not able to procure the funds necessary to complete his son’s treatment and his son develops serious mental disability within a year.

### A.2. Kinky druggist

Heinz and Greta are twenty something college students in the city of Hurva, South Croatia. They are also brother and sister. One day, after getting their father to see a specialist for his migraines, they discover that he has a special kind of brain tumor that, though non-lethal, will cause him to go irrevocably blind within a month’s time. There is one and only one (very pricey) experimental drug that might help him. It is a form of radium, administered intravenously, that the druggist in their very town recently discovered and is about to market world-wide. The drug is expensive to make, and the druggist is charging five times what the drug cost him to produce. Heinz and Greta went to everyone they knew to borrow the money, but were able get only about one half of what the drug costs. . . So they told the druggist all about their situation and asked him to sell the drug cheaper or let them pay later. But the druggist, who worked years to develop the drug, said “no”. However, he intimated that they may be able to “work something out” if the siblings were willing to be “morally flexible.”

. . . As the druggist went onto explain, ever since his doomed adolescent romance with his half-sister Jacqueline, he had had a persistent sexual fantasy about watching two attractive college-age siblings cuddle and make love to each other right in front of him. . . He promised that if Heinz and Greta would help him turn this fantasy into a reality by having (protected) sex in front of him, he would sell them the drug for half the cost. The siblings realize that, though a scoundrel and a deviant, the druggist is not a liar and will keep his end of the bargain. . . In fact, Heinz had already declared his willingness to accept the druggist’s offer as the lesser of two evils. However, Greta said “no”: she simply could not and would not be a party to what she regards as a deeply immoral and even heinous act. . . As a result, Heinz and Greta’s father does not get the necessary treatment and goes irrevocably blind within a month.

## References

- Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, 57, 1–29.
- Blair, R. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Personality and Individual Differences*, 27, 135–145.
- Blair, R., Jones, L., Clark, F., & Smith, M. (1997). The psychopathic individual: A lack of responsiveness to distress cues? *Psychophysiology*, 34, 192–198.
- Bloom, P. (2010). How do morals change? *Nature*, 464, 490 (March 25).
- Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment*. Cambridge: Cambridge University Press.

- Corley, M. (2009). Pat Robertson: Gay marriage is 'the beginning in a long downward slide' to legalized child molestation (includes a video file). Available from: <http://thinkprogress.org/politics/2009/05/07/39152/robertson-child-molestation/>. Retrieved June 29, 2011.
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Madison, WI: Brown & Benchmark Publishers.
- Davis, M. H., & Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *Journal of Research in Personality, 25*, 70–87.
- Dolan, M. S., & Fullam, R. S. (2010). Moral/conventional transgression distinction and psychopathy in conduct disordered adolescent offenders. *Personality and Individual Differences, 49*, 995–1000.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology, 96*, 1029–1046.
- Greene, J. D. (2011). Social neuroscience and the soul's last stand. In A. Todorov, S. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 263–273). New York: Oxford University Press.
- Greene, J. D., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108 (September 14).
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–628.
- Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences, 16*, 701–713.
- Hollos, M., Leis, P. E., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology, 17*, 352–374.
- Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), *Psychology and genetic epistemology* (pp. 151–235). New York: Academic Press.
- Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research, and social issues* (pp. 31–53). New York, NY: Holt, Rinehart and Winston.
- Leeman, R. F. (2003). Rational moral judgments are not just "tail-bound": A review of the moral judgment literature and a critique of the social intuitionist perspective. Unpublished manuscript, University of Pennsylvania.
- Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society B: Biological Sciences, 270*, 819–826.
- Madden, T. (1992). Cultural factors and assumptions in social reasoning in India. Unpublished doctoral dissertation. Berkeley: University of California.
- Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology, 11*, 99–111.
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition, 84*, 221–236.
- Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. New York, NY: Oxford University Press.
- Nichols, S. (2008). Sentimentalism naturalized. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The cognitive science of morality: Intuition and diversity* (Vol. 2, pp. 255–274). Cambridge, MA: MIT Press.
- Nucci, L. P. (2001). *Education in the moral domain*. Cambridge, UK: Cambridge University Press.
- Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuck, C. N., Lohr, J. M., et al. (2007). The disgust scale: Item analysis, factor structure, and suggestions for refinement. *Psychological Assessment, 19*, 281–297.
- Prinz, J. (2008). *The emotional construction of morals*. New York: Oxford University Press.
- Royzman, E. B., Baron, J., & Goodwin, G. P. (in preparation). Is disgust a moralizing emotion? Exploring the sufficiency thesis, Manuscript in preparation, University of Pennsylvania.
- Royzman, E. B., Leeman, R. F., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition, 112*, 159–174.
- Royzman, E. B., Leeman, R. F., & Sabini, J. (2008). "You make me sick": Moral dyspepsia as a reaction to third-party sibling incest. *Motivation and Emotion, 32*, 100–108.
- Royzman, E. B., & Rozin, P. (2006). Limits of symhedonia: The differential role of prior emotional attachment in sympathy and sympathetic joy. *Emotion, 6*, 82–93.
- Rozin, P. (2008). Hedonic adaptation: Specific adaptation to disgust/death elicitors as a result of dissecting a cadaver. *Judgment and Decision Making, 3*, 191–194.
- Rozin, P., Haidt, J., McCauley, C., Dunlop, L., & Ashmore, M. (1999). Individual differences in disgust sensitivity: Comparisons and evaluations of paper-and-pencil versus behavioral measures. *Journal of Research in Personality, 33*, 330–351.
- Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–83). Chicago: University of Chicago Press.
- Singer, P. (1981). *The expanding circle*. Oxford, UK: Clarendon Press.
- Smetana, J. G. (1981). Reasoning in the personal and moral domains: Adolescent and young adult women's decision-making regarding abortion. *Journal of Applied Developmental Psychology, 2*, 211–226.
- Smetana, J. G. (1983). Social-cognitive development: Domain distinctions and coordinations. *Developmental Review, 3*, 131–147.
- Smetana, J., & Braeges, J. (1990). The development of toddlers' moral and conventional judgements. *Merrill-Palmer Quarterly, 36*, 329–346.
- Smetana, J. G., Campione-Barr, N., & Yell, N. (2003). Children's moral and affective judgments regarding provocation and retaliation. *Merrill-Palmer Quarterly, 49*, 209–236.
- Song, M. J., Smetana, J. G., & Kim, S. Y. (1987). Korean children's conceptions of moral and conventional transgressions. *Developmental Psychology, 23*, 577–582.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
- Turiel, E. (2002). *The culture of morality: social development, context, and conflict*. Cambridge, UK: Cambridge University Press.
- Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155–244). Chicago: University of Chicago Press.
- Turiel, E., & Wainryb, C. (1994). Social reasoning and the varieties of social experience in cultural contexts. *Advances in Child Development and Behavior, 25*, 289–326.
- Wainryb, C. (1991). Understanding differences in moral judgments: The role of informational assumptions. *Child Development, 62*, 840–851.
- Wainryb, C., & Turiel, E. (1993). Conceptual and informational features in moral decision making. *Educational Psychologist, 28*, 205–218.
- Westermarck, E. (1906). *The origin and development of the moral ideas*. London: McMillan.