Judgment and Decision Making, Vol. 10, No. 4, July 2015, pp. 296–313

The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect

Edward B. Royzman* Kwanwoo Kim† Robert F. Leeman‡

Abstract

The paper critically reexamines the well-known "Julie and Mark" vignette, a stylized account of two college-age siblings opting to engage in protected sex while vacationing abroad (e.g., Haidt, 2001). Since its inception, the story has been viewed as a rhetorically powerful validation of Hume's "sentimentalist" dictum that moral judgments are not rationally deduced but arise directly from feelings of pleasure or displeasure (e.g., disgust). People's typical reactions to the vignette are alleged to support this view by demonstrating that individuals are prone to become *morally dumbfounded* (Haidt, 2001; Haidt, Bjorklund, & Murphy, 2000), i.e., they tend to "stubbornly" maintain their disapproval of the act without supporting reasons. In what follows, we critically reassess the traditional account, predicated on the notion that, among other things, most subjects simply fail to be convinced that the siblings' actions are truly harm-free, thus having excellent reasons to disapprove of these acts. In line with this critique, 3 studies found that subjects 1) tended *not* to believe that the siblings' actions were in fact harmless; 2) notwithstanding that, and in spite of holding a number of "counterargument-immune" reasons, subjects could be effectively maneuvered into exhibiting all the trademark signs of a morally dumbfounded state (which they subsequently recanted), and 3) with subjects' beliefs about harm and standards of normative evaluation properly factored in, a more rigorous assessment procedure yielded a dumbfounding estimate of about 0. Based on these and related results, we contend that subjects' reactions are wholly in line with the rationalist model of moral judgment and that their use in support of claims of moral arationalism should be reevaluated.

Keywords: incest, moral dumbfounding, moral judgment, disgust, rational, emotion, reason.

1 Introduction

Cassie and Bernie are officemates. One day, in honor of their second week anniversary working together, Bernie presents Cassie with a can of imported wild-caught tuna in lightly sweetened Ponzu sauce. While duly appreciative of the gesture, Cassie politely declines the offer, reminding Bernie that she is committed to consuming only sustainably harvested dolphin-free tuna and that Bernie's can, splendid as it may be, is lacking the discernibly marked dolphin-free label. Bernie retorts that, having anticipated Cassie's concerns, he had thoroughly researched the brand and can avow that the tuna housed within *this* can is sustainably harvested Skipjack. Since dolphins do not associate with Skipjack, this tuna is dolphin-free by default. Cassie seems to comprehend Bernie's reasons, but remains steadfast in her refusal to welcome the gift.

Is she being unreasonable? The hallmark of reason, after all, is *sensitivity to reasons*. On the other hand, not any old reason will suffice. For instance, the matter of assessing Cassie's reasonableness would be greatly muddled if it turned out that she had an unstated rule against accepting gifts from officemates or some deep-seated doubt about the quality of Bernie's on-line research skills. *Suppose*, however, it could be ascertained that Cassie's ethical reservations are *solely* a function of her worries over tuna's dolphin-free pedigree; suppose we could further establish that Cassie shares the full range of Bernie's empirical beliefs and, generally speaking, trusts his judgment completely and unequivocally. *Under these circumstances*, Cassie's continued refusal to take the can ("because it just doesn't feel right") could be rightfully construed as a case of ethical fetishism at its finest, a reliable indicator that we are dealing with someone whose ethical thought process has genuinely strayed from the path of rational discourse. In the theoretical idiom of the moment, Cassie would appear to be *morally dumbfounded* to the hilt.

According to Haidt, Bjorklund and Murphy (2000), *moral dumbfounding* (MD) refers to "the stubborn and puzzled maintenance of a moral judgment *without supporting rea-*

An earlier version of this paper was presented at the March 2015 EPA symposium on moral judgment and emotion and benefited greatly from the participants' comments. We are especially grateful to Jonathan Baron, Philip Dunwoody, Johannes C. Eichstaedt, Geoffrey Goodwin, Yoel Inbar, Justin Landy, Chaz Lively, and Paul Rozin, with additional thanks to Xuan Gao, Daniel Jacobson, Matt Ruby, and Sydney Scott for their counsel and support along the way.

*Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Solomon Lab Building, Philadelphia, PA 19104. Email: royzman@psych.upenn.edu.

†University of Pennsylvania

‡Yale University School of Medicine

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

sons" (Haidt et al., 2000, p. 1, emphasis added) (Haidt, 2001; see also Haidt, Koller & Dias [1993]). Originally reported by Haidt et al. (2000), MD has been featured prominently in Haidt's (2001) influential "Emotional Dog and its Rational Tail" (see Pizarro & Bloom, 2003, for an early analysis and critique), where it is iconically illustrated via the "Julie and Mark" vignette (a.k.a., *Incest*), a sly and epigrammatic tale of sibling love and family vacation gone awry (Haidt, 2001, p. 814):

> Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?

Though only a decade and a half old, Incest has risen to become a fixture in psycho-philosophical debates on the role of reason and passion in moral cognition (Huebner, 2011; Jacobson, 2013; Pinker, 2002; Singer, 2005), commanding levels of attention previously reserved for the likes of Kohlberg's "Heinz" (Colby & Kohlberg, 1987) and Thomson's "Footbridge" (Thomson, 1986) (see also Greene, 2013). Like the latter it has been viewed as a rhetorically powerful validation of Hume's sentimentalist dictum that, akin to judgments of taste, moral assessments are not logically deduced from higher-order beliefs (e.g., "Causing interpersonal harm is wrong", "This is interpersonal ham", "This is wrong"), but arise directly from a feeling of pleasure or displeasure at the object in hand:[1] "So that when you pronounce any action or character to be vicious, you mean nothing, but that from the constitution of your nature you have a feeling or sentiment ... from the contemplation of it" (Hume, 1739–1740/1978, p. 469; see also Hume, 1739–1740/1978, p. 471).

The aim of the paper is to critically reexamine what "Julie and Mark" (and others of its ilk) has to tell us about moral cognition in general and its ties to reason in particular. We begin by reviewing some key aspects of the story and the findings that sealed its repute. We then proceed to report a series of studies that pit our deflationary alternative against its well-established counterpart—*the moral dumbfounding narrative.*

[1]Throughout this paper, we accept a view of Hume' moral philosophy that is extremely common in empirical moral psychology, but that almost certainly fails to capture the full complexity of Hume's moral-philosophical ideas (Hume, 1739–1740/1978) and their evolution in later works (e.g., Hume, 1751/1983)

1.1 The moral dumbfounding narrative

Perhaps, the most celebrated aspect of the "Julie and Mark" vignette is its alleged freedom from harm. As Haidt and colleagues (2000) contend, the story "was carefully written *to be harmless*...[so that] the participant would be prevented from finding the usual 'reasoning-why' about harm that participants in Western cultures commonly use to justify moral condemnation" (Haidt et al., 2000, p. 8, emphasis added). The participants rendering a negative evaluation of Julie and Mark's activities were thereupon questioned by a "devil's advocate" instructed to push back against the initial disapproval of the act by calling attention to various *harm-negating provisos* embedded within the narrative: "For example... if the participant responded that what the person or persons in the story did was wrong, the main counter argument was that no harm was done, and that the fact that an act is disgusting does not make it wrong" (Haidt et al., 2000, p. 9).

What Haidt and colleagues (2000) seemed to have found was nothing short of remarkable:

> Most people who hear the above story immediately say that it was wrong for the siblings to make love, and they then begin searching for reasons (Haidt, Bjorklund & Murphy, 2000). They point out the dangers of inbreeding, only to remember that Julie and Mark used two forms of birth control. They argue that Julie and Mark will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them. Eventually, many people say something like, "I don't know, I can't explain it, I just know it's wrong" (Haidt, 2001, p. 814).

Elsewhere, Haidt uses the image of rummaging for an object in one's pockets and coming up empty-handed as a metaphor for moral dumbfounding as a state defined by lack of all and any discernable reasons to support the moral evaluation that one supports:

> The most common reasons involve genetic abnormalities or that it will somehow damage their relationship. But we say in the story that they use two forms of birth control, and we say in the story that they keep that night as a special secret and that it makes them even closer... And it's only when they reach deep into their pockets for another reason, and come up empty-handed, that they enter the state we call "moral dumbfounding." ...They're *surprised* when they don't find reasons [to support their on-going disapproval]... So it's a cognitive state where you "know" that something is morally wrong, but... can't find reasons to justify your belief... [So] you just say:

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

> "I don't know, I can't explain it, I just know it's wrong" (Sommers, 2009, pp. 155–156).

Aside from some commonly referenced indicants of MD—bouts of confusion/disorientation, "unsupported declarations" (alleging that the act was "just" or "plain" wrong, e.g., "It's just wrong to do that!" [Haidt et al., 2000, p. 12]), a tendency to drop arguments, and, ultimately, the declaration of dumbfounding itself—one prominent feature of Haidt et al.'s (2000) results is the sheer *prevalence* of subjects' reluctance to reverse themselves in light of countervailing reasons. According to Table 1 (Haidt et al., 2000), only 20% of the subjects initially stated that Julie and Mark's actions were Ok. By the end of the session this number increased to 32%, suggesting a moral dumbfounding estimate of 68%.

Where is all the perseverance coming from? Haidt et al (2000) and Haidt (2001) offer a typically Humean reply: when encountering Incest "one feels a quick flash of revulsion... and one knows intuitively that something is wrong" (Haidt, 2001, p. 814). Thus, ultimately, Haidt's *moral dumbfounding narrative* appears to be comprised of two mutually supportive counterparts: the sentimentalist claim that *subjects' judgment against Julie and Mark's dalliance is a direct result of subjects' physical revulsion at the act* and the kindred claim that *subjects' inability to offer **any** subjectively warrantable reason* (i.e., a reason that would be warranted in light of existing normative traditions and that makes sense in the mind of the person advancing it) *in support of their disapproval of the act is unlikely to have any discernable impact on their readiness to give up the disapproval as such,* amounting to a conspicuous breach of the rationalist credo that one "should not hold a judgment in the absence of reasons" (Haidt et al., 2000, p. 6). Theoretically speaking, these two components fit remarkably well, for clear and reasonable as the devil's advocate's appeals might have been, they could hardly been expected to undo, ameliorate, or even finesse the primordial revoltingness of the act (Royzman, Leeman & Sabini, 2008; Royzman, Atanasov, Landy, Parks & Gepty, 2014).

1.2 Critique of the moral dumbfounding narrative

Though the moral dumbfounding narrative seems to offer a reasonably attractive and internally coherent account of MD, one of its key components has been recently called into doubt (Royzman, Leeman & Baron, 2009; Royzman, Goodwin & Leeman, 2011). Using Haidt's (2001) original vignette and a trait measure of disgust sensitivity (DS; Haidt, McCauley & Rozin, 1994), Royzman et al. (2009) found no significant association between individual differences in trait disgust and individual tendencies to moralize Julie and Mark's behavior. At the same time, incest moralization *was* significantly predicted by perceived harm (see also Gray, Schein & Ward, 2014; Turiel, 2002) after taking into account disgust sensitivity, sex, and age, and subjects' sibling status, with a number of subjects directly commenting on the difficulty imagining how the siblings' relationship would remain unaffected in the aftermath of the act (see Huebner, 2011, p. 58 for comparable anecdotal reports of disbelief from some of his students and his conclusion that the "credulity" of Haidt's subjects must have been seriously strained).

Haidt's (2001) own report indicates that a substantial number of subjects initially grounded their condemnation of Incest in appeals to relational harm. Haidt's standard construal of these appeals (Haidt, 2001; Haidt, 2012; Haidt et al., 2000; Sommers, 2009, pp. 155–156) as mere signs of confusion or justificatory despair slights the fact that people routinely anchor fictional content in real-world knowledge, finding it difficult to comprehend information about a fictional universe that contradicts their real-world assumptions (Ferguson & Sanford, 2008; Ferguson, Scheepers & Sanford, 2010) (this appears to be the case even if the key fantastical event [e.g., cats eating carrots] has been set against the backdrop of a fittingly fantastical universe [e.g., cats are vegetarians] [Ferguson & Sanford, 2008]). In the special case of Incest, the failure to accept the "lived happily ever after"proviso is not particularly surprising given the universally dim view of incest as carrying "significant non-biological costs" (Shor & Simchai, 2009, p. 1834) and jeopardizing "both the integrity of the family as a whole and [subjects'] own ability to maintain regular family relationships" (Shor & Simchai, 2009, p. 1834).

It is true, of course, that both the experimenter's appeals and the harm-negating provisos within the vignette were expressly framed to coax all negative real-world preconceptions to the side. However, as discussed elsewhere (Royzman, Cassidy & Baron, 2003), there is now a large and methodologically diverse body of evidence to suggest that individuals are only marginally effective at discounting their prior ideas or beliefs. As argued elsewhere, this epistemtic egocentism (Royzman et al., 2003) or *curse of knowledge* (Camerer, Loewenstein & Weber, 1989) is a robust feature of human cognition and has been found both in children and adults. For example, Baron and Hershey (1988) reported a series of tightly controlled experiments demonstrating that the "privileged" outcome information (the information that subjects were normatively required to set to the side) significantly affected their ratings of a person's decision quality, the finding analogous to the "knew it all along" corollary of the hindsight bias (Fischhoff, 1975) (see also Baron, 2008). In an another important study, Anderson, Lepper and Ross (1980) presented subjects with a set of hypothetical cases suggesting either a positive or negative relationship between risk taking and success as a firefighter. The reputed evidence for this link was then "totally discredited" via a debriefing

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

session. According to Anderson et al. (1980), the debriefing session had only a "minimal" impact on the subjects' subsequent judgments, which were made as if the once stipulated link between risk-taking and being a successful firefighter was still in effect.

Indeed, as mentioned above, Haidt's (2001) own depiction of the results indicates that an unspecified number of subjects did appeal to the likelihood of relational harm early on in the procedure, but had their appeals overruled by the pre-programmed reminder that "no harm was done" (Haidt et al., 2000, p. 9).

Haidt et al. (2000) do not report the exact wording they employed, but it is a reasonable conjecture that being told (in whatever terms) to "try again" as one's initial response failed to take into account the harm-free nature of the act would amount to an implied request to frame all subsequent answers under the assumption that all and any harmful consequences of the siblings' actions have been forestalled, thus rendering any further reference to harm conversationally otiose. A subject continuing to express his or her incredulity beyond this point would not only run the risk appearing uncooperative (see Norenzayan & Schwarz, 1999 on how, "in an attempt to be cooperative communicators, subjects actively monitor and try to provide information tailored to the researchers' interests"), slow, and uncouth (see Bonnefon, Feeney & De Neys, 2011 on politeness as an obstacle to effective communication) (Goffman, 1955), but would also find themselves swimming against two of the mightiest currents in the psychology of social influence—a tendency to defer to the epistemic position of the "man in charge" (Milgram, 1974) and a tendency to pay lip service to the judgments of one's peers, even when these are patently at odds with the evidence of one's senses (e.g., Asch, 1956) (Sabini, 1995).

One other noteworthy complication in Haidt et al.'s (2000) approach is their unstated assumption that *were* subjects to reason their way from a higher-order principle to a case-specific judgment in accordance with the rational deductive model ("It is wrong to do X; this a case of X; this is wrong") the relevant higher-order principle would *need* to be comprised of some variant of the "no harm, no foul" rule. While Haidt et al. do not communicate this point directly, it can be logically inferred from the study's core methodological conceit, i.e., the belief that subjects' ability to retrieve and adduce any subjectively warrantable reasons in support of their judgment of wrong should be quite effectively neutralized via the narrative proviso that the "customary" implications of the intra-familial sex will simply fail to materialize in this particular case.

Yet, as Jacobson (2013) pointed out, Incest and other scenarios of its kind could be condemned from virtually every conceivable normative standpoint within the Western philosophical tradition, including deontology, virtue ethics, and rule-utilitarianism (see Royzman, Landy & Leeman, 2015). Indeed, in some of our previous studies (Royzman et al., 2008; Royzman et al., 2009; Royzman et al., 2011), verbal and written appeals to the likelihood of emotional harm were regularly co-mingled with appeals to the basic counternormative nature of the act ("It is inherently wrong", "Because you are not supposed to have sex with a relative", "Because of the incest taboo") as well as unappealing character traits ("impulsive", "irresponsible"). And, as Taylor and Wolfram (1968) observed some 45 years ago, at the deeper, "foundational" (Kagan, 1998) level of analysis (see Footnote 1), an individual's inherent commitment to, say, telling the truth (Kant, 1785/1959) may be grounded in the view that "the world is so arranged that" telling the truth [or not bedding one's next of kin] "ultimately works out to the general good, whether or not this is clear to the agent or not" (Taylor & Wolfram, 1968, p. 243). For subjects hailing from one of these "alien" normative positions (lay deontology, lay virtue ethics, lay rule-utilitarianism), the study's continued emphasis on *realized* harm as the only legitimate basis for ethical assessment may have spelled further normative disorientation, leading them, willy-nilly, to affirm that they did not in fact have any sound "arguments" to adduce.[2]

1.3 Overview of the hypotheses

The present studies were designed to address four main hypotheses (along with a set of sub-hypotheses). **First**, we anticipated that, being mindful of incest's real-world implications, subjects would reject some (though not necessarily all) of the story's harm-negating provisos, including the key stipulation that Julie and Mark's decision to have sex would leave their relationship unscathed. **Second**, we anticipated that subjects' incredulity regarding this and related aspects of the narrative would remain intact following a detailed counterargument, even as subjects went on to exhibit all the trademark signs of a morally dumbfounded state, including confusion, apparent non-responsiveness to reasons and the declaration of dumbfounding itself.[3] These predic-

[2]We note that similar points apply to "Cadaver", the second (and, in our view, considerably more problematic) of the two moral cognition narratives used by Haidt et al., 2000). In an odd twist, the story features a cannibalistically inclined vegetarian lab assistant who decides to take home and consume a piece of the cadaver placed in the assistant's care. In this case, the added complicating factor is the stated and unstated "metaphysical" beliefs regarding the continuity of psychological functioning after death, the beliefs that college students who hold them might find too "juvenile" to express. The phenomenon has been extensively documented by Bering (2006). Most pertinently, Bering (2002) found that, among undergraduates asked to assess the psychological states of a protagonist who had just experienced a sudden death in car crash, even subjects who subsequently categorized themselves as "extinctivists" (i.e., those who endorsed the view that the conscious self ceases permanently with the death of the body) acknowledged that, at some level, the dead person *knew* that he was dead and thus could potentially be a subject of good or bad treatment from others (see also Rozin & Stellar, 2009).

[3]Comments voiced by subjects in our previous studies (Royzman et al., 2008; Royzman et al., 2009) indicate that they were largely in agreement

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

tions were examined in Studies 1 and 2, respectively. **Third**, we hypothesized that, with credulity and other relevant considerations properly factored in, physical disgust would no longer be a significant predictor of subjects' disapproval of the act and, **last**, that, as the more conceptually stringent criteria for the diagnosis of MD proper are applied, the phenomenon would turn out to be either entirely non-existent or highly irregular, at best.

2 Study 1: The credulity check

2.1 Method

2.1.1 Subjects

Twenty four undergraduates (nine female; M_{age} = 21.96, *SD* = 4.55, *median* = 20) enrolled in a seminar-style Social Psychology course took part in the study in exchange for extra credit. Subjects completed the task during a class break. The time commitment (including debriefing) was 3–5 minutes.

2.1.2 Materials and procedure

The survey consisted of Haidt's (2001) original Incest story[4] (sans the normative judgment probe) followed by five questions. Subjects were asked to read the vignette carefully and respond at their own pace. No time pressure was exerted.

The first four questions (each rated on a 0-to-100 scale, with 0 indicating "Not believable at all" and 100—"100 percent believable") were: "Given the facts of the story, how believable do you find that Julie and Mark will honor their decision not to have sexual relations ever again?" (*Abstain*); "Given the facts of the story, how believable do you find that Julie and Mark will keep what happened between them a secret?" (*Secret*); "Given the facts of the story, how believable do you find that Julie and Mark's having sex with each other will not negatively affect the quality of their relationship or how they feel about each other later on?" (*Relationship*); "Given the facts of the story, how believable do you find that Julie and Mark's having sex with each other will have no bad consequences for them personally and/or for those close to them?" (*Consequences*). Additionally, subjects were asked to speculate on what (if any) effect Julie and Mark's decision to have sex "would have on their lives in the real world" (*Real world*). The three response options (the first two counterbalanced for order) were: "It would have a negative effect" (coded as −1), "It would have a positive effect" (coded as +1), and "It would have no effect either way" (coded as 0). Subjects were also asked to rate their level of confidence in their judgment (0 = Not confident at all; 100 = Extremely confident). The confidence-adjusted ratings of "real world consequences" (*Real world*) were then computed by multiplying subjects' categorical judgments (−1, 0, +1) by their stated confidence in these judgments.

The survey began with two items (Abstinence, Secret) that were expected to garner relatively high believability ratings (with the relative ordering determined at random), while the Real world item, expected to elicit a very "negative" appraisal (and one that could bias all subsequent responses in the direction of lower credulity ratings), was always presented last.

2.2 Results and discussion

Means and 95% CIs for each of the four credulity probes are displayed in Table 1a. Minimum believability ratings by number of subjects collapsed across the four credulity probes are presented in Table 1b. In line with Haidt et al.'s expectations (2000), subjects were largely willing to accept that the siblings would keep their sexual encounter a secret. On the other hand, subjects were generally inclined to reject the harm-negating provisos assessed by Relationship and Consequences, while remaining slightly less certain about the siblings' prospects for not repeating the act in the future.

As Table 1b indicates, the highest minimum believability score for any given subject was only in the 30s (on a 0 to 100 scale), with Relationship and Consequences being the two main drivers of skepticism (see the mean and lowest believability ratings in Table 1a). Lastly, subjects seemed to be generally of the opinion that the real world consequences of Julie and Mark's actions would be quite severe. The mean confidence-adjusted rating for Real world was −68.33 (*SD* = 31.39), 95 % CI [−81.59, −55.07], significantly below 0 (t [23] = −10.66, $p < 0.001$).[5]

All in all, the study results were strongly in line with our prior expectation (Royzman et al., 2009) that a substantial proportion of college-age adults would find it difficult to accept the reputedly "harmless" events of Incest as being truly and credibly harm-free. The purpose of our next study was to ask whether subjects' incredulity would remain intact following a series of targeted counterarguments modeled after those employed by Haidt et al. (2000).

with Jacobson's (2013) point that the emphasis on various means of contraception is a venerable "red herring: a salient but irrelevant point that distracts from the real issue", the real peril being "of course that Julie and Mark will do irreparable harm to their relationship as siblings" (Jacobson, 2013, p. 301). Thus subjects' perception of the effectiveness of contraception received relatively little attention in our studies.

[4]Here and henceforth, the sentence "At the very least it would be a new experience for each of them" was omitted to preclude (in line with some pilot subjects' comments) the impression that Julie and Mark's incestuous encounter marked their initiation into sexual intimacy as such.

[5]We note a significant zero-order correlations between subjects' confidence-adjusted Real world ratings and their believability ratings for Relationship (r = 0.44; p = 0.03) and Consequences (r = 0.63, p = 0.001), respectively, with more negative real-world expectations translating into greater reluctance to accept that Julie and Mark's relationship would remain as unscathed.

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

Table 1a: Mean believability ratings, 95% confidence intervals, and number of subjects who gave their lowest believability rating for Secret, Abstain, Relationship, and Consequences in Study 1.

Credulity probe type	Mean	95% C.I.	Number of subjects who gave their lowest believability rating*
Secret	83.00	72.32–93.67	1
Abstain	37.00	25.53–48.47	5
Relationship	24.75	11.56–37.93	12
Consequences	20.00	9.22–30.77	17

Note: All mean ratings were significantly above/below the scale's midpoint (the point of uncertainty) (alpha = 0.05).

* Overall frequency is greater than 24 because 10 subjects gave the same lowest believability rating to 2 or 3 credulity probes.

Table 1b: Minimum believability ratings by number of subjects collapsed across the four credulity probes.

Score range	N
0-10	16
11-20	2
21-30	2
31-40	4
41-100	0

Note: Mean of subjects' minimum believability rating collapsed across credulity probe: 11.88, SD = 14.45.

Absent such a demonstration, Haidt and colleagues could justifiably assert that, while a certain measure of disbelief was an integral part of the subjects' initial response, it was precisely the counterargument's job to lay any such doubts to rest, further citing their subjects' tendency to give up (or, at least, not to reaffirm) their harm-based reasons as prima facie evidence that the devil's advocate's counterclaims worked just as intended.

3 Study 2: Manufacturing unreason

3.1 Method

3.1.1 Subjects

Twenty-eight undergraduates (19 female; M_{age} = 21.64, *SD* = 2.49, *median age* = 21) enrolled in a seminar-style psychology course (Judgment and Decisions) took part in the study in exchange for extra credit. Subjects were screened for prior knowledge of the vignette. Four subjects reported having seen the story before as part of a class survey and/or as a test item, but reported no knowledge of the underlying theoretical claims or related empirical results. Thus, their data were retained in the sample.

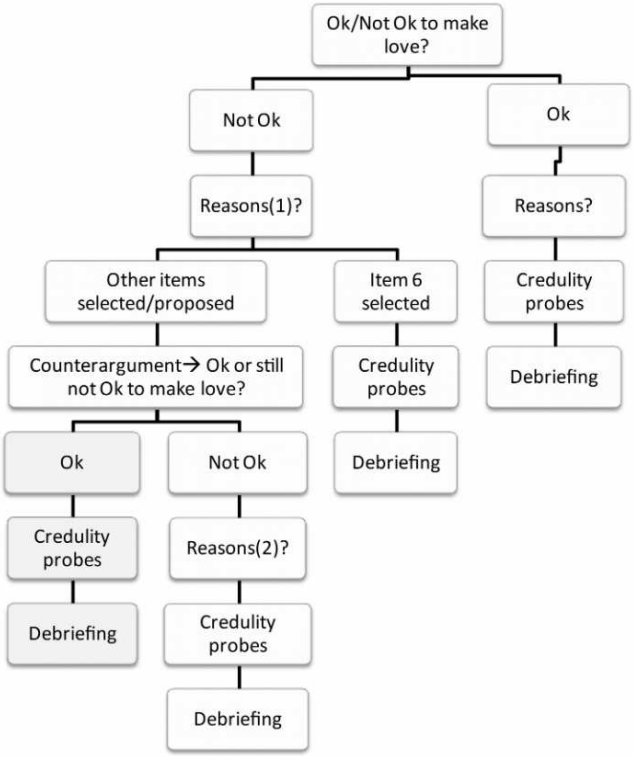
3.1.2 Materials and procedure

Subjects were interviewed individually. With some deliberate exceptions (see below), the protocol was modeled after that in Haidt et al. (2000). All subjects were told that they would hear a story that they might or might not find morally objectionable. They were asked to make a judgment about the events it described. Subjects were told that, once they gave their judgment, the experimenter would play "devil's advocate" by questioning their reasons and that the subject was free to respond in any manner that they saw fit. After informing the subjects of their right to withdraw from the study and obtaining their consent to proceed, the experimenter read a slightly modified version of Incest, then asked subjects to indicate whether, in their personal opinion, it was "Ok for Julie and Mark to make love?" (with "Yes, it was ok" and "No, it was not ok" as the two response options). The version used in this study was identical to that used in Study 1 except for the next-to-last sentence, which read: "*They both enjoy making love and have no regrets about it, but they decide not to do it again.*" The "no regrets" proviso was added to render the story even more "harm-proof" and to bring it in line with the text of the counterargument that followed, which was partly modeled after that cited in Piazza & Sousa (2014). The rest of the procedure varied considerably depending on the subject's answer to the initial evaluative probe (see Figure 1 for the diagrammatic overview). Subjects were also monitored for signs of confusion (e.g., the "self-doubt face" described in detail in Haidt et al., 2000, p. 13) and other non-verbal cues.

Subjects who did not object to the siblings' actions were asked to confirm their answer, then to give a reason or reasons for the judgment they made. They were then directed to the final page of the booklet containing the two credulity items detailed below. Subjects who disapproved of the siblings' actions were first asked to confirm their answer, then *to cite a reason or reasons "supporting [their] judgment that it was not Ok for Julie and Mark to make love.*" To enable accurate accounting of changes in reasons offered by subjects through the course of the study, each subject was provided with an "experimental booklet" that contained a list of 5 harm-based reasons generated based on prior pilot work as well as previously published results (Haidt et al., 2000; Royzman et al., 2008; Royzman et al., 2009; Royzman et al., 2011). The five putative reasons were: "*1. Because it will harm them emotionally/psychologically."; "2.*

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

Figure 1: Diagrammatic overview of the interview protocol in Study 2.



Note: "Reasons(1)?" describes the interviewer's initial request for reasons following the original judgment of "not Ok"; "Reasons(2)?" describes the interviewer's second request for reasons, following the judgment of "not Ok" in the response to the counterargument. Since all subjects maintained the judgment of "not Ok" following the counterargument, the shaded area represents a path of inquiry that was not taken with any subject within this study.

Because it will harm those close to them."; "3. Because it could have harmed them emotionally/psychologically."; "4. Because it could have harmed those close to them."; "5. Because of the dangers of inbreeding." Subjects were told that they were free to nominate all five reasons, none of the reasons, or some combination of reasons ("for example, you can say '1' and '4' "). While this feature of the study diminished its viability as a direct replication of Haidt et al. (2000), it actually enhanced its viability as a conceptual replication of their procedure. The current procedure afforded us a quantitatively precise measure of the "reasons dropped" variable, while further conducing to the study's objective of determining whether subjects' readiness to disavow harm-related reasons of different types (leading up to the all-important declaration of dumbfounding) is compatible with their continued representation of the siblings' actions in a harm-laden manner.

Moreover, the present approach allowed us to make a key conceptual distinction between those harm-related considerations that appeal solely to actual or *realized* harm (e.g., Reasons 1, 2) versus those that appeal mainly to the destructive *potential* of the act (e.g., Reasons 3, 4), with only the former being potentially affected, i.e., neutralized or markedly attenuated, by the "no harm done" argument alone. (By analogy, few would want to dispute that an individual grounding her condemnation of a knowingly reckless act, e.g., driving while under the influence, in the destructive potential of that act would be fully within her rights to keep both her disapproval *and* her reasons even after being informed that, on this particular occasion, "no harm" was incurred.) Subjects were also asked to cite any additional reason or reasons of their own, which we expected to be largely deontological in nature (i.e., appeals to existing norms) and thus also *counterargument-immune*.

Finally subjects were told that if they felt that *none* of the listed reasons applied and that they *also* could find no good reason or reasons of their own, they might consider endorsing option 6: *"6. I don't have a strong reason at this point, but I just feel it's wrong for them to do what they did."*

With the interviewee's responses reviewed and confirmed, the following standardized statement was recited verbatim: "Ok, I will now enact my role as 'devil's advocate' by calling your attention to some aspects of the story… For example, the story makes it clear that Julie and Mark's actions were fully consensual and they both had no regrets about what they did. They also took every conceivable precaution. They used two forms of birth control. As a result, there was no risk of pregnancy. Also, the act was conducted in private, was kept private, and they both enjoyed it. They also never did it again. In short, it's a given that no harm has occurred or will occur." The subject was then asked: "Do you still think Julie and Mark's action of having sex with one another was not Ok?" (with "*Yes, I still think it was not ok" and "No, I now think that it was ok" as the two response options).* It was expected that, as in the original study, most, if not all, subjects would retain their prior evaluation of the act. After confirming subjects' second (post-counterargument) evaluative judgment, the experimenter instructed them to turn to the booklet's next-to-last page. The page contained the same menu of harm-based reasons that subjects referred to earlier in the study; once again, subjects were asked to cite any "reason or reasons" for disapproving of the act, while bearing in mind that, as the story makes it clear, "no harm has occurred or will occur." We anticipated that this simple locution would be markedly effective in restricting subjects' expression of harm-based reasons, causing them to become "morally dumbfounded" in the end. Subjects were then asked to turn to the last page of their booklet containing a two-item credulity probe.

Item one (Relationship) read: "Having read the story and considering the arguments I presented, are you able to **be-**

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

lieve that Julie and Mark's having sex with each other will not negatively affect the quality of their relationship or how they feel about each other later on?" Item two (Consequences) read: "Having read the story and considering the argument I presented, are you able to **believe** that Julie and Mark's having sex with each other will have no bad consequences for them personally and/or for those close to them?" The response options consisted of "Yes (I am able to believe)" and "No (I am not able to believe)." Subjects responding with a "No" to either credulity probe were prompted to elaborate in their own words.

In addition, subjects responding with a "No" to the Relationship probe were asked whether they considered the expected relational damage to be a form of psychological harm.

Finally, as part of the debriefing process, a subset of subjects judging the siblings' actions to be "not Ok" were queried about evident inconsistencies between their responses to the credulity items (which we expected to be characterized by a strong belief that the siblings' actions will have strongly negative consequences for all concerned) and their tendency to rescind previously endorsed harm-based reasons. It was a priori determined that the use of this inconsistency probe would be contingent on the subject's either (a) endorsing the "declaration of dumbfounding" item 6 (initially or following the counterargument) *and* responding to at least one of the credulity probes in the direction of disbelief (indicating that one did not "buy" that the siblings or others would not be harmed) *or* (b) disavowing all harm-related reasons following the counterargument *and* responding to at least one of the credulity probes in the direction of disbelief. Depending on the specifics of the subject's response, some additional exploratory questions were posed. Subjects were then informed about the rationale for the study, thanked for their participation, and asked if they had any further comments.

3.2 Results and discussion

The key descriptive statistics are given in Table 2a.

As expected, the vast majority of subjects (21 out of 28 or 75%) disapproved of the siblings' actions ($p = 0.01$ by the binomial test), with not a single respondent reversing his or her judgment following the counterargument. Also, as expected, there was a substantial difference in the average number of listed reasons cited before and after the counterargument ($M_{Before} = 2.04$, $SD = 1.24$, median = 2; $M_{After} = 0.28$, $SD = 0.64$, median = 0). The difference was statistically significant by a paired t-test: $t\ (20) = 5.72$, $p < 0.001$. Intriguingly, this pattern remained largely intact ($M_{Before} = 0.95$, $SD = 0.86$, median = 1; $M_{After} = 0.14$, $SD = 0.47$, median = 0; $t\ (20) = 3.44$, $p = 0.003$) after the comparison was limited to a subclass of counterargument-immune reasons (Reason items 3 and 4), those that appealed solely to the destructive *potential* of the act, without any consideration for its actual results.

Proportions of subjects who endorsed each of the five "listed reasons" are given in Table 2b. Some subjects (n = 10) also offered additional reasons of their own, all centered on the counternormative nature of the act—with the majority of statements (6 out of 10) initially taking the form of "unsupported declarations" (e.g., "It's immoral", "It is morally wrong") (Haidt et al., 2000). Once subjects were prompted to elaborate, all six declarations were "unpacked" into what could be construed (based on the invocation of norms or codes of conduct) as logically coherent deontological claims (with a given subject stating, for example, that, in his view of things, incest was inherently immoral and, given that this is what the siblings did, their actions were also immoral).

Crucially, the majority of those citing counterargument-immune reasons (n = 17) went on to disavow one or more of these reasons (15/17 or 88%) following the counterargument, with 13 out of 17 (76%) moving on to endorse the "declaration of dumbfounding" option 6. (Most strikingly, the majority [n = 7] of subjects citing "deontological" reasons during the first half of the interview [n = 10], the reasons that subjects themselves chose to put forth as something supplementary to harm-related considerations, declared themselves dumbfounded shortly following the experimenter's assertion that "that no harm has occurred or will occur").[6]

Having been largely successful in replicating Haidt et al.'s (2000) original effect, we now turn to the all-important question of whether subjects' overwhelming endorsement of item 6 (the declaration of dumbfounding) toward the tail end of the study may be construed as an accurate reflection of their genuine acceptance of Incest as a harm-free event. The results suggest otherwise (see Table 2a, Note **): in line with our expectations, all but two subjects reported incredulity regarding lack of harm related to both Relationship and Consequences, with the remaining two reporting incredulity regarding Consequences only. (During the debriefing, subjects tended to re-affirm their "yes" and "no" answers by reiterating their belief that the siblings' relationship would be negatively affected in the end, while occasionally citing what could be construed as rule- and character-based considerations as further reasons for their disapproval of the act). As expected, lack of credulity regarding the harm-free nature of the act and disapproval of the siblings' act were strongly and positively associated (see Table 2c for details) ($p < 0.001$ by Fisher's exact test).

The final phase of the study was designed to explore subjects' own take on the apparent inconsistency between their inclination to impute harm and their observed tendency to

[6]Subjects also tended to display the "self-doubt face" (essentially, sustained frowns) detailed by Haidt et al. (2000, p. 13) and made verbal remarks indicative of confusion, e.g., "This is confusing."

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

Table 2a: Key descriptives for Study 2.

Total number of subjects	28
Subjects who thought the act was not Ok	21
Subjects who thought the act was not Ok and were exposed to the counterargument	19
Subjects who reversed their judgment following the counterargument	0
Subjects who thought the act was not Ok and dropped one or more prior reasons following the counterargument	17
Subjects who thought the act was not Ok and endorsed "a declaration of dumbfounding"*	15
Subjects who thought the act was not Ok and failed to accept the harm-negating provisos**	21
Subjects who thought the act was not Ok and offered counterargument-immune reasons***	17
Subjects who thought the act was not Ok and cited deontological reasons	10
Subjects in the above category who made a declaration of dumbfounding following the counterargument	7
Total number of subjects with supporting reasons****	21
Subjects whose responses warranted the inconsistency probe (see Method for details)*****	17

* This includes 13 subjects who made their declaration of dumbfounding (item 6) following the counterargument and 2 additional subjects whose declarations preceded the counterargument (resulting in the fact that only 19 of 21subjects heard the counterargument and had a chance to change their views in its wake).

** The count represents 21 individuals who indicated a lack of beliefs on both of the credulity probes toward the tail end of the study (with 19 out 21 reporting incredulity regarding Relationship and 21 out of 21 reporting incredulity regarding Consequences [both $ps < 0.001$ by the binomial test], with all incredulous subjects further indicating that they considered the likely negative effect on the siblings' relationship to be a form of psychological harm).

*** For present purposes, counterargument-immune reasons were those comprised of (1) appeals to deontological considerations: rules/inherent immorality of the act and (2) appeals to the harm-inducing *potential* of the act (see items 3 and 4 from the "reasons" menu).

**** The "supporting reasons" count is comprised of all the subjects with counterargument-immune reasons as well as any subject who maintained his/her belief in the harmful implications of the siblings' actions following the counterargument (as assessed by the credulity probes).

***** The subjects in question, all exhibiting a configuration of response tendencies that met the a priori conditions for the application of the inconsistency probe specified in Method , included (a) 15 subjects (the majority) who dropped all harm-based reasons *and* endorsed the "declaration of dumbfounding" item 6 while also responding to at least one of the credulity items in the direction of disbelief (i.e., indicating that they did not "buy" that the siblings or others would not be harmed) and (b) 2 subjects who dropped all of their harm-based reasons and responded to both of the credulity items in the direction of disbelief.

Table 2b: Proportions (and counts) of subjects endorsing each of the five listed reasons for why Julie and Mark's actions were not Ok (in order of descending frequency).

"Because it could have harmed them emotionally/psychologically"	57.1%	(12/21)
"Because it will harm them emotionally/psychologically"	47.6%	(10/21)
"Because it could have harmed those close to them"	38.1%	(8/21)
"Because of the dangers of inbreeding"	38.1%	(8/21)
"Because it will harm those close to them"	23.8%	(5/21)

disavow all or most harm-related reasons that they initially endorsed. With their attention called to the fact, *all* subjects in question (17 out of 21 interviewees) (see Table 2, Note ***** for details) acknowledged that their prior disavowal of harm-related reasons, including and especially those informed by the destructive *potential* of the act, was unjustified. While 6 out of 19 (31 %) seemed unable to account for the inconsistency (e.g., "I am not sure", "I was confused"), the remaining majority tended to state that they said what they said because they felt pressured to and/or inferred that

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

Table 2c: Relationships between subjects' approval/disapproval of the act and their acceptance of the harm-negating provisos in Study 2.

	No	Yes
Accepted lack of harm with respect to relationship?		
Disapproval of act	19	2
Non-disapproval of act	1	6
Accepted lack of harm with respect to individual consequences?		
Disapproval of act	21	0
Non-disapproval of act	5	2

they were required to respond "under the assumption" that no harm has occurred or will occur.

Since we failed to anticipate the full extent of subjects' tendency to disavow their norm-based reasons following the counterargument, the interview protocol had no specific provisions in that regard. However, the issue was broached on an ad-hoc basis during the debriefing session, leading us to conclude that laying stress on harm-negating considerations during and after the counterargument phase was what caused some subjects to judge or infer that non-harm-related reasons were conversationally "irrelevant", just as harm-related reasons were conversationally "proscribed".

All in all, this pattern of results indicates that, while the interviewing procedure had hardly any discernable effect on what subjects were willing *to believe*, it had a very substantial effect on what they were willing *to express*. On the whole, the procedure appears to have rather serious limitations as a means of assessing the presence of a morally dumbfounded state as it has been formally defined (Haidt et al., 2000), being evidently unable to discriminate between the cases in which the criterial features of the moral dumbfounding response (judgment *without supporting reasons*) are genuinely met from those in which they only appear to be met (supporting reasons are abundant but remain *unexpressed*).[7]

4 Study 3: Will the truly morally dumbfounded please stand up!

4.1 Method

4.1.1 Subjects

53 undergraduates (32 female)[8] enrolled in two concurrent sections of a seminar-style psychology course (Judgment and Decisions) took part in the study. Subjects were compensated with extra credit.

4.1.2 Materials and Procedure

The primary study materials consisted of three surveys (completed by all of the 53 subjects involved in the study). The surveys were administered at three different points in time over the course of a semester. The first and second surveys (containing the normative judgment probe and the credulity probe, respectively) were administered four weeks apart. The second and the final survey were administered two weeks apart. These intertemporal delays offered several advantages, including reduced likelihood of post hoc justification, reduced reactivity, and more manageable survey administration time.

The first survey included the original (Study-1) version of the "Julie and Mark" vignette (Haidt, 2001), followed by an evaluative judgment probe taken verbatim from Haidt (2001): "Was it Ok for Julie and Mark to make love?" (p. 814) (with "*Yes, it was ok" and "No, it was not ok" as the two response options).* Subjects were then asked to say "why" they responded as they did.

The second survey consisted of two parts. In Part 1, subjects were asked to read a series of statements, then to **select one that they "identified with" most or saw as** being "most consistent" with their view "on how **a person** may appropriately reason about his/her negative evaluation of an act." The first two statements were designed to sort subjects into two broad normative orientation camps: those who endorsed the no-harm-no-foul orientation and those who did not. The statement designed to convey the no-harm-no-foul orientation consisted of a claim that "violating an established moral norm just for fun or personal enjoyment is wrong only in situations where someone is harmed as a result, but is acceptable otherwise." The alternative stipulated that "violating an established moral norm just for fun or personal enjoyment is inherently wrong even in situations where no one is harmed as a result." The statements were counterbalanced for order. It was verbally underscored that the key distinction is between believing that acts that violate a moral norm are wrong only if they result in harm

[7] In this regard, a declaration of dumbfounding may be viewed as something akin to a false confession, with a psychologist rather than a detective "smoothing out" the process (see Benforado, 2015 on the commonality of psychologically induced false confessions in the present-day criminal justice system).

[8] Age information was not collected from this sample based on a request from some of the older subjects. We estimate the age range for the majority within this sample to be comparable to that reported in Studies 1 and 2, i.e. 18–22 years of age, with 6 additional individuals in their 30s and 40s.

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

and the view that acts that violate a moral norm are wrong even if they do not result in harm. Subjects were also presented with a third statement designed to serve as an attention check. Subjects who expressed an affinity for the "inherently wrong" position were then asked to describe a further reason for endorsing the normative position that they endorsed. This additional probe was inspired by the work of Shelly Kagan (1998) (see also Taylor and Wolfram, 1968) who speculates that reason-giving may operate at two different levels, with a given case-specific judgment of wrong ("It was wrong for Mark to break his promise to Paul") being commonly grounded in a pertinent intermediate-level rule (e.g., "Breaking promises is wrong"), the level at which many ordinary people's reason-giving is thought to operate (Harman, 2010; Kagan, 1998), which may, in turn, be grounded in the more foundational rule-consequentialist considerations, e.g., consideration of utility to all concerned if the collectively advantageous practice of promise-keeping were upheld.

The resultant statements were coded for evidence of consequential reasoning (see below). Part 2 of Survey 2 was designed to assess subjects' acceptance of the story's harm-negating provisos. To that end, subjects were presented again with the Incest vignette followed by two credulity probes (Relationship and Consequences) similar to those used in Study 2. The questions were counterbalanced for order and were followed by two response options: "Yes, I am able to believe this" and "No, I am not able to believe this."[9]

The third and final survey consisted of a set of items designed to check on alternative interpretations of the findings. The survey began with two standardized trait measures aimed at establishing if any hypothesized associations between the Survey 1 and Survey 2 variables could be explained in terms of social desirability or/and a desire to respond in a psychologically consistent manner (included in the Appendix): a 10-item social desirability scale (**MC-1**, Strahan & Gerbasi, 1972; see, e.g., Bartels and Pizarro, 2011 for prior use) and a brief 9-item version of Preference for Consistency Scale (Cialdini, Trost & Newsom, 1995). Subjects also reported their level of state disgust in response to the "Julie and Mark" vignette using a 5-point scale. In line with previous research, state disgust was assessed via the Oral Inhibition index (henceforth, OI) (see Royzman et al., 2008; Royzman et al., 2014).[10] Subjects were asked to rate their political orientation on a 7-point scale, with "1" signifying "Very Conservative", "7"—"Very Liberal" and "4"—"middle-of-the-road". Finally, subjects were also asked to indicate if they have encountered the Incest vignette before and, if so, under what circumstances. Subjects were then informed that the three surveys were all part of the same project and asked to pen down their best guess as to the project's overarching goal. They were then thanked and fully debriefed.

4.1.3 Interviews

To determine the actual incidence of moral dumbfounding within our sample, a set of "fully convergent" subjects who had previously rendered a negative evaluation of the siblings' actions were interviewed roughly midway between the administrations of Surveys 2 and 3. A subject was deemed "fully convergent" if and only if he/she both (1) endorsed the no-harm-no-foul orientation in Part 1 of Survey 3 and (2) responded affirmatively to both of the credulity probes (indicated that, in his/her view, Julie and Mark's actions caused no harm). Further details of the interview protocol and its findings are discussed below.

4.2 Results and discussion

The key descriptive statistics are displayed in Table 3a. A series of exploratory analyses confirmed that all non-categorical variables met the assumption of normality.

Correlational analyses (Table 3b) revealed significant associations between the evaluative response: incest permissibility (Ok/not-Ok to make love) and the following six variables: Relationship, Consequence, Harm/Foul, Politics, OI, and Sex, with greater permissiveness (greater tendency to judge the actions "Ok") expressed by individuals more willing to accept the harm-negating provisos, individuals identifying with the no-harm-no-foul ethic (Harm/Foul), politically liberal individuals, individuals with lower disgust ratings, and males. There were also significant associations of Relationship with Consequence, Harm/Foul, and Sex.

Most importantly, incest permissibility assessed in Survey 1 was strongly and significantly associated with Relationship assessed in session 2, with greater incredulity corresponding to greater likelihood that a subject would disapprove of the act. Furthermore, Relationship was not significantly associated with any of the following: Consistency, Social desirability, Prior exposure to the vignette, OI, and Politics (Table 3b), with the first three variables being also unrelated to Consequence, or the original permissibility judgment.

The correlational analyses were followed up with a hierarchical binary logistic regression, with permissibility as the dependent variable and Relationship, Consequence, Harm/Foul, Sex, Politics, OI (all in step 1) and Consistency

[9]The normative orientation check was always presented first to assure that subjects' general-level judgment was not affected by their reaction to the Incest vignette that followed.

[10]OI consists of three items ("gagging", "physically nauseated", "lacking appetite") rated (in this case) on a 5-point scale (Royzman et al., 2008; Royzman et al., 2014).

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

Table 3a: Sample descriptives for Study 3.

Variable	Mean/-Percentage	SD
Permissibility (0 = OK, 1 = Not Ok)	68%	N/A
Relationship (0 = accepting that the relationship will not be affected, 1 = not accepting this)	60%	N/A
Consequence (0 = accepting that the siblings will not be personally affected, 1 = not accepting this)	68%	N/A
Harm/Foul (0 = identifying with the no-harm-no-foul view, 1 = identifying with the no-harm-but-foul view)	42%	N/A
Social desirability average (score range: 0 to 1) ($\alpha = 0.559$)	0.37	0.19
Consistency average (score range: 1 to 9) ($\alpha = 0.822$)	6.17	1.15
Disgust (OI) (score range: 1 to 5) ($\alpha = 0.768$)	2.31	1.09
Politics (score range: 1 to 7, with higher scores indicating greater liberalism)	4.69	1.43

In case of the categorical variables (variables 1 through 4), percentages represent proportions of subjects selecting option coded as 1.

Table 3b: Zero-order correlations among key variables in Study 3. The three variables in bold font are jointly related to permissibility and relationship. A correlation of ±0.271 or above is significant at the 0.05 level (2-tailed) for this sample size (n = 53).

	Relationship	**Consequence**	**Harm/Foul**	**Sex***	Social desirability	Consistency	OI	Politics
Ok/Not Ok	.600	.567	.318	−.270	−.024	.124	.295	−.430
Relationship		.766	.467	−.369	.075	.128	.232	−.226
Consequence			.431	−.270	−.003	.224	.147	−.345
Harm/Foul				−.150	−.066	.284	−.058	−.467
Sex					−.096	−.079	−.119	−.045
Social desirability						−.157	−.097	.093
Consistency							.168	−.084
OI								−0.008

* 0 = female, 1 = male, with the negative correlation indicating greater permissiveness among male subjects.

Table 3c: Logistic regression coefficients, p-values, and odds ratios for incest permissibility (Ok/Not Ok) in Study 3 as a function of Credulity (Relationship, Consequence), normative identification (Harm/Foul), Sex, OI, Politics with and without the desire for consistency included.

	Model 1: Consistency not included			Model 2: Consistency included		
Variable	B	p (two-tailed)	Odds ratio	B	p (two-tailed)	Odds ratio
Relationship	3.80	.039	44.977	4.433	.034	84.194
Consequence	.278	.849	1.320	.084	.955	1.088
Harm/Foul	−1.484	.322	.227	−2.327	.223	.098
Sex	−1.784	.125	.168	−1.810	.145	.164
OI	.483	.353	1.621	.383	.478	1.467
Politics	−1.494	.019	.224	−1.698	.017	.183
Consistency				.445	.388	1.560

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

(in step 2) as covariates.[11] As seen in Table 3c, Relationship and Politics were the only two significant predictors of permissibility in either model.

To further explore the relative strengths of Relationship and OI as predictors of the evaluative response, we conducted three additional binary logistic regressions enlisting Ok/not judgment as the dependent variable and Relationship and OI as the two predictor variables. While both Relationship (B = 2.96; Odds ratio = 19.33, $p < 0.001$) and OI (B = 0.66; Odds ratio = 1.93, $p = 0.04$) were individually significant predictors of the permissibility (Ok/not Ok) response, Relationship was the only significant predictor when the two variables were entered as covariates (Relationship: $p < 0.001$; OI: $p = 0.16$). A follow-up analysis showed that these associations (permissibility—Relationship vs. permissibility—OI) were significantly different from each other by Steiger's z test for dependent correlations: $z = 2.08$, $p = 0.037$.

Finally, a two-person coding procedure (82% initial intercoder agreement; differential code assignments resolved through discussion) established that appeals to global negative consequences were the most common (70.8%) "foundational" reason offered by subjects espousing the view that "violating an established moral norm is inherently wrong" (Part 1 of Survey 2).[12] This result suggests that, at least among college undergraduates, truly committed deontologists—"deontologists all the way down"—may be few and far between.

Additional Analyses: "Unsupported declarations" and the moral dumbfounding estimation.

"Unsupported declarations" (Haidt et al., 2000) were the largest conceptually coherent category of statements (n = 20) generated in response to the Survey 1 request for reasons, with subjects either restating the relevant moral norm ("Incest is fundamentally wrong", "Brothers and sisters should not make love. Even it is a secret, it is still morally wrong", "Regardless of its being safe sex. They brother and sister. And that is just wrong", "It is immoral") or classifying the act in a manner that would warrant the application of that norm ("Incest", "Incest taboo"). As noted earlier, while one approach would be to regard such statements as further evidence of a morally dumbfounded state, our previous results (Study 2) indicate that these could also be viewed as colloquially phrased/under-articulated deontological claims. Consistent with this latter interpretation, we found a significant positive association between a tendency to make putative "unsupported declarations" in Survey 1 and the Survey 2-assesed likelihood of favoring a normative position designating acts in violation of an established moral norm as "inherently wrong" (chi-square = 6.85, p = 0.009). That is, a tendency to render "unsupported declarations" (e.g., "Incest is fundamentally wrong") was systematically and positively related to a tendency to identify with the view that violating an established moral norm is "plain" wrong, i.e., wrong irrespective of any harmful implications that could ensue.

The analyses reported in the remainder of this section were designed to provide a formal re-assessment of the incidence of moral dumbfounding defined as "a stubborn and puzzled maintenance of a moral judgment without supporting reasons" (Haidt et al., 2000, p. 8). In accordance with the rationale articulated in the Method, only those subjects who were "fully convergent" (in this case, 14 out 53 or 26.4 % of the sample) and thus truly "without supporting reasons" were considered eligible for further scrutiny. Only 4 of these 14 "fully convergent" subjects (i.e., those who both believed both that the siblings' actions were free of harm *and* that harm-free acts are not subject to disapproval) judged that Julie and Mark's behavior was "not Ok". All four of these subjects (two females, two males) were subsequently interviewed with the goal of determining how many, if any, would satisfy the remaining criteria of Haidt et al.'s (2000) definition by maintaining their disapproval in "a stubborn and puzzled" manner.

In each case, a subject was first presented with a printed summary of their earlier (Survey 1 and Survey 2) responses, and asked if they remembered and/or endorsed these responses as applying to the present case. All subjects were found to endorse their previous responses whether or not they remembered them. In step two, subjects were simply advised to carefully review and, if appropriate, revise any of their earlier judgments with particular attention being drawn to the normative relevance of harm. For subjects failing to make any adjustments at that point, the inconsistency between their Survey 1 and Survey 2 responses were pointed out directly.

In the course of the interview, two of the four subjects fully acknowledged the inconsistency between their Survey 1 and Survey 2 responses, with one subject reversing her case-specific judgment and the other reversing her prior endorsement of the no-harm-no-foul standard. One of the two male subjects almost immediately disqualified himself from the "fully convergent" classification by stating that his Survey 1 selection of "not Ok" option was intended merely as a descriptive statement indicating his awareness of the prevailing norm rather that a personal judgment that Julie and Mark's behavior was wrong: a judgment that he said he did not endorse. Finally, one male interviewee explicitly acknowledged the inconsistency between his "fully convergent" Survey 2 response set and his Survey 1 case-specific judgment that Julie and Mark's actions were "not Ok". Un-

[11] Consistency was entered in step 2 to explore the possibility that, as stipulated by the sentimentalist component of the moral dumbfounding narrative, disgust (OI) was the key determinant of both subjects' permissibility and (mediated by consistency) their unwillingness to accept the siblings'actions as genuinely harm-free.

[12] The two independent coders were the first author and a first-year undergraduate student with no prior knowledge of the hypothesis or background literature (see http://journal.sjdm.org/15/15405/supp1.pdf for verbatim statements and coding details).

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

like his male counterpart, this interviewee acknowledged that his judgment of "not Ok" did convey a personal moral disapproval of the act and, unlike his two female counterparts, he was either unable or/and unwilling to resolve the inconsistency by altering one or more elements of his overall response pattern.

In sum, with the requisite manipulation checks on credulity and normative orientation factored in, only 14 of 53 individuals involved in the study were classifiable as lacking supporting reasons, and only 3 of these 14 individuals genuinely disapproved of the siblings' decision to have sex. Furthermore, only 1 of these 3 "dumbfounding-qualified" subjects maintained his disapproval in the "stubborn and puzzled" manner, giving us a moral dumbfounding estimate of 1/53 (1.88 percent), not significantly greater than 0/53 ($z = 1.00$, $p = 0.32$).

5 General discussion

Three studies utilizing two different versions of the "Julie and Mark" vignette revealed that, contra the key assumption of the moral dumbfounding narrative, subjects were generally reluctant to accept the siblings' actions as harm-free (Study 1); notwithstanding this, and in spite of having other subjectively warrantable reasons to disapprove of the act, subjects went on to exhibit all the trademark signs of a morally dumbfounded state, including confusion, a tendency to withdraw reasons, and the declaration of dumbfounding itself (Study 2).[13] Finally, subjects' beliefs (their credulity) regarding the non-occurrence of certain types of harm, but not their level of physical disgust, strongly and uniquely predicted their disapproval of the act (Study 3). Expressions of incredulity, though somewhat varied from one study to the next, remained high irrespective of whether the credulity check was performed immediately upon reading the scenario (and in the absence of any normative evaluation of the act) (Study 1), at the end of a study session, following a detailed counterargument and repeated appeals to the harm-free nature of the act (Study 2), or (Study 3) as long as 4 weeks after the permissibility judgment was obtained. Moreover, these close-ended endorsements were in synch with subjects' spontaneous (pre-credulity-check) remarks about the imagined harmful implications of the act (e.g., the siblings finding it difficult to form romantic ties with other people, undergoing "a crisis" at some future date, and/or being tormented by their secret), mirroring similar remarks in Haidt et al. (see Haidt, 2001, Sommers, 2009) as well as in some prior work of our own (e.g., Royzman, 2009).

A key contribution of Study 3 was its attempt to assess the true incidence of MD, guided by Haidt et al.'s (2000) original definition of the term. We began by limiting our pool of candidates to those and only those (14 out of 53) whose unique configuration of normative endorsements (no harm, no foul) and empirical beliefs (no harm) left them truly "without supporting reasons" to disapprove of the act. Only 3 of these 14 individuals disapproved of the siblings having sex and only 1 of 3 (1.9%) maintained his disapproval in the "stubborn and puzzled" manner.

None of this is to deny that reason-givers may have a bias (see Baron, 2008, on myside bias). (But there is a world of difference between saying that one's reasons are somewhat biased and saying that one has *no reasons* whatsoever). Nor do we harbor any doubts that perceptions of harm and wrong *can* interact. Gray and colleagues' (Gray et al., 2014) recent analysis of "harmless wrongs" (watching animal sex to become aroused, sexually defiling a corpse) suggests that individuals tend to quickly and automatically infer that there is a harm where there is a wrong (just as individuals may presumably infer overwork from next day's fatigue or romantic attachment from a stab of jealous thoughts), making it feasible that, perhaps, a substantial proportion of subjects in Study 2 drew on their antecedent judgments of wrong to inform their study-long belief that the siblings were bound to "pay the price". Were such beliefs truly and genuinely held? We have little reason to think otherwise.[14] This suggests that, whether or not some of the relevant beliefs could be ultimately shown to lie upstream of the judgments of wrong, the fact that those holding these beliefs (and thus having subjectively warrantable reasons to disapprove of the act) still went on to display the trademark signs of MD, including the declaration of dumbfounding itself, reinforces the key deflationary points of Studies 2 and 3: Haidt et al.'s intuitively compelling approach to the diagnosis of a morally dumbfounded state is simply not the discriminantly valid measure that it was purported to be, with a more rigorous counterpart (one taking the precaution to filter out all those with real expectations of future harm and other subjectively warrantable reasons to disapprove of the act) yielding a dumbfounding estimate of 1.

Needless to say, it remains to be seen how the findings we report may change as a function of future studies that employ a different set of stimuli and a larger, less WEIRD (White, educated, industrialized, rich, democratic; Henrich, Heine & Norenzayan, 2010), non-collegiate sample. However, given that all three studies we discuss were intended as conceptual replications of Haidt et al. (2000), it bears

[13] In Bayesian terms, the problem could be described as one of a far-too-low diagnostic specificity rate (see Table 2a), creating the impression of a morally dumbfounded state even among those who clearly possessed (and knew that they possessed) subjectively warrantable reasons for disproving of the act.

[14] There is no a priori reason to doubt that these beliefs were any less genuinely held than the evaluative judgments with which they link; we also checked on this point more formally in Study 3, showing no significant association between beliefs about harm and either an established measure of social desirability or that of response consistency; there was also no relationship between either of these measures of "response authenticity" and judgments of wrong.

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

mention that Haidt et al.'s original conclusions derive entirely from interviewing a small (N = 31) and prototypically WEIRD subset of UVA undergraduates, with no cross-cultural replications having been reported at this date.

Furthermore, one could contend that, given the reputed association between a more "traditional" lifestyle and a "broader"/more "multi-value" moral outlook (Haidt, 2012; Shweder, 1990), isolating MD should prove to be *especially* tricky among the world's more representative populations who, on Haidt's (2012) current view (Moral Foundations Theory or MFT), would be able to access and adduce a far wider range of reason-giving considerations than their non-traditional counterparts. To illustrate, consider *the Flag*, one of the five "taboo violation" vignettes developed by Haidt as part of his doctoral work on the cultural underpinnings of moral judgment (Haidt, 1992; Haidt et al., 1993). The story centers on a woman who cuts up an old American (or Brazilian, when subjects were Brazilian) flag into rags, which she then uses for cleaning the bathroom. Though, along with other "taboo violations" in the set, the Flag has never been utilized as part of a formal dumbfounding interview (Haidt et al., 2000), its juxtaposition of an inanimate patient with a solitary agent makes it seem like a highly potent variation on the "harmless wrong" motif (Yoel Inbar, March 7, 2015, private communication). In our recent use of the Flag (N = 26, 19 female), we found that relatively few University of Pennsylvania undergraduates deemed the agent's behavior "morally wrong" (27%) and even fewer judged that the agent "should be punished" (4%) (willingness to punish being one of Haidt et al.'s [1993] two indicators that the action was viewed as genuinely immoral). The few who did judge the action to be morally wrong cited the flag's significance as a symbol of the nation's history and appealed to the principles of respect for that history as key considerations guiding their choice (e.g., "The values of loving your country and the representation of freedom that the flag signifies for me makes this morally wrong", "According to my morals, disrespecting or defacing a sacred symbol of national pride is symbolically not okay") (see http://journal.sjdm.org/15/15405/supp2.sav for the raw data and complete verbatim "explanations"). Lack of interpersonal harm was the key reason cited by those voicing no moral disapproval of the act. (Similar considerations would apply to other "taboo violations", e.g., *the Chicken* scenario—a man has sex with a dead chicken, then cooks and eats it; as Haidt [1992] pointed out, people's response to this act is grounded in two separate taboo violations [p. 31], bestiality and necrophilia, whose joint capacity to attract moral condemnation may rival that of incest).[15]

These findings are generally in synch with those reported by Haidt et al. (1993) some 23 years ago: among the relatively liberal University of Pennsylvania undergraduates and other high-SES Philadelphians, flag-cutting, chicken sex, and the like "were not [considered] morally wrong, as long as these actions were perceived to have no harmful interpersonal consequences" (Haidt, 1992, p. 45). Both the Flag and the Chicken *were* moralized by the low-SES respondents (especially, in Brazil), but, again, in Haidt's own interpretation, these other groups' disapproval was supported by their "broader construction of morality" (p. 45), defined by their endorsement of various codes of interpersonal conduct commanding respect for authority, tradition, and compliance with the natural law (Shweder, 1990) (see also Haidt et al., 1993 and Haidt, 2012), making them an *especially* unlikely population within which to bare symptoms of moral dumbfounding as such.

More generally, as the forgoing analysis illustrates, a definitionally pristine bout of MD is likely to be a extraordinarily rare find, one featuring a person who doggedly and decisively condemns the very same act that she has no prior normative reasons to dislike. From the Bayesian perspective, this means that any future reports of MD, especially those alleging it to be a common (or easily demonstrable) feature of moral cognition, should be treated with utmost caution and skepticism.

Ultimately, Haidt et al.'s (2000) success in "revealing" high incidence of MD among their subjects is attributable to two main factors. The first concerns the aforementioned social/conversational dynamics of the interviewing process (see Study 2). (Related to this factor is the subject's possible concern over not being able to fully articulate his or her position and/or coming across as "inattentive" or "stubborn", as well as, perhaps, the sheer desire to end the monotony of the interview, all leading to "I don't knows", etc. as an easy way out).

Second, at least some portion of the alleged dumbfounding effect can be explained by Haidt et al.'s decision not to view certain subjectively warrantable reasons as such. This tendency (a kind of normative hegemony)[16] has been already exemplified by Haidt et al.'s (2000) penchant for interpreting apparent deontological claims as cases of "unsupported declarations" (see Study 2 and 3 for the discussion

[15]These data highlight the difficulty that the Flag scenario (and others of its kind) present for dumbfounding research. Clearly, no pertinent dumbfounding interview can be conducted with a subject who deems the act "not morally wrong." But it is equally unclear how one would proceed in the case of the subject whose morals dictate that an important symbol of the nation's history must not be casually "defaced." Would one contest this person's apparent normative commitment to tradition and authority or would one contend (as other, more permissive subjects have argued) that it is up to each individual flag owner to decide if the cloth they own is truly a symbol or just a cloth? The ultimate strategy remains unclear.

[16]An example of this phenomenon from a non-academic context would be the commonly voiced complaint about the substandard quality of service offered at various dining establishments of Eastern and Central Europe, where servers are said to be as inattentive ("she never checked on us once") as they are slow ("it takes ages to get a cheque"). What such criticisms and grievances commonly overlook is the nature of the local hospitality norms, which prescribe that, as a matter of respect, guests must be allowed to eat in peace and not "rushed out" the moment they set down their forks.

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

and the evidence). An even more striking example is afforded by the "physical" dumbfounding task (adapted from Rozin, Millman & Nemeroff [1986]) in which a subject was invited to drink from a glass of water or apple juice into which a "sterilized" cockroach was momentarily immersed. Haidt et al. (2000) explain that the task was "designed to produce the same cognitive situation as the moral intuition tasks: a clear 'seeing-that' the act was wrong or undesirable, coupled with a difficulty in finding 'reasoning-why' to justify one's refusal" (p. 8). Indeed, when a subject refused to partake of the juice, the experimenter argued, Incest-style, that the roach was thoroughly sterilized and posed no risk of disease. Further refusals were interpreted as a sign that the subject was "clearly dumbfounded" (Haidt et al., 2000, p. 14) and beyond the ken of rational persuasion. That is, in the minds of these researchers, the sheer psychological *unpleasantness* of taking in the recently "roached" juice (see Royzman & Sabini, 2001 on disgust as a "cognitively impenetrable" response to concrete elements of a situation) did not qualify as a subjectively warrantable reason for saying "No!" to the juice[17] (just as presumably the momentary physical distress caused by a mild electric jolt would not impress them as a subjectively warrantable reason for saying "No!" to the jolt).

Indeed, Haidt's more recent work on the foundations of moral cognition (Haidt, 2012) indicates that appeals to disgustingness (unnaturalness, weirdness, and the like) may function as proper reasons even amidst a moral dumbfounding interview. As previously discussed, the signature feature of MFT (Haidt, 2012) (see also Haidt et al., 1993) is its inclusion (and normative legitimization) of a set of non-utiltiarian considerations, e.g., Purity/Sanctity, that allow for a deed or a mode of conduct to be censured or soundly condemned based on "disgustingness" alone. Two of the total of six items developed by Haidt and colleagues to assess a subject's endorsement of the value of Purity (a.k.a., the Purity/Sanctity foundation) are couched in the language of disgust: "[It is morally relevant] whether or not someone did something disgusting" and "People should not do things that are disgusting, even if no one is harmed" (see http://www.yourmorals.org). This means that, from the pluralistic perspective of MFT, telling subjects that "the fact that an act is disgusting does not make it wrong" (Haidt et al., 2000, p. 9) would almost certainly preclude some (more traditional) individuals from accessing the very language or mode of expression that they would need to ally themselves with sexual impropriety-linked reasons of Purity, and, thus, with reasoned condemnation as such.

Finally, in line with some subjects' comments, we speculate that at least a part of the confusion surrounding the subject-experimenter interactions in Haidt et al. (2000) is attributable the interactants' widely divergent views on the nature of the justificatory process, with some subjects using appeals to, say, familial discord or "dangers of inbreeding" (Haidt, 2001, p. 814) not so much as *proximate* reasons for their antecedently acknowledged disapproval of *the act*, but rather as *foundational* reasons for upholding the proscriptive norm (i.e., the incest taboo) that they assumed to be tacitly invoked by making their disapproval heard.

5.1 Conclusion

All in all, the data gathered across three studies and one pilot study demonstrate that, contra the received wisdom, subjects' seemingly arational reactions to the "Julie and Mark" vignette are largely in line with the rationalist ideal of moral evaluation espoused by all major scholars of moral cognition from Kant to Kohlberg (and beyond). More generally, the paper highlights (a) the need for more robust manipulation checks on whether the cognitively taxing demands embedded in many a scenario-based moral judgment task have been fully or even partially met as well as (b) the need for a lucid and thoughtful discussion *on what may or may count as supporting reasons* in the context of a moral judgment task, with an eye toward articulating clearer normative benchmarks whereby future candidate cases of moral unreason may be rationally selected and assessed. Furthermore, our three studies bring to light some fairly nuanced ways in which harm or harm-related considerations may enter the process of moral evaluation, while also drawing attention to the general importance of giving due weight to subjects' own standards of judgment, empirical beliefs, and conceptualization of the justificatory process, *all* of which may differ considerably from those favored by the researchers masterminding the study or the scientific community at large.

References

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39,* 1037–1049.

Asch, S. E. (1956). Studies of independence and submission to group pressure: I. A minority of one against a unanimous majority. *Psychological Monographs*, *70*(9, Whole No. 417).

Baron, J. (2008). *Thinking and deciding (4th ed.).* Cambridge, UK: Cambridge University Press.

[17]The implicit premise of the "Roach" is that the deeper evolutionary "reason" for subjects' feelings of revulsion (and resultant avoidance)—a need to steer clear of pathogens and those that move them around—is effectively nullified once the critter is rendered germ-free (making all those continuing to say "No" dumbfounded). However, it seems that, by the same token, dumbfounding would have to be imputed to a group of "naïve" male subjects whose stated eagerness to bed an attractive female confederate remains unabated even after being informed that her current contraceptive regiment makes her utterly unable to conceive.

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology, 54,* 569–579.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*, 154–161.

Bering, J. M. (2002) Intuitive conceptions of dead agents' minds: The natural foundations of afterlife beliefs as phenomenological boundary. *Journal of Cognition and Culture*, *2*, 263– 308.

Bering, J. M. (2006). The folk psychology of souls. *Behavioral and Brain Sciences*, *29*, 453–462.

Bonnefon, J. F., Feeney, A., & De Neys, W. (2011). The risk of polite misunderstanding. *Current Directions in Psychological Science*, *20*, 321-324.

Benforado, A. (2015). *Unfair: The New Science of Criminal Injustice*. Crown.

Camerer, C., Lowenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97,* 1232–1254.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology, 69*, 318–328.

Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment*. Cambridge: Cambridge University Press.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye- movement investigation. *Journal of Memory and Language*, *58*, 609–626.

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes*, *25*, 297–346.

Fischhoff, B. (1975). Hindsight $\neq$ foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299.

Goffman, E. (1955). On Face-work: An Analysis of Ritual Elements of Social Interaction. *Psychiatry: Journal for the Study of Interpersonal Processes, 18*, 213-231.

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*, 1600–1615.

Greene, J. (2013). *Moral tribes*. Penguin Press.

Haidt, J. (1992). *Moral judgment*, *affect*, *and culture*, or is it *wrong to eat your dog*? Unpublished doctoral dissertation. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon.

Haidt, J., Bjorklund, F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason*. Unpublished manuscript, University of Virginia.

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.

Harman, G., Mason, K., & Sinnott-Armstrong, W. (2010). Moral reasoning. In J. Doris & the Moral Psychology Research Group (Eds.), *The moral psychology handbook* (pp. 206–245). New York: Oxford University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010).The weirdest people in the world? *Behavioral and brain sciences*, *33*, 61–83.

Huebner, B. (2011). Critiquing empirical moral psychology. *Philosophy of the social sciences*, *41*, 50–83.

Hume, D. (1978). *A treatise of human nature.* Oxford, UK: Oxford University Press. (Original work published 1739-1740).

Hume, D. (1983). *An enquiry concerning the principles of morals*. J.B. Schneewind (Ed.). New York: Hackett. (Original work published 1751).

Jacobson, D. (2013). Moral dumbfounding and moral stupefaction. M. Timmons (Ed.), *Oxford Studies in Normative Ethics*, *Volume 2* (pp. 289–316). Oxford: Oxford University Press.

Kagan, S. (1998). *Normative ethics.* Boulder, CO: Westview Press.

Kant, I. (1959). *Foundations of the metaphysics of morals (L. W. Beck, Trans.)*. Indianapolis: Bobbs-Merrill (Original work published 1785).

Milgram, S. (1974). *Obedience to Authority*. New York: Harper and Row.

Norenzayan, A., & Schwarz, N. (1999). Telling what they want to know: Participants tailor causal attributions to researchers' interests. *European Journal of Social Psychology*, *29*, 1011–1020.

Piazza, J., & Sousa, P. (2014). Religiosity, political orientation, and consequentialist moral thinking. *Social Psychological and Personality Science, 5,* 334-342.

Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A reply to Haidt (2001). *Psychological Review*, *110*, 193–196.

Pinker, S. (2002). *The blank slate*. London: Penguin Classics.

Royzman, E., Atanasov, P., Landy, J. F., Parks, A., & Gepty, A. (2014). CAD or MAD? Anger (not disgust) as the predominant response to pathogen-free violations of the Divinity code. *Emotion*, *14*, 892–907.

Royzman, E., Cassidy, K., & Baron, J. (2003). "I know, you know": Epistemic egocentrism in children and adults. *Review of General Psychology*, *7*, 38–65.

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. (2015). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive science*, *39*, 325–352.

Royzman, E. B., Leeman, R., & Baron, J. (2009). Unsentimental ethics: Towards a content-specific account of the moral-conventional distinction. *Cognition*, *112*, 159–174.

Royzman, E. B., Leeman, R., & 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., & Sabini, J. (2001). Something it takes to be an emotion: The interesting case of disgust. *Journal for the Theory of Social Behaviour, 31,* 29–59.

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic magic in disgust and other domains. *Journal of Personality and Social Psychology*, *50,* 703–712.

Rozin, P., & Stellar, J. (2009). Posthumous events affect rated quality and happiness of lives. *Judgment and Decision Making*, *4*, 273–279.

Sabini, J. (1995). *Social psychology (2nd ed.).* New York: Norton.

Shor, E., & Simchai, D. (2009). Incest Avoidance, the Incest Taboo, and Social Cohesion: Revisiting Westermarck and the Case of the Israeli Kibbutzim. *American journal of sociology*, *114*, 1803-1842.

Shweder, R. A. (1990). In defense of moral realism: Reply to Gabennesch. *Child Development, 61,* 2060–2067.

Singer, P. (2005). Ethics and intuitions. *Journal of Ethics, 9,* 331-352.

Sommers, T. (2009). *A very bad wizard.* Believer Books.

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crown Social Desirability Scale. *Journal of Clinical Psychology*, *28*, 191–193.

Taylor, G., & Wolfram, S. (1968). The self-regarding and other-regarding virtues. *The Philosophical Quarterly*, *18*, 238–248.

Thomson, J.J. (1986). *Rights, restitution and risk*. Cambridge, MA: Harvard University Press.

Turiel, E. (2002). *The culture of morality: social development, context, and conflict.* New York: Cambridge University Press.

Appendix: Additional scales used

Social desirability scale (MC-1) (Strahan & Gerbasi, 1972)

Personal Reaction Inventory

Listed below are a number of statements concerning personal attitudes and traits. Please read each item and decide whether the statement is true (circling T) or false (circling F) as it pertains to you personally.

1. I'm always willing to admit it when I make a mistake.
2. I always try to practice what I preach.
3. I never resent being asked to return a favor.
4. I have never been irked when people expressed ideas very different from my own.
5. I have never deliberately said something that hurt someone's feelings.
6. I like to gossip at times.
7. There have been occasions when I took advantage of someone.
8. I sometimes try to get even rather than forgive and forget.
9. At times I have really insisted on having things my own way.
10. There have been occasions when I felt like smashing things.

The Preference for Consistency Scale (Cialdini, Trost & Newsom, 1995)

Listed below are a number of statements. You will probably disagree with some of them and agree with others. In front of each item below, please write the number: 1 if you strongly disagree, 2 if you disagree, 3 if you somewhat disagree, 4 if you slightly disagree, 5 if you neither agree nor disagree, 6 if you slightly agree, 7 if you somewhat agree, 8 if you agree, and 9 if you strongly agree. Please answer each question as honestly and accurately as you can, but don't spend too much time thinking about each answer.

1. It is important to me that those who know me can predict what I will do.
2. I want to be described by others as a stable, predictable person.
3. The appearance of consistency is an important part of the image I present to the world.
4. An important requirement for any friend of mine is personal consistency.
5. I typically prefer to do things the same way.
6. I want my close friends to be predictable.
7. It is important to me that others view me as a stable person.
8. I make an effort to appear consistent to others.
9. It doesn't bother me much if my actions are inconsistent

https://doi.org/10.1017/S193029750000512X Published online by Cambridge University Press