



Discussion

Unfounded dumbfounding: How harm and purity undermine evidence for moral dumbfounding



Steve Guglielmo

Department of Psychology, Macalester College, 1600 Grand Avenue, St. Paul, MN 55105, United States

ARTICLE INFO

Article history:

Received 21 April 2017

Revised 3 August 2017

Accepted 4 August 2017

Available online 10 August 2017

Keywords:

Morality

Dumbfounding

Purity

Wrongness

Moral Foundations Theory

Theory of Dyadic Morality

ABSTRACT

Two paradigm-shifting ideas have gained widespread influence in current accounts of moral cognition: (a) that moral judgments are pluralistic, extending beyond domains of harm and fairness, and (b) that people's judgments are driven primarily by intuition, such that people are "morally dumbfounded" about the reasons behind their own judgments. An ongoing debate has emerged regarding the former claim of moral pluralism, with opposing sides in disagreement about whether moral judgments are best understood as reflecting multiple moral domains vs. a single moral domain. The current analysis demonstrates that however this debate concerning pluralism is resolved, evidence of moral dumbfounding is undermined. This evidentiary basis for intuitive moral judgment is therefore not well supported, and additional evidence indicates that moral judgments are more reasoned and malleable than the dumbfounding account would allow.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Current perspectives on moral cognition have been heavily influenced by two paradigm-shifting claims: (a) that moral judgments are pluralistic, reflecting concerns not just about harm and fairness but also about purity (among other concerns), and (b) that people are "morally dumbfounded" by their own moral judgments, ignorant of the factors that drive these judgments. The latter claim of moral dumbfounding has been taken as evidence that moral judgments stem primarily – if not near-exclusively – from processes of intuition rather than reasoning (Haidt, 2001). The former claim of moral pluralism has led to an ongoing debate, with opposing sides (Graham, 2015; Schein & Gray, 2015) in disagreement about whether purity is a distinct moral domain or whether it is subsumed under a unifying superordinate domain of harm. The current analysis illustrates how moral dumbfounding intersects with the moral pluralism debate; more specifically, it shows that however this debate is resolved, evidence of moral dumbfounding is undermined.

2. Moral pluralism and purity

It has long been held that morality reflects considerations of harm and fairness: immoral behaviors are those that are perceived as harmful or unfair toward individuals (or groups). A great deal

research in the past decade, however, illustrates that the scope of morality extends beyond these two domains. Moral Foundations Theory (MFT), the most prominent account of such moral pluralism, holds that moral judgments additionally reflect considerations of purity – or whether behaviors are unnatural, degrading, or disgusting – as well as loyalty and authority (Haidt & Joseph, 2007). Building upon earlier work that highlighted the moral significance of purity, especially among less westernized cultures (Haidt, Koller, & Dias, 1993), more recent studies under the MFT framework have provided systematic evidence for this moral domain, even among highly westernized cultures (Graham, Haidt, & Nosek, 2009; Graham et al., 2011; Koleva, Graham, Iyer, Ditto, & Haidt, 2012).

Purity concerns are represented in people's everyday experience of immoral acts (Hofmann, Wisneski, Brandt, & Skitka, 2014), and purity is often linked to specific emotional experiences, particularly involving disgust. Purity-based moral judgments are associated with higher levels of trait disgust (Horberg, Oveis, Keltner, & Cohen, 2009; Inbar, Pizarro, & Bloom, 2009) and are exacerbated by inducing disgust via various sensory modalities (Eskine, Kacinik, & Prinz, 2011; Schnall, Haidt, Clore, & Jordan, 2008; Wheatley & Haidt, 2005; see Landy & Goodwin, 2015 for a meta-analysis of these effects). Notably, some evidence has shown that these induction effects emerge only for participants who are highly aware of the presence of the disgusting stimuli (Wisneski & Skitka, 2017) or of their own bodily sensations (Schnall et al., 2008), suggesting that people may perceive their own purity-based moral judgments to be normatively appropriate. Finally,

E-mail address: sgugliel@macalester.edu

purity is dissociable from harm in terms of both its elicitors and its consequences — for example, as compared to harm, moral perceptions of purity are less sensitive to an agent's intent (Chakroff et al., 2016; Young & Saxe, 2011), more likely to lead to person-based attributions (Chakroff & Young, 2015), and better predictive of moral judgments about particular violations such as suicide (Rottman, Keleman, & Young, 2014). Taken together, this work suggests that purity is a distinct moral domain that itself helps to explain people's everyday experience and expression of morality.

3. Moral dumbfounding: Demonstrated and undermined

Beyond illustrating the moral importance of purity, the literature has also advanced the claim that people are morally dumbfounded by their own judgments. According to this widely influential phenomenon, people often view behaviors as moral violations yet are unable to provide sufficient reasons why. In the paradigmatic illustration of this effect (Haidt, Björklund, & Murphy, n.d.), participants read “intuition” scenarios designed to depict harmless yet morally objectionable behavior (e.g., a brother and sister having consensual sex while using multiple forms of birth control; a woman eating a disease-free cadaver), as well Kohlberg's (1969) Heinz dilemma, which involves tradeoffs between prototypical considerations of harm and rights. Participants exhibited notably different responses to these cases, reporting that their judgments were based primarily on reasoning for the Heinz dilemma, but primarily on intuition for the incest and cannibalism cases. Moreover, participants not only frequently judged the incest and cannibalism behaviors to be wrong but also persisted in these judgments when challenged by the experimenter, typically appealing to (ostensibly non-existent) harmful consequences or steadfastly affirming that the behavior is wrong (Haidt et al., n.d.). This persistence in perceiving wrongness despite lacking compelling articulable justifications was said to indicate that participants were morally dumbfounded: that they were ignorant of the actual moral principles that drove their judgments. Similar dumbfounding patterns have emerged for other purity-violating but putatively harmless behaviors (e.g., engaging in unusual forms of masturbation: Haidt & Hersh, 2001).

The phenomenon of moral dumbfounding plays a central role in Haidt's (2001) influential Social Intuitionist Model, which argues that moral judgment is driven almost exclusively by intuition rather than reasoning.¹ More generally, the paradigmatic findings have gained widespread acknowledgement and acceptance in the literature as a hallmark of the intuitive nature of moral judgment (e.g., Ditto, Liu, & Wojcik, 2012; Teper, Zhong, & Inzlicht, 2015; Usoof-Thowfeek, Janoff-Bulman, & Tavernini, 2011; Van Bavel, Packer, Haas, & Cunningham, 2012).

Despite the prominence of moral dumbfounding in the literature, evidence for the phenomenon is sparse and not widely generalizable. Just a single study, though itself not specifically designed to assess the phenomenon, offers pertinent evidence of dumbfounding for non-purity violations (Cushman, Young, & Hauser, 2006). This study examined whether participants had conscious, explicit access to three principles shown to shape their harm-based moral judgments (e.g., that actions are morally worse than omissions). Cushman et al.'s (2006) results revealed that just one principle exhibited patterns consistent with dumbfounding (i.e., substantial rates of insufficient justifications for participants' own judgments, indicating an implicit, rather than explicit, usage of the principle). Consequently, there is little support for a broad

interpretation that dumbfounding is a pervasive feature of moral cognition. But even when considering only a narrower conception of moral dumbfounding (i.e., for paradigmatic purity violations), there are still strong reasons to doubt the veracity of this phenomenon. These reasons arise from competing perspectives about whether purity constitutes its own distinct moral domain.

The theory of dyadic morality (TDM) argues that perceived harm lies at the core of all moral judgments (Gray, Young, & Waytz, 2012; Schein & Gray, 2017); thus, purity is not its own moral domain but rather is a specific manifestation of the superordinate domain of harm. Evidence supporting this view shows that harm is the most cognitively accessible and morally impactful domain (Schein & Gray, 2015), and that even ostensibly harmless purity violations are perceived to be harmful (Gray, Schein, & Ward, 2014). In contrast, the pluralistic MFT account maintains that purity constitutes a moral domain distinguishable from harm (Graham, 2015). Evidence supporting this view shows that, as compared to perceptions of harm, perceptions of purity or sacredness are often stronger (Frimer, Tell, & Haidt, 2015), sensitive to different input factors (Dungan, Chakroff, & Young, 2017; Young & Saxe, 2011) and better predictive of certain moral judgments, such as those concerning self-directed violations (Dungan et al., 2017; Rottman et al., 2014). A fierce debate continues between these competing perspectives; but the truth of *either* perspective spells trouble for moral dumbfounding.

According to the TDM perspective, the putatively harmless purity violations typically used to assess moral dumbfounding (e.g., consensual incest) might not actually be perceived as harmless. Thus, people's moral objections to such behaviors might reflect their concerns about harm. In fact, participants often perceive actual or probable harm in the consensual incest scenario, which thereby explains their moral disapproval of the act (Royzman, Kim, & Leeman, 2015; see also Jacobson, 2012). Consequently, people's common insistence that moral dumbfounding behaviors are both immoral and harmful indicates not that they are dumbfounded by their own judgments but that perceived harm is precisely their basis for viewing the behaviors as morally objectionable. However, when people offered harm-based explanations, these were interpreted as indicative of dumbfounding, with the experimenter countering such explanations by insisting that “no harm was done” (Haidt et al., n.d.). Royzman et al. (2015) showed that this procedural approach of repeatedly countering participants' harm-based justifications yields merely the *appearance* of dumbfounding. With their justifications unflinchingly rejected by the experimenter, participants eventually admit that they lack any further basis for their moral objections; however, they privately remain unconvinced by the experimenter's counterarguments, holding firm to their original harm-based justifications (Royzman et al., 2015).

According to the pluralistic MFT perspective, purity is categorically distinct from harm, constituting its own moral domain. References to purity, then, should constitute valid psychological justifications of people's moral judgments.² That is, we should expect people to appeal to impurity, disgustingness, or unnaturalness when explaining their moral objection to supposed dumbfounding behaviors. And indeed they often do, but Haidt et al. (n.d.) again characterized such explanations as invalid: After participants judged that various purity violations were immoral, the experimenter's “main counter argument was that no harm was done, and that the fact that an act is disgusting does not make it wrong.” Purity-based explanations were therefore prevented — by researcher fiat — from constituting valid justifications for purity-based judg-

¹ The model nonetheless posits that reasoning can have substantial interpersonal influence over others' moral judgments (see also Mercier & Sperber, 2011), despite having minimal intrapersonal influence over one's own judgments.

² The question of whether reasons are psychologically valid — whether, descriptively, they were the inputs to one's judgment — is of course distinct from the question of whether these reasons are normatively valid or defensible.

ments. When people persisted in their commitment to purity concerns by, for example, maintaining that incest is just wrong (Royzman et al., 2015), they were categorized as morally dumbfounded (Haidt et al., n.d.).

To accept moral judgments as expressions of people's genuinely-held subjective beliefs is to grant explanatory power to whichever factors give rise to these judgments. When someone judges a behavior to be a moral violation, there is, necessarily, some explanatory factor(s) to which this judgment is attributable. In the case of the paradigmatic dumbfounding behaviors (such as consensual incest), the plausible explanatory factors are perceived harm or perceived impurity. That is, the descriptive explanation for people's moral objections to such behaviors is that they are perceived as harmful or as violating inviolable standards of purity. However, the paradigmatic demonstration of moral dumbfounding closed off both such explanations from constituting valid justifications for people's moral objections. People justified their moral objections by appealing to harmfulness; but since the behaviors in question were ostensibly harmless ("...no harm was done": Haidt et al., n.d.), these objections were taken to indicate dumbfounding. People also justified their objections by appealing to impurity; but since impurity is allegedly not a valid basis for moral condemnation ("...the fact that an act is disgusting does not make it wrong": Haidt et al., n.d.), these objections, too, were taken to indicate dumbfounding. Accordingly, if people cited either of the principles that, descriptively, are capable of accounting for their moral objections, they were said to be dumbfounded.

The ongoing debate regarding moral pluralism — although itself not yet resolved — therefore reveals that claims of moral dumbfounding have been greatly exaggerated. If purity-violating behaviors are perceived as genuinely harmful, as claimed by the TDM account, then citing harm-based reasons for moral disapproval cannot constitute dumbfounding. If purity is its own genuine moral domain, as claimed by the MFT account, then citing purity-based reasons for moral disapproval cannot constitute dumbfounding. In citing these two sets of reasons (as in Haidt et al., n.d.), participants clearly recognized the psychological basis of their moral objections.

4. Conclusion

The influential moral dumbfounding account, despite its limited and non-generalizable evidence, holds that people are ignorant of the factors that drive their own moral judgments, and has been taken as central evidence for the view that moral judgments are driven primarily by intuition. However, the present reinterpretation of this phenomenon challenges this claim, showing instead that people in fact recognize and articulate the grounds for their moral judgments. Countering the dumbfounding implication that judgments are hasty or unreasoned, the present conclusion is consistent with a host of other work demonstrating that moral judgments can often be structured and flexible. People adeptly utilize and identify the rules that underlie their moral judgments (Bennis, Medin, & Bartels, 2010; Nichols & Mallon, 2006; Trémolière, De Neys, & Bonnefon, 2017), systematically acquire new information to help inform their moral judgments (Guglielmo & Malle, 2017), and likewise adjust their initial judgments in response to receiving new information (Monroe & Malle, 2017). Such flexibility is pivotal in enabling us to reconsider our moral positions, providing the basis for genuine moral change and progress (Bloom, 2016; Pizarro & Bloom, 2003). Thus, moral judgments are far more reasoned, complex, and malleable than the dumbfounding account would allow. To be sure, our moral judgments often show bias or partiality (Mazar, Amir, & Arieli, 2008; Valdesolo & DeSteno, 2007), but the claim that we are dumb-

founded by them — that we are oblivious to the reasons behind our own moral judgments — is unfounded.

References

- Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, 5, 187–202.
- Bloom, P. (2016). *Against empathy: The case for rational compassion*. New York: Ecco.
- Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, 11, 476–484.
- Chakroff, A., & Young, L. (2015). Harmful situations, impure people: An attribution asymmetry across moral domains. *Cognition*, 136, 30–37.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, 17, 1082–1089.
- Ditto, P. H., Liu, B., & Wojcik, S. P. (2012). Is anything sacred anymore? *Psychological Inquiry*, 23, 155–161.
- Dungan, J. A., Chakroff, A., & Young, L. (2017). The relevance of moral norms in distinct relational contexts: Purity versus harm norms regulate self-directed actions. *PLoS One*, 12, e0173405.
- Eskine, K. J., Kaciniuk, N. A., & Prinz, J. (2011). A bad taste in the mouth: Gustatory disgust influences moral judgment. *Psychological Science*, 22, 295–299.
- Frimer, J. A., Tell, C. E., & Haidt, J. (2015). Liberals condemn sacrilege too: The harmless desecration of Cerro Torre. *Social Psychological and Personality Science*, 6, 878–886.
- Graham, J. (2015). Explaining away differences in moral judgment: Comment on Gray & Keeney (2015). *Social Psychological and Personality Science*, 6, 869–873.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029–1046.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101, 366–385.
- 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.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101–124.
- Guglielmo, S., & Malle, B. F. (2017). Information-acquisition processes in moral judgments of blame. *Personality and Social Psychology Bulletin*, 43, 957–971.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.
- Haidt, J., Björklund, F., & Murphy, S. (n.d.). Moral dumbfounding: When intuition finds no reason. Unpublished manuscript.
- Haidt, J., & Hersh, M. (2001). Sexual morality: The cultures and emotions of conservatives and liberals. *Journal of Applied Social Psychology*, 31, 191–221.
- Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 367–391). New York: Oxford.
- 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.
- Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, 12, 1340–1343.
- Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of purity. *Journal of Personality and Social Psychology*, 97, 963–976.
- Inbar, Y., Pizarro, D. A., & Bloom, P. (2009). Conservatives are more easily disgusted than liberals. *Cognition and Emotion*, 23, 714–725.
- Jacobson, D. (2012). Moral dumbfounding and moral stupefaction. *Oxford Studies in Normative Ethics*, 2, 289–316.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally.
- Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality*, 46, 184–194.
- Landy, J. F., & Goodwin, G. P. (2015). Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspectives on Psychological Science*, 10, 518–536.
- Mazar, N., Amir, O., & Arieli, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633–644.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34, 57–74.
- Monroe, A. E., & Malle, B. F. (2017). Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General*, 146, 123–133.
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100, 530–542.
- Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: Comment on Haidt (2001). *Psychological Review*, 110, 193–196.
- Rottman, J., Keleman, D., & Young, L. (2014). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition*, 130, 217–226.

- Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, *10*, 296–313.
- Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*, 1147–1163.
- Schein, C., & Gray, K. (2017). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*.
- Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*, 1096–1109.
- Teper, R., Zhong, C.-B., & Inzlicht, M. (2015). How emotions shape moral behavior: Some answers (and questions) for the field of moral psychology. *Social and Personality Psychology Compass*, *9*, 1–14.
- Trémolière, B., De Neys, W., & Bonnefon, J. F. (2017). Reasoning and moral judgment: A common experimental toolbox. In L. J. Ball & V. A. Thompson (Eds.), *International handbook of thinking and reasoning*. Hove, UK: Psychology Press.
- Usoof-Thowfeek, R., Janoff-Bulman, R., & Tavernini, J. (2011). Moral judgments and the role of social harm: Differences in the role of automatic versus controlled processing. *Journal of Experimental Social Psychology*, *47*, 1–6.
- Valdesolo, P., & DeSteno, D. (2007). Moral hypocrisy social groups and the flexibility of virtue. *Psychological Science*, *18*, 689–690.
- Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same action. *PLoS One*, *7*, 1–14.
- Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science*, *16*, 780–784.
- Wisneski, D. C., & Skitka, L. J. (2017). Moralization through moral shock: Exploring emotional antecedents to moral conviction. *Personality and Social Psychology Bulletin*, *43*, 139–150.
- Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*, 202–214.