

# Are good reasoners more incest-friendly? Trait cognitive reflection predicts selective moralization in a sample of American adults

Edward B. Royzman\*    Justin F. Landy†    Geoffrey P. Goodwin†

## Abstract

Two studies examined the relationship between individual differences in cognitive reflection (CRT) and the tendency to accord genuinely moral (non-conventional) status to a range of counter-normative acts — that is, to treat such acts as wrong regardless of existing social opinion or norms. We contrasted social violations that are *intrinsically* harmful to others (e.g., fraud, thievery) with those that are not (e.g., wearing pajamas to work and engaging in consensual acts of sexual intimacy with an adult sibling). Our key hypothesis was that more reflective (higher CRT) individuals would tend to moralize selectively — treating only intrinsically harmful acts as genuinely morally wrong — whereas less reflective (lower CRT) individuals would moralize more indiscriminately. We found clear support for this hypothesis in a large and ideologically diverse sample of American adults. The predicted associations were not fully accounted for by the subjects' political orientation, sensitivity to gut feelings, gender, age, educational attainment, or their placement on a sexual morals-specific measure of social conservatism. Our studies are the first to demonstrate that, in addition to modulating the intensity of moral condemnation, reflection may also play a key role in setting the boundaries of the moral domain as such.

Keywords: moral/conventional, CRT, harm, rational, judgment.

## 1 Introduction

The notion that *thinking well* can make a difference to a person's moral outlook has a long and illustrious history in the annals of Western thought (e.g., Plato [e.g., *Protagoras and Meno*]; Kant, 1785/1959; Rawls, 1971; Singer, 2005). Ironically, it was David Hume, the reputed uber-sentimentalist, who penned one of the most impassioned testimonials in its defense:

The final sentence, it is probable, which pronounces characters and actions amiable or odious, praise-worthy or blameable. . . . It is *probable*, I say, that this final sentence depends on some internal sense or feeling. . . . But in order to pave the way for such a sentiment, and give a proper discernment of its object, it is often *necessary*, we find, *that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained* . . . . In many orders of

beauty, particularly those of the finer arts, *it is requisite to employ much reasoning, in order to feel the proper sentiment*; and a false relish may frequently be corrected by argument and reflection. There are just grounds to conclude, that moral beauty partakes much of this latter species . . . . (Hume, 1751/1983, p. 5, italics added)

More recently, Edvard A. Westermarck — an anthropologist, a philosopher, and Hume's fellow sentimentalist — offered an intriguing and largely untested conjecture: two parties or peoples who share core moral ideals may nevertheless find themselves in a state of pervasive moral disagreement owing strictly to the degree of cognitive sophistication with which they apply these ideals to an issue at hand.

Most people follow a very simple method in judging of an act. Particular modes of conduct have their traditional labels many of which are learnt with language itself and the moral judgment commonly consists simply in labeling the act according to certain obvious characteristics . . . . But a conscientious and intelligent judge proceeds in a different manner. He carefully examines all the details connected with the act, the external and internal conditions under which it was performed, its consequences, its motive and since the moral estimate in a large measure depends upon the regard paid to these circumstances his judgment may differ greatly from

We are grateful to Gordon Pennycook and an anonymous reviewer for their comments on an earlier version of this draft and to Jonathan Baron for his comments, editorial advice, and contributions to various aspects of this project.

Copyright: © 2014. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

\*Department of Psychology, University of Pennsylvania, 3720 Walnut St., Solomon Lab Bldg., Philadelphia, PA, 19146. Email: royzman@psych.upenn.edu.

†Department of Psychology, University of Pennsylvania.

that of the man in the street even though the moral standard which they apply be exactly the same (Westermarck, 1906, pp. 9–10).

Thus, divergent as their philosophical outlooks might have been in other respects (see Westermarck, 1906), Hume and Westermarck appear to be largely in agreement on one key point: *a general aptitude for thinking well is likely to exert a profound, even foundational, influence on a person's moral outlook.*

In what follows, we further elaborate and offer an empirical assay of this point by making use of what has long been taken to be one of the most reflexive and cognitively impenetrable moral cognitions ever (see Pennycook, Cheyne, Barr, Koehler, & Fugelsang [2014] for discussion) — the widespread repudiation of consensual sibling incest — and of one of the most demonstrably valid and widely used measures of “thinking well” — the Cognitive Reflection Test (CRT).

### 1.1 Repudiation of sibling incest and the social intuitionist model

The notion that principled moral opposition to third-party sibling incest is a result of automatic “negative imprinting” enabled by early childhood experiences (whose development both bypasses and is impervious to rational reflection) goes at least as far back as Westermarck's *Histories of Marriage* (e.g., Westermarck, 1921) (see Lieberman, Tooby, & Cosmides [2003] for a recent revival of this perspective, but see Royzman, Leeman, & Baron [2009] and Royzman, Goodwin, & Leeman [2011] for a contrarian viewpoint). The putatively non-reflective quality of this opposition is front and center in the lead paragraphs of Jonathan Haidt's (2001) influential “The emotional dog and its rational tail”. The paper opens with a duo of college-age siblings, Julie and Mark, opting for a night of non-committal sex while vacationing abroad. In Haidt's (2001) positive analysis, when reading the story, “one feels a quick flash of revulsion . . . and one knows intuitively that something is wrong” (p. 814), with subsequent reasoning being largely utilized to marshal a range of post-hoc arguments for the validity of the initial impression, while remaining largely impotent to alter the impression itself (see Haidt, Bjorklund, & Murphy, 2000). That is, in Haidt's politically minded update of Hume's metaphor, moral reflection, can (at least, in this particular case) aspire to no higher office than that of the press-secretary of the “passions”.<sup>1</sup> (In Haidt's model, “intuitions”, products of a fast, automatic, hard-to-recover, System-1 processes,

<sup>1</sup>It is important to underscore that, contrary to some of its critics, Haidt's model puts considerable weight on what might be called “external reasoning”, i.e., reasoning via argument or negotiation (Link 3 of the model; Harman, Mason, & Sinnott-Armstrong, 2010), as well as direct social “pressure” (Link 4). Nor does it entirely preclude the possibility

take place of the “passions”; nevertheless, many of his candidate “affect-laden” intuitions are clearly what Hume and others would recognize as “passions” par excellence). Since press secretaries are neither paid nor invited to tell their masters what to think, the natural corollary of Haidt's metaphor is that, while more characteristically reflective individuals may display a superior knack for defending their moral cognitions, *whatever they may be* (as well as influencing others' moral cognitions in the process), there should be no unique association between one's general capacity for private reflection and the content of one's moral beliefs. This prognosis is strikingly at odds with that of the newer wave of “reflectionist” dual process accounts of moral cognition (see Bucciarelli, Khemlani, & Johnson-Laird [2008], Paxton et al. [2012], and Pennycook et al. [2014], among others; see Saunders [2013] for a thoughtful discussion of the nature of moral reasoning), which, different as they are in some respects, concur that reason is far more than the ever-conniving spin doctor of the passions, but is rather an active and independent contributor to the moral process itself.

### 1.2 Cognitive Reflection Test

The Cognitive Reflection Test (CRT) (Frederick, 2005) consists of three mathematical puzzles thought to gauge an individual's ability and/or willingness to spend more time and mental effort examining and solving a problem characterized by an intuitively compelling but inaccurate response. Despite being only a few years old and three items short, CRT has proved itself as one of the most useful measures of cognitive performance available. To take a few examples, it has been found to predict resistance to common biases in judgments and decisions (Toplak, West, & Stanovich, 2011), including resistance to logical fallacies and overconfidence (Oechssler, Roeder, & Schmitz, 2009), more elaborate and thorough heuristic search (Cokely & Kelly, 2009), greater future-mindedness (Frederick, 2005), enhanced forecasting ability (Mellers et al., 2014), a less robust belief in paranormal phenomena (Pennycook et al., 2012), and (under specified conditions) a propensity toward utilitarian moral judgment (Paxton et al., 2012, but see Royzman, Landy, & Leeman, in press). Particularly germane to the present project, Toplak et al. (2011) found CRT to be uniquely related to a variant of the classic outcome bias task (Baron & Hershey,

of using private reflection to influence our initial judgment (Links 5 and 6), but (and this is the crucial proviso in the absence of which the model would be next to impossible to falsify and thus scientifically inert) such occurrences are said to be highly atypical or rare (2001, p. 815). Thus our predictions concerning Haidt's model should be treated as those that the model would make under “normal” or “typical” conditions, whatever these may be. An important aspect of these typical conditions for most people we know would be lack of any direct social pressure or “reasoned persuasion” to adopt a more tolerant view of ad-hoc sibling sex.

1988), with high CRT subjects tending to be significantly more resistant to the outcome bias than lower scoring subjects. This finding is of importance given that the outcome bias task's key cognitive demand lies in being able to set aside one's privileged knowledge of a decision's end result while forming a fair assessment of an actor's a priori decision-making skill (see also Pennycook et al., 2012). This, as we argue below, is also the key cognitive demand involved in one of the most widely used measures of mature moral judgment — the Moral-Conventional Distinction Task.

### 1.3 The Moral-Conventional Distinction Task (and a tentative typology of harms)

The Moral-Conventional Distinction Task (henceforth, MCDT) was developed as a methodological companion to what has become one of the most generative and influential models of moral cognition to come out of the structural-developmental tradition: the *social domain theory* (SDT) (see Nucci, 2001; Smetana, 1983; Turiel, 1983; Turiel, Hilderbrandt, & Wainryb, 1991; Turiel, 2002 for overviews, and Blair, 1995; Haidt, Koeller, & Dias, 1993; Nichols, 2002; Shweder, Mahapatra, & Miller, 1987 for further applications and critique). SDT's fundamental supposition is that socially counter-normative acts may be deemed impermissible in two qualitatively distinct ways. On the one hand, there are prototypically moral transgressions, such as a child engaging in thievery or physical assault; on the other hand, there are prototypically conventional transgressions, e.g., a child coming to school wearing swimwear in lieu of a uniform (Turiel, 2002). Both types of acts are routinely judged inappropriate (or “not OK”) by a majority of children and adults, with the differences encompassing both form and content. With regards to form, conventional transgressions are routinely considered to be less serious and less wrong than their moral counterparts. Most critically, a prototypically immoral act is generally judged to be “not OK” even when its performance is entertained under a set of counterfactual conditions designed to negate its status as a subversion of the existing normative regime.<sup>2</sup> Thus, an individual may be asked to imagine that the act at hand is no longer “against the rules,” that it has the backing of a legitimate authority, or, perhaps most meaningfully, that it has taken place within a cultural milieu where its performance has been normatively sanctioned by popular sup-

<sup>2</sup>The notion that children and adults spontaneously mutate proscriptive rules (as in “would this act still be wrong if the rule against it did not exist?”) in their quest to construct separable domains of morality and convention is not unique to Turiel (1983). Hampshire (1982, see pp. 151–152), Blair (1995), and Nichols (2002, 2004) seem to acknowledge this as well.

port (e.g., Turiel et al., 1991; Turiel, 1983, see Nucci, 2001 and Turiel, 2002 for an overview of the congruence among these measures). While such “what if” propositions tend to generally erode the perceived inappropriateness of socio-conventional misdeeds (e.g., inappropriate dress code), they leave the perceived wrongness of prototypically immoral acts (e.g., stealing) relatively intact, giving them their signature *socially transcendent* character. This basic differentiation has been observed in a remarkably wide range of subject populations separated by class, religion, geography, cognitive ability, and age (e.g., Hollos, Leis, & Turiel, 1986; Madden, 1992; Nucci, 2001; Smetana & Braeges, 1990; Yau & Smetana, 2003; see Turiel, 2002 for a review).<sup>3</sup>

One key (and often overlooked) distinction within SDT (e.g., Turiel, 1983) is the relatively nuanced way in which conventional and moral transgressions are thought to differ with respect to their content, with each giving rise to a different sort of harm. Prototypically moral transgressions involve acts that are (perceived to be) *intrinsically* harmful to others (Turiel, 1983, e.g., p. 44, p.221). In contrast, while prototypically conventional transgressions may cause harm (which is real enough, if for no other reason that they cause disruptions of the established routine and/or hold a potential to undermine the reigning institutional authority), the harms they cause are *contingent* upon aspects of an existing normative regime that could themselves be deemed arbitrary and modifiable by consensus.

Consider: within the continental United States, right-hand vehicular traffic is the norm; those who breach it would rightly be deemed a danger to others and to themselves. Yet, the act of driving on the left does not carry harmful consequences as such — it causes harm only within a particular normative regime — the socio-conventional environment where the right-hand driving is the norm. As a consequence, the act of driving on the left is typically not considered wrong in and of itself, but is regarded instead as a conventional violation (albeit one of a very serious kind). And the right-hand driving norm itself is regarded as being arbitrary and revisable: few of us would be shocked (much less call for a moral crusade) upon first learning of a place where the converse rule is in

<sup>3</sup>Most recently, Bryce Huebner and colleagues (Huebner, Lee, and Hauser, 2010) used a large multi-cultural internet sample and a wide range of newly designed adult-appropriate cases of hypothesized moral and conventional transgressions to determine (via cluster analysis) if transgressions of different types sort together in accordance with the SDT-specific criteria. The answer was a strong “yes” for the putative moral cases and a somewhat weaker “yes” for the putative conventional cases (due largely to some amorously rated cases involving traffic laws); nevertheless, “the majority of hypothesized conventional transgressions also form[ed] an obvious cluster” (Huebner et al., 2010, p. 1), with a number of central cases whose content matched that of the paradigmatic conventional offences (e.g., inappropriate dress code, inappropriate mode of address) employed in the SDT-inspired developmental research.

effect. Similar considerations hold for another prototypically conventional (not intrinsically harmful) violation: wearing pajamas to work (henceforth, *Pajamas*, Huebner et al., 2010; please see the Appendix for the details of this and other key vignettes).

Contrast this with cruelty to animals. Like breaches of vehicular traffic rules, puppy torture is a normative no-no. However, along with other “prototypically immoral” acts, e.g., killing and fraud, it has a further feature that grants it a unique status within SDT’s normative hierarchy: it is *intrinsically* hurtful to the puppies. Unlike the harms resulting from a contravention of the right-hand traffic rule or contingently inappropriate dress code, the physical pain entailed in the puppy torture case cannot be voted or legislated out of existence by a sheer feat of social reorganization. It remains constant under a variety of conceivable normative regimes. Similar considerations hold for subtler varieties of intrinsic harm, such as those instantiated in Huebner et al.’s (2010) *Tickets* vignette, where an individual callously sells counterfeit concert tickets to the public knowing full well that those who purchased them shall arrive at the concert venue filled with anticipation that will come to nothing.

Thus, contingent harms, unlike intrinsic (and, thus, socially transcendent) harms, can be psychologically nullified under an alternate normative regime. If asked whether the act of wearing pajamas to work ceases to be problematic in a wearing-pajamas-to-work-is-splendid universe, those able to effect the requisite *normative regime nullification*, i.e., those able to override their existing epistemic viewpoint (a capacity that, as we noted earlier, is likely predicted by CRT), should return an unqualified Yes. By contrast, those who are unable or unwilling to effect normative regime nullification should perceive a contingently harmful act (i.e., *Pajamas*) as intrinsically harmful and thus transcendentally wrong. In other words, cognitive reflectivity (as measured by CRT) should correlate with a tendency to treat acts comprised of contingent harms as merely conventional offenses, rather than offenses that are universally, unalterably wrong.

In contrast, intrinsically harmful acts such as puppy torture, or theft (e.g., Greene et al.’s [2001] *Wallet* vignette), would be expected to yield a *CRT-insensitive* correlational pattern, since the harm or violation of rights that they produce remains intact regardless of any counterfactual conjectures. Thus, in cases of this kind, those who reflect deeply and effectively would be expected to return the very same verdict as those who reflect little or hardly at all.

In sum, to the degree that individuals’ capacity to effect normative regime nullification is reliably indexed by the CRT, and to the degree that normative regime nullification affects contingently harmful offenses, but not their intrinsically harmful counterparts, two empirically

testable hypotheses result. First, there should be a significant *inverse* association between individuals’ CRT scores and their tendency to assign socially transcendent/non-conventional status to paradigmatically conventional (contingently harmful) offenses (i.e., *Pajamas*); second, no such association should obtain between CRT and paradigmatically immoral (intrinsically harmful) acts (i.e., *Tickets*).<sup>4</sup>

The situation, however, is rendered somewhat more complex by the fact that SDT also recognizes a third class of violations — violations of the so-called *nonprototypical* class (Turiel et al., 1991). These violations tend to be sexual in nature and feature consenting adults performing acts devoid of violence, exploitation, or compulsion behind closed doors. Perhaps the most-studied of nonprototypical transgressions is incest.<sup>5</sup> Here, we are speaking specifically of reproductively inert, exploitation-free physical intimacy between consenting adults, as typified in the following *Kissing* vignette adapted from Haidt, Koeller, and Dias (1993):

Dave and Laura are college seniors. They are also brother and sister. They are quite affectionate with each other and like to cuddle and kiss each other on the mouth. When nobody is around, they find a secret hiding place and kiss each other on the mouth passionately.

On the surface, the vignette is a moral violation par excellence. When introduced in the context of several class discussions, it tended to provoke invariably negative reactions, with the majority of students judging the siblings’ actions “wrong” and “very bad.” Indeed, Turiel et al. (1991) found that a vignette containing a depiction of non-reproductive consensual incest was judged more seriously wrong by his subjects than any other “nonprototypical” offense (including abortion, homosexuality, and pornography).

<sup>4</sup>The use of *Tickets* and *Pajamas* as instantiations of intrinsic and contingent harm, respectively, was further validated via a structured discussion with and a follow-up survey administered to a group of college undergraduates ( $n = 22$ , 14 female) in an intermediate-level psychology course. The students represented a variety of majors and cultural backgrounds. First, “contingently harmful” behaviors vs. “intrinsically harmful” behaviors were defined and illustrated. “Contingently harmful” behaviors were defined as those whose major harmful consequences are not “in the nature of the act” but are rather dependent on a particular social practice or convention. “Intrinsically harmful” behaviors were defined as those that inherently give rise to a victim or a possible victim. The students were subsequently presented (in counterbalanced order) with the texts of *Pajamas* and *Tickets* and asked to indicate if the act depicted in each vignette was “intrinsically” or “contingently” harmful. It was stressed that the acts could either belong to the same or two different categories. 22 out of 22 participants designated *Pajamas* as contingently harmful, while 21 out of 22 designated the *Tickets* as intrinsically harmful. That is, save for one individual (who thought that the acts depicted in both vignettes were only contingently harmful), all responses conformed to the expected pattern.

<sup>5</sup>Other cases in point include pornography and homosexual relations.

Notwithstanding this, “the majority of subjects [in Turiel et al.’s sample] judged that incest should be legal both in the United States (53%) and in another country (68%)” (Turiel et al., p. 53; see table 15). Moreover, even Turiel et al.’s most conservative subjects (devout high-school Catholic students) tended to treat incest as a socially contingent and nongeneralized offence — that is, they did not regard incest as continuing to be wrong under an alternative normative regime in which it was permitted. Rape, killing, and theft, on the other hand, were judged as non-contingent and generalized by virtually all (and nearly all subjects also believed that these three should be universally illegal). Moreover, in justifying their negative evaluations of incest, subjects commonly appealed to “custom and tradition” as underlying normative standards, the appeals that have been shown to be intimately tied to conventional act justifications (Turiel, 1983; Turiel, et al., 1991, p. 3; see also Southwood, 2011).

Thus, notwithstanding strong social prohibitions against sibling incest, the existing evidence suggests that consensual sibling sex belongs to a set of “nonprototypical” (Turiel et al. 1991) issues that elicit an MCDT response that is more characteristic of contingently harmful, conventional offences (a la *Pajamas*) than of their prototypically immoral (intrinsically harmful) counterparts, such as *Tickets* and *Wallet*, featuring fraud and thievery, respectively. As such, the argument of the preceding section suggests that individuals’ reactions to a vignette such as *Kissing* will yield a correlational pattern more similar to a (contingently harmful) conventional offence (i.e., *Pajamas*) than to an intrinsically harmful offense (e.g., *Tickets*).

This expectation of distinctive correlational patterns for different categories of counternormative acts highlights the important theoretical distinction between our theoretical perspective and that recently articulated by Pennycook et al. (2014). Pennycook et al. predicted that higher-scoring CRT performers may ultimately render less severe moral wrongness judgments against consensual sibling incest and sexual intercourse with a dead chicken. However, according to their account, there is “no implied necessary association between rationality and rejection of the wrongness of vignettes. Rather, the hypothesis is that greater analytic ability and style is related to a greater tendency to reflect upon the details of the vignette and hence, given that the intuitive default is of considerable wrongness, lead to . . . reductions in the judgments of wrongness.” (Pennycook et al., 2014). On this account, low and high CRT individuals alike begin with a rapid condemnation of the act, followed by some downward adjustment for the high CRT performers, but little or none for the low CRT performers. Thus, low CRT performers would be expected to remain tethered to their initial normative default, rendering

significantly more severe judgments across the board. Notably, Pennycook et al. offer no particular argument why this process should work any differently for acts of sexual deviance than for acts of thievery or physical violence.

Why, then, focus on incest and sex with dead poultry? In Pennycook et al.’s view, the importance of examining such nonprototypical offenses (Turiel et al., 1991) in the context of the study of individual differences in cognitive style resides largely in the rhetorical value of doing so: being able to show that rational reflection matters *here*, where affect and intuition seem to reign supreme, signals that it should matter even more elsewhere. That is, as Pennycook et al. point out, because these types of stimuli “elicit a strong disgust based response [, they] provide a very strong test of the role of individual differences in moral judgement because such moral vignettes should be particularly sensitive to intuitive processing and resistant to reflective processing” (Pennycook et al., 2014, p. 194). Thus, viewed in the context of the present MCDT-based research design, Pennycook et al.’s (2014) predictions diverge substantially from ours, yielding no apparent expectation of differential correlational patterns for intrinsically and contingently harmful offenses.<sup>6</sup>

## 2 Study 1

### 2.1 Method

**Participants.** Seven hundred and twenty participants located in the United States were recruited, paid, and debriefed through Amazon’s Mechanical Turk website. Participants were paid \$1.00 to complete a survey that was expected to take roughly ten minutes. Twenty-one participants failed to complete their entire survey and were excluded from analysis, leaving a final sample of  $N = 698$ . Five hundred seventy-two of these participants were randomly assigned to participate in Study 1, while the remaining 126 were assigned to participate in Study 2, which was run concurrently.

**Procedure.** In order to maintain participants’ attention, they were told at the beginning of the study that a special “code word” would appear at one point in the survey, and that they would need to enter this word at the end of the study in order to receive credit. Such a word did, in fact, appear just before the conclusion of the study.

<sup>6</sup>It is worth noting, of course, that Pennycook et al. (2014) did not engage their subjects in the moral/conventional distinction task. Thus, their data leave it open whether more reflective (higher CRT) subjects are actually less likely to view the acts in question as non-conventionally (morally) wrong or if they simply tend to view them as less wrong, with no alteration in their tendency to perceive these acts as transgressions of non-conventional kind.

An ancillary goal of Study 1 was to examine whether any effects of CRT would be moderated by the presence of incidental disgust, an emotional state that has been argued to increase the severity of moral condemnation (see, e.g., Wheatley & Haidt, 2005; Schnall, Haidt, Clore, & Jordan, 2008; Horberg, Oveis, Keltner, & Cohen, 2009). To the best of our knowledge, there has not yet been a single study, published or otherwise, that examined how feelings of disgust interact with trait cognitive reflectivity to predict moral judgments of wrong. Participants in Study 1 were randomly assigned to either experience one of two such inductions — one involving images of disgusting stimuli such as feces and vomit, and one involving ingesting a food or drink that the participant found to be very disgusting — or a control induction involving ingesting plain crackers. The participants in Study 2 received no induction and served as the control group for the participants who received the disgusting-picture induction. There were no main or interactive effects of this manipulation either on the scenarios reported here or additional scenarios used solely for the purpose of assaying the boundaries of the disgust induction, a result which we report in more detail and elaborate upon elsewhere (see Baron, Royzman, & Goodwin, 2013). Because this affect induction had no effects whatsoever (notwithstanding successful manipulations checks) and is peripheral to our primary hypothesis here, we do not discuss it further. A brief methodological description of the procedures can be found in Landy and Goodwin (2014), and additional materials/data pertaining to this aspect of the study are available from the first author upon request.<sup>7</sup>

After completing the disgust induction procedure, participants were told that they had completed the “first study” and that they would now be moving on to the “second study.” They were told that they would be reading several vignettes and making judgments regarding the behaviors described. The vignettes, all presented in a randomized order, included the three key scenarios mentioned earlier: *Pajamas* (a non-intrinsically harmful conventional violation), *Kissing* (sibling incest), and *Tickets* (an intrinsically harmful, moral violation)<sup>8</sup> as well as some additional

vignettes. One of the vignettes (*Muffin*) was designed to represent a harm-free “personal” domain act (see Turiel 1983, Turiel et al., 1991, Turiel, 2002) — an individual switching from a banana muffin to a blueberry muffin as his breakfast meal. The other, the aforementioned *Wallet*, described a poor man taking money from a wealthy man’s wallet). In this vignette, the intrinsic harm caused by taking another’s property is “balanced out” by the overall gain in utility the act produces (18 out of 18 students in a preliminary study judged that overall utility would indeed be maximized if the poor man kept the cash). We therefore included this vignette to explore whether high CRT individuals responded to it in a different way than they responded to the paradigmatically moral *Tickets* case (Turiel, 1983), in which the intrinsic harm is unmitigated. The text of the four key vignettes is presented in the Appendix.

After reading each scenario, participants rated the wrongness of each act on a 100-point scale anchored by “Not at all wrong” on the one end and “Extremely wrong” on the other. This was followed by the key MCDT probe, which asked participants to “suppose that there were foreign country A where some time ago everyone came together and decided that a behavior such as this was OK. In your view, would it be wrong or not wrong for [the protagonist’s name] to do what he/she did, assuming he/she was raised and lived in country A?” In keeping with the traditional MCDT format, this question was framed as a dichotomous “wrong”/“not wrong” choice, with “wrong” and “not wrong” responses being categorized as “moralizing” and “non-moralizing” responses, respectively.<sup>9</sup>

After responding to all vignettes, participants completed the three-item CRT and a brief demographic questionnaire, including a measure of political orientation (responses were made on a nine-point scale ranging from “Extremely conservative” to “Extremely liberal” with the midpoint labeled “Middle-of-the-road” [Jost, 2009]). Also included with the demographic questionnaire was the five-item Private Body Consciousness scale (PBC; Miller, Murphy, & Buss, 1981), a widely used measure of a tendency to attend to and make use of the feelings emanating from one’s body, which has been found in some cases to interact with moral judgment, e.g., in one study high-scoring PBC participants (but not their low-scoring counterparts) made harsher moral judgments in a filthy work area than in a clean work area (Schnall et al., 2008).

effects.

<sup>9</sup>This relativistic “other country” measure of the moral/conventional distinction goes to the heart of the theoretical construct of a “socially transcendent rule” (Turiel, 1983) and has been shown to correlate with other MCDT questions formats, e.g., those that appeal to suspension of rules (“no norm”) or nullification by consensus in one’s own country (Royzman, Leeman, and Baron, 2009, p. 170). We scored social transcendence on the basis of the answer to this question alone.

<sup>7</sup>Depending on the method of disgust-induction, additional materials included either written prompts designed to elicit gustatory disgust by having participants imagine consuming an unpleasant food substance (with eating a plain cracker as a control) or visual representations of core-disgusting substances (e.g., feces, vomit), as well two additional vignettes (Council; Wheatley & Haidt, 2005; Kitten; Schnall et al., 2008) that were deemed relevant for purposes of testing the disgust induction hypothesis but not, due to classificatory vagueness, the selective moralization hypothesis.

<sup>8</sup>In selecting *Tickets*, we rejected other paradigmatically moral, intrinsically harmful vignettes cited in Huebner et al. (2010) insofar as they contained elements of direct physical violence (e.g., a man hitting another man in the face, a man driving a car through a crowd trying to hit as many people as possible), due to (pre-tested) concerns over ceiling

Table 1:  
a. Descriptive Statistics for Study 1 (50.9% female).

	CRT	PBC	Politics	Age	Wrongness ratings			
					Tickets	Pajamas	Kissing	Wallet
Mean	1.34	21.43	5.77	31.77	93.08	57.31	82.33	76.48
Median	1	22	5	28	100	60	99	86
Mode	0	20	5	23	100	100	100	100
Std. Dev.	1.20	4.26	2.01	11.26	14.59	33.00	27.69	26.67
Cronbach's $\alpha$	0.75	0.69						
Minimum	0	9	1	17	0	0	0	0
Maximum	3	30	9	74	100	100	100	100

b. The percentage of subjects moralizing (treating as a socially transcendent offence) each act in Study 1, in order of increasing magnitude. Moralization refers to a tendency to deem a violation wrong in the context of the stipulated counterfactual reality where the violation is considered to be okay.

Scenario	Percentage of subjects moralizing
Muffin	1.0
Pajamas	19.6
Kissing	64.3
Wallet	64.5
Tickets	84.4

Table 2: Descriptive statistics for CRT in Study 1.

CRT Score	Frequency	Percentage
0	206	36.0
1	108	18.9
2	111	19.4
3	145	25.3
No response	2	0.3
Total	527	100.0

## 2.2 Results and discussion

The descriptive statistics for the key variables and the percentage of subjects who rated each scenario as a socially transcendent offence can be found in Tables 1a and 1b, respectively. The key descriptive statistics for CRT are presented in Table 2 and are similar to those obtained by Frederick (2005). The zero-order correlations between CRT, Demographics, Politics, PBC, and the degree of moralization (attribution of social transcendence) for each of the key scenarios (*Tickets*, *Pajamas*, *Kissing*, *Wallet*) are re-

ported in Table 3.

Consistent with our hypothesis, there was a significant inverse correlation between CRT and the extent to which subjects treated *Pajamas* as transcendentally wrong on the MCDT task ( $r = -0.144$ ,  $p = 0.001$ ), combined with a (non-significantly) positive association between CRT and MCDT responses on the *Tickets* vignette ( $r = 0.002$ ,  $p = 0.966$ ). The correlations were significantly different by Steiger's z test ( $p = 0.004$ , one-tailed).<sup>10</sup> Thus, consistent with one of our main hypotheses, we found that individual differences in deliberative capacity (CRT) were unrelated to the moralization of a paradigmatically moral, intrinsically harmful act, but significantly predicted a tendency to moralize a paradigmatically conventional, contingently harmful offence, with the less deliberative subjects being

<sup>10</sup>Steiger's test was used throughout to assess differences between dependent correlations. Given that the application of the method always followed the discovery of two correlations of specified magnitudes (making a null result or a particular directional result the only two possible outcomes of a comparison), a one-tailed approach was adopted to enhance power. The advantages and relative stringency of Steiger's method are analyzed in Steiger (1980) and further discussed in Meng, Rosenthal, and Rubin (1992), who also indicate the appropriateness of one-tailed tests in this context (see Meng et al. [1992], Examples 1 and 3).

Table 3: Zero-order correlations between CRT, Demographics, Political ideology, PBC, and the degree of moralization (attribution of social transcendence) for each of the four key scenarios in Study 1.

	Sex	Age	Politics	PBC	Tickets	Pajamas	Kissing	Wallet
CRT	.183***	.013	.030	-.235***	.002	-.144**	-.173***	-.056
Sex		-.268***	-.006	-.249***	.006	.036	-.080	-.025
Age			-.090*	.054	.061	.163***	.180***	.093*
Politics				.114**	-.102*	-.104*	-.184***	-.135**
PBC					.051	.056	.014	.058
Tickets						.115**	.264***	.417***
Pajamas							.165***	.136**
Kissing								.249***

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

Table 4: Regression coefficients,  $p$ -values, and odds ratios for the four key social transcendence/moralization variables (range: 0–1) as a function of CRT, with Age, Sex, Politics, and PBC as covariates.

Scenario	B	$p$ (2-tailed)	Odds ratio
Tickets	0.052	.613	1.05
Pajamas	-0.357	.000	0.71
Kissing	-0.301	.000	0.74
Wallet	-0.055	.484	0.94

more inclined to treat the act as a socially transcendent (moral-like) wrong.

Consistent with our further prediction, we found a significant inverse association between MCDT responses to *Kissing* and the CRT ( $r = -0.173$ ,  $p < 0.001$ ). This association was significantly different from that between CRT and *Tickets* ( $p = 0.0003$ , one-tailed, by Steiger’s method), but not statistically different from that between CRT and *Pajamas*. And, as would be expected, no significant association was found ( $r = 0.013$ ,  $p = 0.748$ ) between CRT and the level of moralization accorded to the paradigmatic “personal” act — a decision to switch from one breakfast food to another (*Muffin*).<sup>11</sup>

The significant inverse association between CRT and *Kissing* (and CRT and *Pajamas*), accompanied by a non-significantly positive association between CRT and *Tick-*

*ets*, is clearly in concert with our general hypothesis. However, as argued earlier, it is also consistent with a set of potentially valid third-variable accounts. While a list comprising all such third-variable candidates would be unmanageably (perhaps, infinitely) long, we identified three candidate variables — a person’s political ideology, his/her sex, and his/her sensitivity to gut feelings — as being of particular theoretical interest and concern. To explore these possible alternatives for the conjoint variation in CRT and MCDT, we ran a series of logistic regressions with *Kissing* and *Pajamas* social transcendence ratings as our target variables and Age, Sex, Politics (scored [1–9 scale] in the direction of greater liberalism), and PBC as covariates. As can be seen from Table 4, with all four covariates in place, CRT remains a significant inverse predictor of *Kissing* and *Pajamas* (but not of *Tickets* or *Wallet*).<sup>12</sup> With two exceptions (Politics and Sex), none of the covariates were significant predictors of the target variables. Politics was a significant inverse predictor of all four variables [Odds Ratios ranging between 0.889 and 0.832;  $ps$  ranging between  $< 0.001$  and 0.03], while females tended to be significantly more moralizing about *Pajamas* than males [Odds Ratio = 1.954,  $p = 0.006$ ]. Indeed, putting the results of the regression to the side, none of the four aforementioned variables could have conceivably mediated the link between moralization and CRT since, as indicated in Table 3, neither Age nor Politics correlated with CRT, while neither PBC nor Sex correlated with any of the target variables.

On the whole, the results are quite straightforward and fall readily in line with the hypothesized pattern. Yet, some lingering questions remain.

<sup>11</sup>Mirroring the aforementioned pattern of selective association between CRT and attributions of social transcendence for *Pajamas* and *Kissing*, CRT was significantly and inversely associated with the wrongness ratings for *Pajamas* and *Kissing*, but not with the wrongness ratings for *Tickets*. Zero-order correlations between CRT and wrongness ratings for *Tickets*, *Pajamas*, and *Kissing* were 0.032 ( $p = 0.448$ ),  $-0.218$  ( $p < 0.001$ ), and  $-0.179$  ( $p < 0.001$ ), respectively.

<sup>12</sup>CRT was also a significant inverse predictor of the wrongness ratings for *Kissing* and *Pajamas* (standardized beta =  $-0.123$ ,  $p = 0.004$  and standardized beta =  $-0.188$ ,  $p < 0.001$ , respectively), but not of the wrongness ratings for *Tickets*, with all the same covariates in place.



First and foremost, it may be asked whether, based on the evidence available from this study, we should not conclude that the more reflective subjects tend to deliver a more “relativist” (“not wrong in another country”) answer with respect to *Kissing* and *Pajamas* simply because they see these acts as less serious infractions even under existing conditions. If so, this may cause such individuals to judge these transgressions as “not wrong” when faced with making a dichotomous choice for the key MCDT question. To address this, we re-ran the aforementioned logistic regression models with *Kissing* wrongness ratings and *Pajamas* wrongness ratings as additional covariates. As would be expected, the effects were somewhat weaker but remained quite strong (Odds Ratio = 0.766 (vs. 0.742 in the original regression,  $p = 0.004$ ) for *Kissing*, and Odds Ratio = 0.777, (vs. .711,  $p = 0.017$ ) for *Pajamas*, when wrongness ratings were accounted for statistically. Thus, it appears that, consistent with our general proposal, high CRT individuals’ tendency *not* to treat *Kissing* and *Pajamas* as genuinely moral (socially transcendent) offenses cannot be fully accounted for by how severely wrong individuals judge these offenses to be under existing normative conditions. (And, again, it is possible that the small reduction in the CRT effect can be explained by an effect of perceived social transcendence on the wrongness judgments themselves; an act may be considered less wrong if it would not be wrong at all somewhere else.) This result, combined with the other disambiguating findings, is broadly congruent with our original conjecture that, at least in part, CRT’s relation to moralization results from high CRT individuals’ capacity to distinguish between contingently and intrinsically harmful acts, treating only the latter as genuinely moral (socially transcendent) offenses.

Second, we should note that, though the results displayed in Table 3 bear out our main theoretical predictions, they also signal some other intriguing correlational patterns that we did not anticipate. One concerns the uniform tendency for political liberals to be less likely to view the counter-normative behaviors of interest (*Tickets*, *Pajamas*, *Kissing*, and *Wallet*) as transcendentally wrong. The crucial difference between this correlational pattern and that found for the “hyper-reflective” (high CRT) subjects is that, while the high CRT subjects exhibited their “relativistic” orientation in a “selectively amoral” fashion (going easy on the practitioners of illicit sex and dress-code rebels, but being as tough as the low-CRT subjects on acts of thievery and fraud), liberals tended to be more “uniformly amoral”, exhibiting a more relativistic normative orientation across-the-board. Similarly, the liberally oriented subjects tended to rate *all* behaviors (including theft and fraud) as significantly less wrong under existing normative conditions.

## 3 Study 2

### 3.1 Method

**Participants.** As described above, 126 participants located in the United States were randomly assigned to complete the Study 2 procedure rather than the Study 1 procedure, though the two studies were run concurrently.

**Procedure.** As in Study 1, participants were told at the beginning of the study that a special “code word” would appear at one point in the survey, and that they would need to enter this word at the end of the study in order to receive credit. Such a word did appear just before the conclusion of the study.

Participants in Study 2 did not receive any of the disgust inductions included in Study 1, and immediately performed the MCDT with respect to the five vignettes described above. They subsequently completed a number of items that Study 2 shared with Study 1. These included the CRT, PBC, and the demographic questionnaire. They next completed a filler task, followed by two novel items. One new instrument was a 6-item two-facet measure of conservative/liberal orientation (Kemmelmeir, 2008), which included three items concerning evaluative attitudes toward homosexuality and abortion; both issues were investigated by Turiel et al. alongside incest for their 1991 monograph and represent defining sexual morality issues of America’s cultural right (Baumgartner, Francia, Morris & Scavo, 2008). For each item, the subjects indicated their agreement/disagreement on a 5-point scale (varying from “Strongly disagree” to “Strongly agree”). The scale items, given below, were presented to the subjects in a randomized order and scored in the direction of greater conservatism (R = reverse scored): 1. *The federal government should do more to control the sale of handguns* (anti-regulatory; R); 2. *Colleges should prohibit racist/sexist speech on campus* (anti-regulatory; R); 3. *Wealthy people should pay a larger share of taxes than they do now* (anti-regulatory; R); 4. *It is important to have laws prohibiting homosexual relations* (social); 5. *Same sex couples should have the right to legal marital status* (social; R); 6. *Abortion should be legal* (social; R). Kemmelmeir’s original principal component analysis uncovered two factors that accounted for 55.6% of the overall variance, with the social conservatism factor being signaled by high factor loadings of sexual morality items (see items 4, 5, 6) and the second factor characterized by high factor loadings of “small government”/“anti-regulatory” items (items 1,2,3) (see Kemmelmeir, 2008). Our follow-up principal axis factoring analysis with oblique rotation (oblimin) and 0.5 as the conservative cut-off for rotated factor loadings largely confirmed the two-factor solution; the two factors accounted for 63.6% of the variance, with the social con-

servatism factor being signaled by high factor loadings of all three sexual morality items and the second factor characterized by high factor loadings of two of the three “small government”/“anti-regulatory” items. Consistent with the measure’s reported two-factor structure, the social conservatism composite ( $\alpha = 0.783$ ) and the anti-regulatory conservatism composite ( $\alpha = 0.451$ ) were not significantly correlated with each other ( $r = .12, p = 0.18$ ). The second new measure assessed participants’ level of educational attainment, in years of formal schooling. The new measures were included to rule out potential third-variable explanations of any association between CRT and moralization that could not be ruled out in Study 1.

### 3.2 Results

The analyses reported here were designed to further examine the relative potency of CRT as a predictor of sibling incest moralization (Royzman et al., 2008; Royzman et al., 2009; Royzman, et al., 2011) by pitting the predictive powers of CRT against a more nuanced, sexual morality-centric measure of conservative/liberal orientation and a measure of educational attainment (see Method for details). The three social conservatism items of the conservatism scale (Kemmelmeier, 2008) concerned judgments about abortion and homosexuality, widely believed to be the defining issues of America’s culture wars (with opposition to both said to be strongly rooted in Christian biblical beliefs) (Baumgartner et al., 2008).

The composite scores for each factor/subscale were formed by adding the three individual items. As reported above, consistent with the measure’s posited two-factor structure, the social conservatism composite and the anti-regulatory conservatism composite were not significantly correlated with each other ( $r = .12, p = 0.18$ ). Because of its low internal consistency and because it was conceptually unrelated to our main interest, we excluded the anti-regulatory composite from further consideration.

Table 5 lists the results of a logistic regression with *Kissing* (social transcendence) as the dependent variable and CRT, Age, Gender, PBC, and the two new items (Social Conservatism and Education) as covariates.

We entered all of the available covariates into a stepwise logistic regression (using the method of conditional backward elimination based on Wald’s test). Only two variables were retained in the final model, Social Conservatism ( $B = 0.175$ , odds ratio = 1.191,  $p = 0.023$ ) and CRT ( $B = -0.564$ , odds ratio = 0.569,  $p = 0.001$ ), with CRT being the stronger of the two predictors.<sup>13</sup> Thus, CRT once

again emerges as a predictor of *Kissing* after the other variables of interest have been accounted for statistically, thus replicating the pattern observed in Study 1. We should further note that, as in Study 1, the association between CRT and *Kissing* transcendence was not fully accounted for by the *Kissing* wrongness ratings when the latter was included as a covariate ( $B = -0.467$ ; odds ratio = 0.627;  $p = 0.034$ ).

The rest of the pattern was similar to that reported in Study 1, with one exception. Relatively few participants (13.5 percent) in this sample judged that the *Pajamas* scenario would still be wrong under a different normative regime, which may explain the lack of any significant association between CRT and the *Pajamas* transcendence rating ( $B = -0.163$ ; odds ratio = 0.850;  $p = 0.489$ , accounting for Sex, Age, and Education). The remainder of the associations were as expected, including a significant inverse association between CRT and *Pajamas* wrongness ratings (standardized beta =  $-.198, p = .04$ ) accompanied by no significant association between CRT and the wrongness/transcendence ratings for either of the intrinsic harm scenarios.

## 4 General discussion

The main thrust of the present research was to join prior work in social domain theory with a recently validated and popular measure of individual differences in deliberative capacity (CRT) to test the hypothesis that an individual’s propensity towards deliberative thought would be significantly associated with a tendency to treat non-intrinsically harmful, but socially inappropriate acts (wearing pajamas to work, engaging in acts of sexual intimacy with a sibling) as matters of convention rather than morality proper, but that no such association would be found between deliberative capacity and a tendency to moralize acts that are intrinsically harmful to others.

To test this hypothesis, we asked a large and ideologically diverse sample of American adults to complete the CRT along with a streamlined version of the MCDT (moral-conventional distinction task) featuring, among others, a prototypical conventional issue (*Pajamas*), a widely discussed nonprototypical issue (*Kissing*), and two vignettes involving prototypically harmful acts against others. We found that, while high CRT performance was significantly associated with a tendency to treat contingently harmful transgressions as arbitrary societal constructions, it bore no relation to the tendency to moralize, or not moralize, acts that are intrinsically harmful to others. Moreover, this pattern remained in-

<sup>13</sup>The zero-order correlations between CRT and Tickets and CRT and *Kissing* were 0.124 (n.s.) and  $-0.312$  ( $p < 0.001$ ), respectively. The correlations were significantly different by Steiger’s method ( $p = 0.0001$ , one-tailed). We also computed correlations between CRT and Tickets/Wallet composites and CRT and *Kissing*/*Pajamas* composites. These

were  $-0.034$  (n.s.) and  $-0.234$  ( $p = 0.008$ , one-tailed), respectively. The correlations were significantly different by Steiger’s method ( $p = 0.02$ , one-tailed).

Table 5: Social transcendence (range: 0–1) for Kissing in Study 2 as a function of Sex, Age, PBC, Social Conservatism, Years of Education (Education), and CRT.

Covariate	B	<i>p</i> (2-tailed)	Odds ratio	Pearson’s <i>r</i>	<i>p</i> (2-tailed)
Sex	0.098	.836	1.11	−0.09	0.274
Age	0.019	.341	1.02	0.11	0.213
PBC	0.022	.662	1.02	0.17	0.052
Social Conservatism	0.181	.026	1.19	0.25	0.005
Education	0.050	.537	1.05	0.00	1.000
CRT	−0.560	.004	0.57	−0.31	0.000

*Note.* Politics was excluded due to concerns about multicollinearity with Social Conservatism; entering the two predictors simultaneously rendered both non-significant, with the CRT level of significance remaining unchanged ( $p = 0.004$ ). Social Conservatism was retained as the more morally pertinent of the two.

tact when taking into account political orientation, sensitivity to gut feelings, gender, age, and, in Study 2, educational attainment and a more nuanced measure of social conservatism centered on sexual morality. Nor was the CRT-moralization link fully explained by subjects’ antecedent ratings of the acts’ permissibility. This indicates that CRT’s predictive relation to our key measure (one that no previous study of cognitive style, including Pennycook et al.’s [2014] has tapped) — whether a counter-normative act will be represented as a “genuinely moral” (socially transcendent) rather than merely conventional offense — does not merely reflect a normative spillover from subjects’ reactions to the wrongness probe.

Our pattern of findings may offer a new way to illuminate a very old question: what accounts for the apparently vast differences in moral opinion across time and place? (To see just how old the question is, one need only refer to Herodotus’s famed account [the Histories, Book III, section 38] of King Darius’s futile attempts to persuade some of his subjects to forfeit their customs for the disposal of the dead in favor of their neighbors’.)

In much recent research, apparent group differences in the moralization of actions have been taken as evidence for the existence of different underlying moral systems that are not shared or may be differentially activated across groups (see, for instance Shweder et al., 1987, 1997; Jensen, 1997, 1998; Graham, Haidt, & Nosek, 2009). One commonly cited case in point is the priestly town of Bhubaneswar, where it is judged *morally* impermissible for a widow to eat fish. At first blush, this seems like a case of fundamentally divergent values par excellence, with the “traditional” Indian informants subscribing to a code of ethics (later dubbed the “Divinity Code” [Schweder et al., 1997]) that their liberal Western counterparts either do not possess at all or possess in a radically attenuated form. However, this interpretation of the

evidence has been questioned (Turiel, Killen, & Helwig, 1987). For, as Shweder’s own ethnographic work indicates (see Turiel et al., 1987 for an overview), there is a viable alternative account of the phenomenon: according to the local beliefs, fish acts as a sexual stimulant, increasing the likelihood that the widow will engage in sexual congress, which her husband’s ever-watchful spirit could witness (and presumably agonize over) from its celestial abode. On this interpretation, the ban on eating fish appears to be but another example of a societal attempt to regulate preventable harm. Similar “informational assumptions” (Turiel et al., 1991) about the fabric of the cosmos, the mind, and, especially, the reciprocal relationship between material and spiritual realms, may apparently account for a wide range of other seemingly indisputable cases of fundamental “moral disagreement” (Brandt, 1954).

Our findings may elucidate this issue by revealing a third and hitherto unexplored reason why different individuals or groups might moralize the same behaviors to varying degrees: the difference may be less a matter of *what* they think than of *how* they think, i.e., their ability (and/or willingness) to counterfactualize the existing normative regime on demand. Different cognitive styles could produce different patterns of moralization, leading, per Westermarck (1906), to recognizably distinct normative profiles, even while holding constant the key value positions and the underlying factual assumptions. Residents of Bhubaneswar might simply be unable (or unwilling) to posit an alternate reality in which widows can properly enjoy a plate of cod or in which a child may wear a non-traditional item of clothing to a priestly ceremony.<sup>14</sup>

<sup>14</sup>Indeed, later field work in India (Madden, 1992) has shown that, compared to the more educated and Westernized Indian college students residing in the same area — whose performance on MCDT was very similar to that of their Western counterparts — the majority of Hindu priests hold an explicit commitment to not even momentarily counterfactualize

Our demonstration that different cognitive styles are associated with different patterns of issue moralization also has some clear methodological implications. For one, it should give pause to any automatic inference from the evidence of an apparent difference in cultural (Shweder et al., 1987), socioeconomic (Haidt et al., 1993), or ideological (e.g., Jensen, 1997) patterns of moral evaluation to a presumption of fundamentally divergent values as the underlying cause (Graham et al. 2009). The differences could simply result from distinctive cognitive styles, which affect how individuals understand and apply their *shared* values to facts on the ground. In fact, the overall pattern we report is highly reminiscent of that found in one previous study (where no controls for cognitive style or cognitive ability were used [Haidt et al., 1993]) in which members of economically disadvantaged, relatively uneducated groups were shown to be significantly more likely than their better educated high-SES counterparts to blur the distinction between morality and convention, with the lower SES group treating “harmless” nonprototypical offenses, including incest, as moral absolutes. Interestingly, the two SES groups were, however, identical in their condemnation of the one and only truly prototypical moral offense included in the stimulus set: a child pushing another child off a swing. Given the increasingly well-documented association between poverty and impaired cognitive performance, as measured both by indices of fluid intelligence and CRT-like tasks requiring individuals to override an intuitively salient response to perform well (see Mani, Mullainathan, Shafir, & Zhao, 2013 for evidence and discussion), the current analysis strongly indicates that the SES-moralization link found in Haidt et al.’s (1993) study would have been at least partly mediated by measures of cognitive style and cognitive ability had such measures been used, suggesting that the moral-conventional distinction gap between Haidt et al.’s high- and low-SES participants is at least as likely to be a product of the low-SES participants’ cognitive limitations as it is of high-SES participants’ purportedly limited access to the full range of foundational values, whether these are framed in terms of Shweder and colleagues’ original “Big Three” (Shweder, Much, Mahapatra, & Park, 1997) or Haidt and colleagues’ more recently proffered “Big five” (Graham et al., 2009).

One further novel finding to emerge from the present research is the striking divergence between the uniform correlational pattern found between liberalism and MCDT response, and the more selective pattern found between the latter and the CRT; while the higher scoring CRT subjects tended to be “less moralistic” in a category-specific way (“going easy” on the practitioners of illicit sex and dress-code violators, but being as “tough” as their low-CRT counterparts on acts of thievery and fraud), liber-

als tended to be more “uniformly amoral”, engaging in less moralization across-the-board. Similarly, the liberally oriented subjects tended to rate *all* behaviors (including theft and fraud) as significantly less wrong. While it is clearly beyond the scope of this paper to adduce a convincing explanation for this effect, we hazard that it partly results from the fact that those judging from within a distinctly liberal mindset may view all deviant acts (including mass fraud) as a shared responsibility between the individual transgressor and the “inauspicious” social environment within which he or she is situationally embedded, while more conservative individuals see the responsibility as resting squarely on the shoulders of the deviant actor alone.

A final and, perhaps, most basic implication of our results is that, in conjunction with the pioneering work of Paxton et al. (2012), Pennycook et al. (2014) and others, they underscore the foundational importance of deliberative thought to the process of moral evaluation. Moreover, the present research adds to both of the aforementioned studies by indicating that, beyond affecting the reported amplitude of one’s moral disapproval, proneness to rational deliberation may play a unique role in setting the boundaries of the moral domain as such, informing one’s very construal of where the realm of convention ends and that of morality proper begins. For people lacking the requisite measure of cognitive sophistication (be they the Brahmins of Orissa<sup>15</sup> [Shweder et al., 1987] or lower-SES Brazilian children [see Haidt, 2012, p.21]) or the will to posit cosmological realities different from one’s own (see Madden, 1993), the line between the two normative domains may be blurred to the point of non-existence.<sup>16</sup>

What does this all mean? It has been widely argued in recent moral judgment literature — not to mention historically — that, by and large, reason “can aspire to no higher office” than that of the resourceful and steadfast, but ultimately clueless spinmeister of the passions: an inveter-

<sup>15</sup>Subsequent research (Madden, 1993; Jensen, 1998) demonstrated that the issues separating Shweder’s Brahmin subjects from his Hyde Park subjects were roughly the same issues that separated conservative and progressive individuals within a given culture, with both Madden (1993) and Jensen (1998) showing that moral judgments of progressive Indians were more similar to those of their progressive American counterparts than to the conservative participants within the Indian sample. In either case, the progressive subjects were substantially better educated than their conservative counterparts. Indeed, Jensen notes that, when it comes to her Indian subjects, “the educational level of the two groups differed markedly. Eighty percent of the orthodox [i.e., conservative] participants had fewer than 11 years of education whereas 90% of the progressivist participants had 16 years or more of education” (p. 95). Moreover, as Jensen (1997) remarks, in analyzing differences between these groups, “education was not entered as a covariate” (p. 96).

<sup>16</sup>As Haidt (2012) points out, “Shweder found almost no trace of social conventional thinking in the sociocentric culture of Orissa” (p. 17), while he and his collaborators (Haidt et al., 1993) found that “the working-class kids [in Recife, Brazil] judged the uniform rebel in exactly the same way they judged the swing-pusher” (p. 21).

the cosmological beliefs grounding their faith.

ate after-the-fact explainer and justifier with little to no involvement in the judgment process itself. The studies reported herein suggest otherwise. They suggest that, aside from likely playing a key role in adjudicating intrapsychic normative disputes (Paxton et al., 2012; Rozman et al., 2011), reason is in fact an active and powerful contributor to the construction of the moral domain itself, the position that David Hume himself held dear some 250 years ago.

## References

- Blair, R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition*, *57*, 1–29.
- Baumgartner, J., Francia, P. L., Morris, J. S., and Scavo, C. (2008). Is it Really Red Versus Blue? Politics, Religion, and the Culture War Within. *American Review of Politics*, *29*, 1–18.
- Brandt, R. B. (1954). *Hopi ethics, a theoretical analysis*. Chicago, IL: University of Chicago Press.
- Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579.
- Baron, J., Rozman, E., & Goodwin, G. (2013). *Is disgust a moralizing emotion?* Unpublished manuscript. University of Pennsylvania.
- Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making*, *3*, 121–139.
- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, *4*, 20–33.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 24–42.
- 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., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*, 2105–2108.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.
- Haidt, J., Bjorklund, F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason*. Unpublished manuscript.
- 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.
- Harman, G., Mason, K., & Sinnott-Armstrong, W. (2010). Moral reasoning. In John Doris and the Moral Psychology Research Group (Eds.), *The Oxford Handbook of Moral Psychology* (pp. 206–245). New York: Oxford University Press.
- Hollos, M., Leis, P., & Turiel, E. (1986). Social reasoning in Ijo children and adolescents in Nigerian communities. *Journal of Cross-Cultural Psychology*, *17*, 352–76.
- Huebner, B., Lee, J., & Hauser, M. D. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture*, *10*, 1–26.
- Hume, D. (1983). *An enquiry concerning the principles of morals*. J. B. Schneewind (Ed.). New York: Hackett. (Original work published 1751).
- Jensen, A. L. (1997). Different worldviews, different morals: America's culture war divide. *Human Development*, *40*, 325–344.
- Jensen, A. L. (1998). Moral divisions within countries between orthodoxy and progressivism: India and the United States. *Journal for the Scientific Study of Religion*, *37*, 90–107.
- Jost, J. T. (2009). "Elective affinities": On the psychological bases of left-right differences. *Psychological Inquiry*, *20*, 129–141.
- Kant, I. (1959). *Foundations of the metaphysics of morals* (L. W. Beck, Trans.). Indianapolis: Bobbs-Merrill. (Original work published 1785).
- Kemmelmeier, M. (2008). Is there a relationship between political orientation and cognitive ability? A test of three hypotheses in two studies. *Personality and Individual Differences*, *45*, 767–772.
- Landy, J. F. & Goodwin, G. P. (2014). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. Manuscript submitted for publication.
- 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 of London. Series B, Biological Sciences*, *270*, 819–826.
- Madden, T. M. (1992). *Cultural context and information assumptions as sources of variation in social reasoning in India*. (Unpublished doctoral dissertation). University of California-Berkeley, Berkeley, CA.
- Mellers, B. A., Unger, L., Baron, J., Ramos, J., Gurcay, B., Fincher, J., Scott, S., Moore, D., Atanasov, P., & Tetlock, P. E. (2014). *Improving Geopolitical Forecasting*. Manuscript in preparation. University of Pennsylvania.
- Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*, 976–980.
- Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172–175.

- Miller, L. C., Murphy, R., Buss, A. H. (1981). Consciousness of body: Public and private. *Journal of Personality and Social Psychology*, *41*, 397–406.
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*, 221–236.
- Nucci, L. (2001). *Education in the moral domain*. New York: Cambridge University Press.
- Oechssler, J., Roeder, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, *72*, 147–152.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–177.
- Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, *123*, 335–346.
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J. & Fugelsang, J. A. (2014). The role of analytic thinking in moral judgments, moral values, and religious disbelief. *Thinking & Reasoning*, *20*, 188–214.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Royzman, E. B., Goodwin, G. P., & Leeman R. F. (2011). When sentimental rules collide: “Norms with feelings” in the dilemmatic context. *Cognition*, *121*, 101–114.
- Royzman, E. B., Landy, J. F., & Leeman, R. F. (in press). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science*.
- Royzman, E. B., Leeman, R., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, *112*, 159–174.
- Saunders, L. F. (2013). What is moral reasoning? *Philosophical Psychology*. <http://dx.doi.org/10.1080/09515089.2013.801007>.
- Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*, 1096–1109.
- 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.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The “Big Three” of morality (autonomy, community, divinity) and the “Big Three” explanations of suffering. In A. Brandt & P. Rozin (Eds.), *Morality and health*. (pp. 119-169). New York, NY: Routledge.
- Singer, P. (2005). Ethics and intuitions. *Journal of Ethics*, *9*, 331–352.
- 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.
- Southwood, N. (2011). The moral/conventional distinction. *Mind*, *120*, 761–802.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245–251.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275–1289.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge: Cambridge University Press.
- Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Inconsistency and consistency in reasoning about social issues. *Monographs of the Society for Research in Child Development*, *56*, Serial No. 224.
- 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. (2002). *The culture of morality: social development, context, and conflict*. New York: Cambridge University Press.
- Wheatley, T., & Haidt, J. (2005). Hypnotically induced disgust makes moral judgments more severe. *Psychological Science*, *16*, 780–784.
- Westermarck, E. (1906). *The origin and development of the moral ideas*. London: McMillan.
- Westermarck, E. (1921). *The history of human marriage* (5<sup>th</sup> Ed.). London: McMillan.
- Yau, J., & Smetana, J. (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, *74*, 647–658.

## Appendix

### Full texts of four key vignettes used in Studies 1 and 2

**Pajamas.** Richard wakes up late on Monday and he can’t decide what to wear to work. After realizing that he doesn’t like any of his pants, he just decides to wear his silk pajamas to the office.

**Incest.** Dave and Laura are college seniors. They are also brother and sister. They are quite affectionate with each other and like to cuddle and kiss each other on the

mouth. When nobody is around, they find a secret hiding place and kiss each other on the mouth passionately.

**Tickets.** Tim, a college freshman, has found a way to counterfeit tickets for a local concert venue. Although he knows that they will not get people in to concerts, he sells them to various people at inflated prices just to make money. Everyone who bought the tickets comes to the concert only to realize that the tickets are fake.

**Wallet.** Sam is walking down the street one day when he comes across a wallet lying on the ground. He opens the wallet and finds that it contains several hundred dollars in cash as well the owner's driver's license. From the credit cards and other items in the wallet it's very clear that the wallet's owner is wealthy. Sam, on the other hand, has been hit by hard times recently and could really use some funds. So he decides to send the wallet back to the owner without the cash, keeping the cash for himself.

### **Sample questions, from the Tickets scenario**

How wrong is it for Tim to sell the counterfeit tickets just to make money?

Suppose there were foreign country A where some time ago everyone came together and decided that a behavior such as this was OK. In your view, would it be wrong or not wrong for Tim to do what he did, assuming he was raised and lived in country A?