UNIVERSITY OF LEEDS

This is a repository copy of *Many moral buttons or just one? Evidence from emotional facial expressions*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/135327/

Version: Accepted Version

Article:
Franchin, L, Geipel, J, Hadjichristidis, C orcid.org/0000-0002-9441-6650 et al. (1 more author) (2019) Many moral buttons or just one? Evidence from emotional facial expressions. Cognition and Emotion, 33 (5). pp. 943-958. ISSN 0269-9931

https://doi.org/10.1080/02699931.2018.1520078

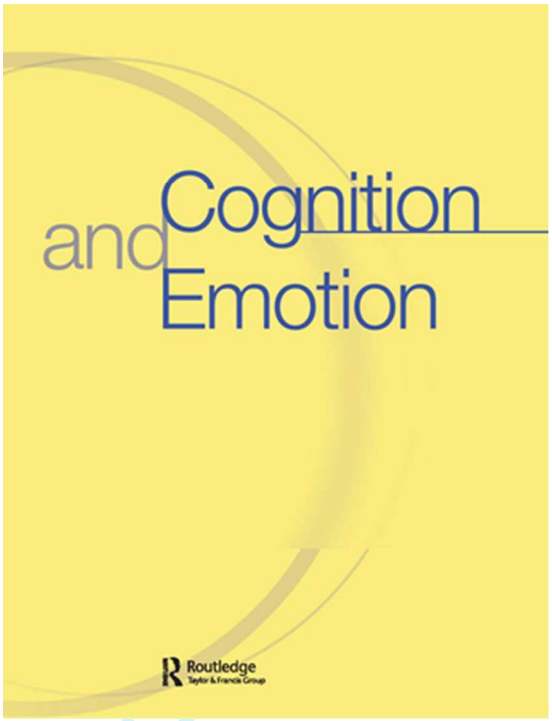
© 2018 Informa UK Limited, trading as Taylor & Francis Group. This is an author produced version of a paper published in Cognition and Emotion . Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/



Many moral buttons or just one? Evidence from emotional facial expressions

Journal:	*Cognition and Emotion*
Manuscript ID	CEM-FA 444.17.R3
Manuscript Type:	Full Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Franchin, Laura; University of Trento Geipel, Janet; University of Chicago Hadjichristidis, Constantinos; University of Trento; University of Leeds, UK Surian, Luca; University of Trento
Keywords:	Facial expression, Moral judgment, Moral Foundation Theory, Harm, Purity

SCHOLARONE™
Manuscripts

Many moral buttons or just one? Evidence from emotional facial expressions

Laura Franchin[1], Janet Geipel[2], Constantinos Hadjichristidis[3,4], and Luca Surian[1]

[1]Department of Psychology and Cognitive Sciences, University of Trento, Italy

[2]Department of Psychology, University of Chicago, USA

[3]Department of Economics and Management, University of Trento, Italy

[4]Centre for Decision Research, University of Leeds, UK

Acknowledgments. We thank students who participated in our study and Ilaria Perrucci for her help in data collection and in coding facial expressions. The authors declare no conflict of interest.

Corresponding Author: Laura Franchin, Department of Psychology and Cognitive Sciences, University of Trento, Corso Bettini, 31, 38068 Rovereto (Trento), Italy; e-mail: laura.franchin@unitn.it, telephone: +390464808633.

Word count = 9087

For Peer Review Only

Abstract

We investigated whether moral violations involving harm selectively elicit anger, whereas purity violations selectively elicit disgust, as predicted by the Moral Foundations Theory (MFT). We analyzed participants' spontaneous facial expressions as they listened to scenarios depicting moral violations of harm and purity. As predicted by MFT, anger reactions were elicited more frequently by harmful than by impure actions. However, violations of purity elicited more smiling reactions and expressions of anger than of disgust. This effect was found both in a classic set of scenarios and in a new set in which the different kinds of violations were matched on weirdness. Overall, these findings are at odds with predictions derived from MFT and provide support for 'monist' accounts that posit harm at the basis of all moral violations. However, we found that smiles were differentially linked to purity violations, which leaves open the possibility of distinct moral modules.

Keywords: Facial expression; Moral judgment; Moral Foundations Theory; Harm; Purity

For Peer Review Only

Many moral buttons or just one? Evidence from emotional facial expressions

One of the central goals in moral psychology is to uncover the mechanisms that underpin moral judgment. To understand these mechanisms, we need to determine the extent to which their input, output and characteristic computations are domain-specific, that is, the extent to which they compute only particular types of information and produce certain kinds of outputs. The theoretical positions given to this question fall into two camps, sometimes called 'pluralists' and 'monists'. Pluralists claim that moral cognition is composed of many domain-specific mechanisms that are selectively triggered by different kinds of violations. According to a famous proposal, the Moral Foundations Theory (MFT), there are at least five specialized mechanisms that deal with violations in the domains of harm, fairness, loyalty, authority and purity (e.g., Graham et al., 2013). They are specialized in their input requirements—they are triggered by different input conditions—but also in their computations and outputs, for example, they elicit distinct emotions. Alternative 'monist' accounts propose that morality is about harm and that all moral disapprovals involve, to a different degree, a concern for harm. Monist accounts may endorse a constructionist perspective on how information is integrated during processing (for a review, see Cameron, Lindquist, & Gray, 2015) or a modular breakdown of processing stages (e.g., Cushman, 2013; Mikhail, 2007). In both cases, harm violations may occur in many different contexts and numerous factors, such as directness and intentionality, can constraint moral judgments.

In the present work, we focus on what is considered to be a crucial test for tackling the issue of domain-specificity: Do different types of moral violations elicit different types of emotions? A positive answer to this question has been one of the initial empirical claims for multiple modules theories. Support for it came from studies using a variety of tasks and paradigms. In their seminal paper, Haidt, Koller, and Dias (1993) presented children and adults with stories purported to include instances of harmless actions that were either disrespecting, and therefore violated the 'Community ethics', or were “unconventional food and sexual practices, designed to trigger disgust” (p. 618), like eating a dog, kissing a sibling in the mouth, or masturbating using a

For Peer Review Only

dead chicken. They found that people in the United States and Brazil judged these acts as immoral, often justifying their judgments by stating that the acts were disgusting. Note that such actions were performed privately, to avoid the criticism that public disrespect and taboo violations may cause harm on witnessing people, because, if they do, then they fall into the domain of harm-based morality (Turiel, 1989; Turiel, Killen, & Helwig, 1987). In another influential study (Rozin, Lowery, Imada, & Haidt, 1999), people were asked to pair different types of violations with different facial emotional expressions or emotion words. Rozin and colleagues (1999) demonstrated that violations of loyalty and group solidarity norms (the 'Community code') were primarily associated with the emotion of contempt, violations involving harmful actions (the 'Autonomy code') were associated with anger, and violations concerning purity (the 'Divinity code') were associated with disgust (hence the name of the model, CAD – Community-contempt, Autonomy-anger, Divinity-disgust).

People's judgments of wrongness and disgust are affected by both content and intentionality and such factors interact in a way consistent with MFT. Young and Saxe (2011) found that intentional purity violations, such as incest, are judged as more wrong, but not more disgusting, than accidental purity violations; and that purity violations are judged as more disgusting, but not more wrong, than harm violations. Converging evidence comes also from studies on spontaneous facial expressions, which reveal links between harm violations and expressions of anger, and between purity violations and expressions of disgust (Cannon, Schnall, & White, 2011). Some functional brain imaging studies also found that moral violations, such as incest and pathogen-related actions, activate brain regions that have been associated with disgust (e.g., Parkinson et al., 2011; Schaich Borg, Kahn, Sinnott-Armstrong, Kurzban, Robinson, & Kiehl, 2013). Rottman, Kelemen, and Young (2014) found that the condemnation of suicide was significantly associated with trait disgust and disgust reported in reading obituaries, regardless of political orientation. Perceived harm did not predict judgments of suicide, suggesting that suicide

For Peer Review Only

is considered immoral not because of its harming consequences on close relatives, but primarily on grounds of norms related to purity.

Despite the empirical support for MFT, the existence of specialized moral modules and of non-harmful moral violations remains controversial. First, numerous results run against the predictions of MFT. For example, Gutierrez and Giner-Sorolla (2007) reported that harm is frequently associated with certain purity violations such as consensual incest. However, they proposed that such concerns may reflect post hoc rationalizations and therefore may be consistent with the social-intuitionist model (Haidt, 2001).

Gray, Schein, and Ward (2014) challenged this conclusion by reporting evidence suggesting that harm is automatically activated by purity violations. More specifically, they presented participants with instances of purity violations (e.g., covering a bible with feces) and non-moral negative actions (e.g., a student failing an exam). Following each action, they presented participants with a Chinese character and asked them to indicate whether its meaning was harmful, sad, or wrong. Relative to neutral actions, both negative non-moral actions and moral violations increased the sadness scores, but only moral violations increased also harmfulness and wrongness scores.

Another challenging evidence for MFT comes from Royzman, Atanasov, Landy, Parks, and Gepty (2016). These authors found that pathogen-free violations (related to the 'Divinity code') were more strongly linked to anger than to disgust. Indeed, according to a recent review on emotion-moral content studies (Cameron et al., 2015) the available evidence provides little support for purity-disgust and harm-anger correspondences. Cameron and colleagues demonstrate that most of the published studies do not report exclusive emotion-moral content links and the 25 studies that report such links fail short of ruling out that they are not due to global factors such as core affect, typicality or severity.

Can such domain-general factors account for effects previously attributed to moral content? Gray and Keeney (2015) claimed that scenarios commonly used to instantiate purity

For Peer Review Only

violations portray actions that are highly atypical (e.g., having a tail attached by plastic surgery) and are, on average, less severe violations than those used to instantiate harm violations (e.g., armed assault). While this may be a common aspect of purity violations, especially in cultures dominated by harm-based ethics (Graham, 2015), this raises the possibility that domain-general aspects, rather than moral content per se, are responsible for effects previously thought to support MFT. Gray and Keeney (2015) found that when controlling for such variables, the differences in responses to purity and harm violations vanished. In their study, the focus was on an effect originally found by Uhlmann and Zhu (2014): Harm violations are rated as more immoral than purity violations, but purity violations are judged as more indicative of an agent's immoral character than harm violations (see also Giner-Sorolla & Chapman, 2017). Gray and Keeney orthogonally manipulated moral content, severity, and typicality, and found that the effect due to moral content disappeared in a new set of violations – which they called 'naturalistic' – that had been generated by lay people rather than by the experimenters.

The aim of the present study was to investigate the association between type of moral violation and type of emotional reaction by focusing on the facial expressions elicited by harm and purity violations. Two types of scenarios were employed: classic MFT scenarios that have been extensively used in previous studies, and the new set of more 'naturalistic' scenarios which has been developed by Gray and Keeney (2015). In the naturalistic scenarios, the harm and purity violations were matched for weirdness. We employed both MFT and naturalistic scenarios to examine whether they would yield similar patterns of results. According to MFT, purity and harm scenarios should evoke differential affective responses because specific moral concerns are associated with specific emotional states. More specifically, anger should be the expression most frequently triggered by harm scenarios whereas disgust the expression most frequently triggered by scenarios depicting violations of purity (e.g., Cannon et al., 2011; Rozin et al., 1999). We name this the strong MFT prediction. We also considered a weaker MFT prediction according to which,

For Peer Review Only

relative to harm scenarios, purity scenarios trigger relatively fewer expressions of anger and relatively more expressions of disgust.

Apart from anger and disgust, we also investigated the presence of contempt, smiling and surprise expressions. Contempt is considered to be the most subtle and coldest of the triad of hostility emotions (Izard, 1977; Rozin et al., 1999). Ekman (1994) considered contempt as 'disapproving of' and 'feeling morally superior to' someone. It is usually claimed to involve a negative evaluation of others and their actions and is associated with the violations of loyalty and group solidarity norms, not with violations of individual rights or purity (Rozin et al., 1999). MFT makes no clear prediction about contempt reactions in the two types of violations, apart that contempt should not be the predominant reaction in either case. The main reason we decided to examine contempt is that it is a crucial emotional response in the CAD model, associated selectively with violations of the 'community morality' norms (Rozin et al., 1999).

Turning to smiling expressions, Cannon et al. (2011) found that some highly negative reactions to purity violations were associated with an increased activity of the zygomaticus muscle (which is associated with smiling). The authors explained this effect as the result of cross-talk activity with the levator muscle that is activated in extreme disgust facial expressions (e.g., Vrana, 1993). However, they also discussed an alternative possibility: "participants may have found some of the more extreme purity behaviors amusing as well as disgusting" (Cannon et al., 2011, p. 330). This is plausible. According to the benign-violation hypothesis (McGraw & Warren, 2010), to elicit humor a situation must be appraised as a violation and, simultaneously, must be perceived as benign. According to McGraw and Warren (2010), situations that elicit humor range from apparent physical threats to violations of personal dignity (e.g., physical deformities), linguistic norms, social norms (e.g., strange behavior), and—related to the present purposes—moral norms (e.g., bestiality or disrespectful behaviors). Because purity violations have been found to trigger smiles, we also examined the presence of smiling expressions.

For Peer Review Only

In addition, given that the traditional purity scenarios are regarded as highly unusual or atypical (Gray & Keeney, 2015), it seems plausible that they might evoke surprise. Thus, we also investigated the presence of surprise expressions. As it was the case with contempt, MFT does not make a prediction about smiling or surprise expressions other than neither should be the most prominent expression in response to either type of violation.

Finally, the reason we decided to study facial expressions, instead of relying on emotional scale ratings, is twofold. First, participants may use and understand emotion words in different ways than the one intended by the experimenter. For example, people may use the word 'disgust' to refer to a state of irritation or anger (Nabi, 2002), or to achieve a rhetorical effect (Bloom, 2004). Therefore, facial expressions provide a less ambiguous indicator of emotion than ratings on emotion scales. Second, evidence from facial expressions is less open to interpretations such that a particular response may be a product of a post hoc rationalization (see e.g., Gutierrez & Giner-Sorolla, 2007). This is especially true in the present research in which we recorded the first spontaneous facial expressions in response to vignettes depicting moral violations—note that participants were simply asked to listen to the scenarios, they were not asked to morally evaluate the depicted actions.

Method

Participants

Thirty-three Italian native speakers (17 females, 16 males, M_{age} = 24.79, age range = 18–38) participated in the study. Participants were recruited via an announcement through the University of Trento mailing list. Two additional participants were excluded from the analysis due to the low quality of the video recordings (i.e., participants were not centered on the camera focus and therefore their facial expressions could not be recorded clearly). The study protocol was conducted in accordance to the principles expressed in the Declaration of Helsinki.

We conducted a power analysis for the within-groups comparison (i.e., a t-test for one sample case) using G-Power (Faul, Erdfelder, Lang, & Buchner, 2007). To detect a medium

For Peer Review Only

effect size (i.e., Cohen's $d = 0.61$, based on Cannon et al., 2011) with alpha set at .05 and a power of .80, a minimum sample size of 24 participants was required. With our final sample size of 33 participants, we therefore expect a power of .92.

Materials and Procedure

The experiment was conducted in a quiet room. Participants were tested individually after being informed about the general aims of the study and having signed a consent form. Each participant was asked to complete two tasks, which were separated by a short break. In the first task, participants were asked to listen to 20 statements describing negative moral actions while their spontaneous facial expressions in response to these statements were video-recorded. In the second task, participants were asked to listen once again to the 20 statements and to answer to a series of questions that were presented in a booklet.

Task 1. Participants were told that the aim of the study was to examine cognitive processes while listening to statements about other individuals, but were not informed about the specific hypothesis. Participants were informed that a video camera (Sony Handycam HDR-SR5) recorded their upper part of the body. The camera was placed just below the monitor, at a distance of 50 cm from the participants. Facial expressions were recorded during the entire session. Participants simply had to look on the screen, read some instructions on it, and listen to all the statements.

Specifically, participants listened to 20 statements: 5 depicting commonly used Moral Foundations Theory (MFT) harm violations, 5 MFT purity violations (see Graham, Haidt, & Nosek, 2009), 5 naturalistic harm violations, and 5 naturalistic purity violations (see Gray & Keeney, 2015). The statements were presented in a pseudo-randomized order (see Table 1, for the full list of the moral violations).

The presentation of the statements was computer-based and was created using the open-source program OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). After a fixation cross that appeared for 2 s, a blank screen appeared and a statement concerning a moral violation

For Peer Review Only

was presented orally (male adult voice). The presentation of moral violations lasted between 4 and 15 s. At the end of each moral violation, the written statement "Please think about this action" appeared on the computer screen for 4 s. Each trial ended with a blank screen (presented for 250 ms), which was followed by the instruction 'Please wait!' for 4 s between trials (see Figure 1, for a schematic representation). During this task, participants were not aware that they would later be asked to make judgments about these statements.

--INSERT FIGURE 1 ABOUT HERE--

Task 2. After a short break, the statements used in Task 1 were presented to the same participants again but in a different pseudo-randomized order. After listening to each statement, participants were asked to respond, in a booklet, to the following questions (Gray & Keeney, 2015): (1) "How morally wrong is this action?" (1 = *not at all wrong*, to 7 = *very wrong*), (2) "How severe is this action?" (1 = *not at all severe*, to 7 = *very severe*), (3) "How atypical (bizarre, weird, odd) is this action?" (1 = *not at all atypical*, 7 = *very atypical*), (4) "How harmful (this implies physical and/or emotional) is this action? (1 = *not at all harmful*, to 7 = *very harmful*), and (5) "How impure is this action?" (1 = *not at all impure*, to 7 = *very impure*).

At the end of the experimental session, participants were asked to guess what was the purpose of the experiment. No one guessed the real purpose of the study. The experiment lasted about 20 minutes.

Coding of facial expressions

We relied on a well-established objective coding system developed for the analysis of facial expressions, the Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002a). This system allows the analysis of minimal units of facial activity, known as action units (AUs), which are anatomically separate and visually distinguishable. We coded the presence of AUs that are diagnostic of disgust, anger, contempt, smiling and surprise expressions, but also certain additional AUs that could help disambiguate between the emotions of interest and other emotions, namely sadness and fear. Specifically, we coded the following action units: AU1

For Peer Review Only

(inner brow raiser), AU2 (outer brow raiser), AU4 (brow lowered), AU5 (upper lid raiser), AU6 (cheek raiser), AU7 (lids tight), AU9 (nose wrinkle), AU10 (upper lip raiser), AU11 (nasolabial furrow deepener), AU12 (lip corner puller), AU13 (sharp lip puller); AU14 (dimpler), AU15 (lip corner depressor), AU16 (lower lip depress), AU17 (chin raiser), AU18 (lip pucker), AU20 (lip stretch), 22 (lip funneler), AU23 (lip tightener), AU24 (lip presser), AU28 (lip suck), AU25 (lips part), AU26 (jaw drop), AU27 (mouth stretch). Below we present the scheme that we used to classify the AUs into emotional expressions.

The facial expression of *disgust* was classified by the presence of AU9 (nose wrinkle), AU10 (upper lip raise), or the combination of AU9 + AU10 (Ekman, Friesen, & Ancoli, 1980; Smith, 1989). We recognize that AU10 is an ambiguous emotional indicator for disgust: bilateral AU10 could also indicate anger, while unilateral AU10 could also indicate contempt (Ekman & Friesen, 1986; Rozin et al., 1999). However, since our main aim was to assess the MFT theory, we followed Rozin et al. (1999, p. 584) and counted both types of AU10 (either in isolation or combined with AU9) as indicative of moral disgust. We also classified disgust by the presence of AU25 (lips part) but only if it appeared in combination with AU9 and/or AU10 (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & van Knippenberg, 2010; Rozin, Lowery, & Ebert, 1994).

The prototypical facial expression of *anger* involves the combination of frowning (AU4), lid tightening (AU7), and lip tightening/lip pressing (AU23/AU24). However, several additional variations of this expression have been reported (Durán, Reisenzein, & Fernández-Dols, 2017). Following previous research, we classified anger by the following AU combinations: AU4 + AU7 + AU23/AU24, AU4 + AU7, AU4 +AU5, AU7 + AU5, AU7 +AU23, AU7 + AU24, AU7 +AU17 (e.g., Durán et al., 2017; Ekman, Friesen, & Hager, 2002b; Langner et al., 2010; Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008; Rosenberg, Ekman, & Blumenthal, 1998; Rozin et al., 1999; Sayette & Hufford, 1995; Smith, 1989; Wiggers, 1982). Furthermore, we also classified anger by the presence of the following AUs in isolation: AU4 (brow lowered;

For Peer Review Only

Cannon et al., 2011; Smith, 1989; Wiggers, 1982), AU7 (lids tight; Smith, 1989), AU23 (lip tightened; Alvarado & Jameson, 2002; Ekman et al., 2002b; Wiggers, 1982), and AU24 (lip presser; Alvarado & Jameson, 2002; Rozin et al., 1999, p. 579).

Unlike disgust, the coding of anger is problematic because certain single AUs or AU combinations that are associated with anger are also associated with sadness and fear (e.g., AU4 is associated with all three emotions, while AU4+AU5 is associated with both anger and fear). For this reason, we coded certain additional AUs—AU1, AU2, AU15 and AU20—whose presence may indicate sadness or fear (Durán et al., 2017; Matsumoto et al., 2008). Specifically, the facial expression of sadness was classified by the presence of AU1+AU4+AU15, whereas the facial expression of fear by the presence of AU1/2+AU4+AU5+AU20 (Durán et al., 2017).

The facial expression of *contempt* was classified by the presence of the unilateral AU14 (dimple) (Ekman & Friesen, 1986; Matsumoto, 1992; Rozin et al., 1999). The *smiling* facial expression was classified by the presence of AU12 (lip corner pull) or AU6 + AU12 (cheek raise with lip corner pull) (Ekman et al., 1980; Ekman, Davidson, & Friesen, 1990; Sayette & Hufford, 1995; Smith, 1989). Finally, the facial expression of *surprise* comprises three components: eyebrow raising (AU1, AU2), eye widening (AU5), and mouth opening (AU25) (Durán et al., 2017; Reisenzein, 2000). We classified this expression by the following combination of AUs—AU1+AU2+AU5+AU25. Note that the expression of surprise shares several AUs (i.e., AU1, AU2, AU5) with the expression of fear.

The discussion of the anger and surprise expressions indicates a potential problem of classification ambiguity as these facial expressions share action units with sadness and fear. To deal with such ambiguities, we employed a similarity-based rule.[1] Whenever the facial expression displayed by a subject in response to a particular item fully matched one emotion (i.e., it contained all AUs associated with a variant of that emotion) while only partially matched another, we classified it as an expression of the former emotion. For example, the combination

[1] We thank an anonymous reviewer for this suggestion.

For Peer Review Only

AU4+AU7+AU20+AU24 was classified as an expression of anger as it included all the features required in one variant of anger (AU4+AU7+AU24) while only partially matched the expression of fear (it matched two of its components [AU4, AU20] but also missed two [AU1/2, AU5]). Whenever a facial expression partially matched two or more emotions then we counted it as an expression of the emotion that it matched more closely. For example, the combination AU1+AU15 was classified as an expression of sadness (as it misses one component for the facial expression of sadness, while it misses three components for the facial expressions of surprise or fear). In cases where some AU/AUs that are not core components of an emotion were present in isolation (e.g., AU18; AU26+AU28), we did not classify them as an expression of a particular emotion.

A certified FACS rater watched all video-recordings. Forty-two percent of the data were also comparison coded by an independent certified FACS coder. Both coders viewed videotapes in slow motion using VLC media player on a laptop and listed the presence of every single AU of interest on a coding sheet. According to MFT, violations automatically activate distinct emotions. Therefore, we focused on the first facial expression, and consequently the first AUs, among those of our interest, that a participant displayed immediately after listening to a violation. Our window of analysis started with the oral presentation of a moral violation and continued until the instruction "Please think about this action" had disappeared from the screen (see Figure 1).

The interrater reliability (Cohen's Kappa) was based on the classification of the AUs, and it was calculated on 42.4% of all data-points (280 out of 660 items). This analysis showed an excellent consistency between the two independent coders for the classification of AU1, AU2, AU4, AU6, AU10, AU15, AU17, AU18, AU20, AU23, AU24, AU26, AU28 ($\kappa > .80$), and a good consistency for the classification of AU5, AU7, AU9, AU12, AU14, AU25 ($.65 < \kappa > .74$) (e.g., Fleiss, 1981). Inconsistencies were resolved through discussion between the coders.

Results

For Peer Review Only

Facial expressions (Task 1). We first examined the emotional expressions elicited by listening to the scenarios. MFT predicts that anger would be the most frequent reaction to harm violations, and disgust the most frequent reaction to purity violations. Consistent with this prediction, in MFT harm scenarios anger was the most frequent expression. It was more frequent than disgust ($t(32) = 7.21$, $p < .001$), smiling expressions ($t(32) = 5.79$, $p < .001$), and contempt ($t(32) = 4.80$, $p < .001$). However, in MFT purity scenarios disgust was not the most frequent expression. It was more frequent than contempt ($t(32) = 2.70$, $p = .011$), but less frequent than anger ($t(32) = 2.90$, $p = .007$) and smiling expressions ($t(32) = 3.82$, $p = .001$). Expressions of surprise were rare, and therefore excluded from the analyses.[2]

Analyses of the naturalistic scenarios produced similar results. As predicted by MFT, anger was the most frequent expression in naturalistic harm scenarios. It was more frequent than smiling ($t(32) = 5.23$, $p < .001$), disgust ($t(32) = 4.99$, $p < .001$), and contempt ($t(32) = 2.81$, $p = .008$). However, the disgust expression was the least frequent expression in response to purity scenarios. There were fewer expressions of disgust than of anger, smiling or contempt (respectively, $t(32) = 5.66$, $p < .001$; $t(32) = 6.49$, $p < .001$; $t(32) = 2.73$, $p = .010$).

We then examined the weaker version of MFT (i.e., the hypothesis that in comparison to harm scenarios, purity scenarios trigger less anger and more disgust) by comparing the presence of anger and disgust expressions across harm and purity scenarios. As before, we first focused on MFT and then on naturalistic scenarios. As predicted by weak MFT, the anger expression was more frequently exhibited in response to MFT harm scenarios than MFT purity scenarios ($t(32) = 2.06$, $p = .048$), while the disgust expression was more frequently exhibited in response to MFT purity scenarios than MFT harm scenarios, $t(32) =$

[2]The expressions classified as surprise were rare. However, there were several cases where one or two components of the surprise expression was present (e.g., AU1, AU1+AU2) (see Table 2). Because such cases are difficult to classify in terms of our similarity-based scheme (as the match with surprise was not higher than that with sadness or fear), we decided to leave them unclassified.

For Peer Review Only

3.08, $p = .004$ (see Figure 2a). In additional analyses, we found that contempt showed the same trend as the anger expression, that is, it was more frequently exhibited in MFT harm scenarios than MFT purity scenarios ($t(32) = 3.436$, $p = .002$). In contrast, smiling expressions followed the opposite pattern: they were more frequent in response to MFT purity scenarios than MFT harm scenarios ($t(32) = 5.50$, $p < .001$, see Figure 2a).

However, the analyses of naturalistic scenarios produced less favorable results for weak MFT. Contrary to its predictions, no difference was found in the frequency of anger and disgust (or contempt) expressions across harm and purity scenarios (see Figure 2b). The only significant difference between naturalistic harm and naturalistic purity scenarios was in smile expressions: smiles were more frequent in response to purity scenarios than in response to harm scenarios, $t(32) = 5.75$, $p < .001$ (see Figure 2b).

--INSERT FIGURE 2 ABOUT HERE--

Tables 2 and 3 show the frequency of action units and facial expressions of interest, respectively, elicited by specific items. As can be seen in Table 3, among the MFT harm scenarios, items HM1, HM2, HM5 elicited selectively the expression of anger. Among MFT purity scenarios, only item PM2 most frequently elicited the expression of disgust. Item PM5 elicited mostly smiles, PM4 elicited anger, while PM1 and PM3 elicited equally smiles and anger. So, contrary to strong MFT, only one of the purity scenarios selectively elicited disgust.

Turning to naturalistic harm scenarios, items H3, H4 elicited selectively the expression of anger. None of the naturalistic purity violations predominantly elicited disgust expressions. Items P2, P3, and P5 elicited mostly smiles, P4 elicited anger and P1 elicited equally smiles, anger and contempt. In sum, in both MFT and naturalistic purity scenarios the expressions of disgust were rare: instead, there were frequent smiles and expressions of anger.

The use of single AUs as indicators of anger is potentially problematic because it makes our similarity-based rule to favor anger classifications. In an attempt to address this issue, in Table 4 we distinguish between anger expressions classified on the basis of a

For Peer Review Only

combination of AUs (and thus are less ambiguous) from those classified on the basis of a single AU (and thus are more ambiguous). Note that the pattern of results remains essentially unchanged if we replace the 'Anger' column of Table 3 with the 'Combination' column of Table 4. Specifically, the emotions that originally emerged as predominant in each of the four types of scenarios, also emerge as predominant even if we restrict anger expressions to cases where a combination of AUs was present. This is also the case if we consider instead individual items.

--INSERT TABLE 1, TABLE 2, TABLE 3 AND TABLE 4 ABOUT HERE--

It is noteworthy to mention that Ekman and Friesen (2003, Chapter 7) have proposed an even more restrictive classification of the facial expression of anger, which they call 'the full anger expression'. According to this classification, a simple combination of AUs is not sufficient to classify an expression as anger; in addition, the facial signals must be present in all three facial areas ("the brow/forehead; the eyes/lids and root of the nose; and the lower face, including the cheeks, mouth, most of the nose, and chin", p. 28). The AU combinations that we observed with this characteristic were: AU4+AU7+AU23/AU24/AU17; AU4+AU5+AU17. Even if we follow this highly restrictive classification scheme, the pattern of results with respect to the hypotheses of interest is similar to that observed from AU combinations (Table 4). In relation to the MFT scenarios, there were more anger expressions for the harm versus the purity items (full facial expression: 10 vs. 3; all AU combinations: 41 vs. 32), whereas in relation to the Naturalistic scenarios, we observed an equal number of anger expressions across the harm and purity items (full facial expression: 2 vs. 3; all AU combinations: 21 vs. 20).

Moral judgments (Task 2). We expected to replicate the findings of Gray and Keeney (2015). Specifically, we hypothesized that, compared to MFT harm scenarios, MFT purity scenarios would be rated as less severe and weirder. We further expected that MFT harm scenarios would be perceived as more harmful and, strangely, more impure than MFT purity

For Peer Review Only

scenarios. We examined these hypotheses via an analysis of variance (ANOVA) with Content (Harm vs. Purity) as a within-subjects factor on four dependent variables: Severity, Weirdness, Harm, and Impurity.

We found that, with respect to MFT purity scenarios, MFT harm scenarios received higher severity ratings, $F(1, 32) = 50.18$, $p < .001$, $\eta_p^2 = .61$, lower weirdness ratings, $F(1, 32) = 44.99$, $p < .001$, $\eta_p^2 = .58$, and higher harmfulness ratings, $F(1, 32) = 66.81$, $p < .001$, $\eta_p^2 = .68$. Strangely, but in line with Gray and Keeney (2015, Study 1), MFT harm scenarios also received higher impurity ratings than MFT purity scenarios, $F(1, 32) = 18.18$, $p < .001$, $\eta_p^2 = .36$. Harm and impurity ratings were highly correlated in the scenarios, $r(8) = .82$, $p = .004$. The findings are fully in line with Gray and Keeney's (2015). MFT purity scenarios were perceived as less severe, less harmful, less impure, and weirder than MFT harm scenarios (see Figure 3).

We next compared the two types of scenarios, MFT versus naturalistic, with respect to weirdness ratings, moral judgment, severity ratings, impurity ratings, and harmfulness ratings. We found that MFT scenarios were judged as more weird than naturalistic scenarios, $t(32) = 7.06$, $p < .001$. In particular, MFT purity scenarios were judged as more weird than their counterpart naturalistic scenarios, $t(32) = 9.34$, $p < .001$ (see Figure 3). These findings are in line with Gray and Keeney's (2015) results. MFT scenarios were also judged as less severe than naturalistic scenarios, $t(32) = 2.21$, $p = .034$. However, analyzing separately harm and purity scenarios, no significant differences were found. Thus, neither MFT harm nor MFT purity scenarios were judged as less severe than their naturalistic scenario counterparts (see Figure 3). In this case, only the general *t* test output is in line with Gray and Keeney's (2015) results. Finally, MFT scenarios were judged less morally wrong, impure, and harmful than naturalistic scenarios, $t(32) = 2.91$, $p = .006$; $t(32) = 3.55$, $p = .001$; $t(32) = 2.78$, $p = .009$, respectively. Moreover, when we conducted paired-sample *t* tests for harm and purity, naturalistic and MFT scenarios, we found that only MFT purity scenarios were judged as less morally wrong, impure

For Peer Review Only

and harmful than their naturalistic counterparts, $t(32) = 3.71$, $p = .011$; $t(32) = 2.61$, $p = .014$; $t(32) = 3.58$, $p = .001$ (Figure 3). In sum, the results of the moral judgment task provide a useful cross-cultural replication of Gray and Keeney (2015, Study 2).

--INSERT FIGURE 3 ABOUT HERE--

Discussion

We tested a strong and a weak claim of the moral foundation theory. According to the strong claim, harm scenarios predominantly elicit anger reactions whereas purity scenarios predominantly elicit disgust reactions. According to the weaker claim, with respect to harm scenarios, purity scenarios elicit fewer anger reactions and more disgust reactions. Participants listened to scenarios depicting harm or purity violations, and their spontaneous facial expressions were analyzed. The scenarios included classic MFT scenarios (e.g., Haidt et al., 1993), and more naturalistic scenarios, in which the harmful and impure actions were matched for typicality ('weirdness', Gray & Keeney, 2015). The results—at least a straightforward interpretation of them—do not support either MFT prediction.

Specifically, we found that the expression of anger was elicited equally often by harm and purity violations, and this was particularly true for naturalistic scenarios. Disgust reactions were rare. In line with weak MFT, disgust reactions were more frequent in response to purity than to harm violations, but only for MFT scenarios. But even for MFT purity scenarios, contrary to strong MFT, disgust expressions were less frequent than anger or smiling expressions. For naturalistic scenarios, disgust expressions were extremely rare for both harm and purity violations. The higher frequency of disgust expressions in response to MFT purity violations could be due to the fact that these violations included physically disgusting elements. The only significant difference between harm and purity violations was in the frequency of smiling expressions: purity violations triggered significantly more smiles than harm violations. Intriguingly, this was the case both for MFT and naturalistic scenarios. Had this been the case only for MFT purity scenarios, one could attribute it to their weirdness.

For Peer Review Only

Why did participants smile more in response to purity than harm violations? The versatile nature of the human smile was highlighted in several studies (e.g., Ekman,1992; Keltner, 1995; Rychlowska, Jack, Garrod, Schyns, Martin, & Niedenthal, 2017). In our case, one possibility is that smiles indicate covert disgust. Under this interpretation, the present results could still be consistent with both the strong and weak MFT. Notice that for purity scenarios (both MFT and naturalistic) the sum of smiling and disgust expressions was higher than that of anger expressions (inspect Figure 2). Another, more plausible, explanation is that participants found purity violations more amusing and less threatening than harm violations (see McGraw & Warren, 2010). This explanation appears to be consistent with a domain general proposal that considers purity violations as involving personal harm, but in a more subtle and indirect way than harm violations. Notice, however, that there were also more smiles in response to naturalistic purity versus naturalistic harm scenarios, which were balanced in terms of perceived weirdness. Thus, a domain-general proposal cannot easily explain the observed association between smiles and purity violations.

Smiles might be associated with pleasure, relief, or amusement (enjoyment smiles), but non-enjoyment smiles also exist. Theorists have identified various types of non-enjoyment smiles such as ‘embarrassed smiles’, ‘masking smiles’ (while experiencing negative emotions, the smile at least partially covers muscular movements associated with another emotion), ‘false smiles’ (appearing as if enjoyment is felt when it is not), ‘anticipatory smiles’ or ‘miserable smiles’ (representing the willingness to endure unpleasant circumstances) (Ekman, 1985; Ekman & Friesen, 1982). These different forms of non-enjoyment smiles could be associated with compliance, embarrassment, shame, grin-and-bear-it (Ekman,1992; Keltner, 1995; Keltner & Buswell, 1996).

To which types of smile do the smiles we found belong to? Although there are no standardized procedures for classifying different types of non-enjoyment smiles, there is a wide agreement in the literature that when the lip corner raising is accompanied by the cheek raising

For Peer Review Only

(commonly known as a Duchenne smile) the smile denotes enjoyment (Ekman et al., 1990). In view of that, we reanalyzed our data on purity scenarios classifying smiles as either Duchenne smiles or non-Duchenne smiles (or simple smiles). We counted about twice as many Duchenne smiles than simple smiles in response to both MFT scenarios (31 Duchenne, 18 simple) and naturalistic scenarios (24 Duchenne, 14 simple). Thus, consistent with McGraw and Warren (2010), these results suggest that participants found some purity violations amusing.

The discussion on smiles highlights two limitations of the present study. The first is that emotional responses could be complex, that is, they could involve a combination of different emotions (e.g., Gross & Levenson, 1995; Kuppens, Tuerlinckx, Russell, & Barrett, 2013). To tackle this critical point, future research could investigate whether the present findings hold also with implicit physiological measures and other behavioral responses. Moreover, future research could use experimental paradigms involving ecological social contexts, which would prompt people to explicitly display spontaneous and stronger facial expressions in order to communicate a social message to others.

A second limitation of the present study concerns the use of facial expressions as an indicator of emotions. Emotions are not always accompanied by a facial expression and this is especially true for anger and disgust—the main emotions of interest in the present study—but less so for amusement (see e.g., Durán et al., 2017). Moreover, the fact that the participants knew that they were videotaped could have reinforced this tendency (Durán et al., 2017). Thus, a defender of MFT could argue that the purity scenarios did elicit mostly disgust, but people did not displayed it in their facial expressions. However, for this argument to go through, the defender of MFT must also assume that expressions of disgust were more strongly inhibited than expressions of anger. Therefore, this objection is not really strong.[3]

[3]We thank an anonymous reviewer for pointing this limitation to us and for discussing its implications for the MFT hypotheses.

For Peer Review Only

However, as this is a critical issue, we decided to address it directly in a follow up study (for details, see Supplementary materials). We presented a new sample of 34 participants with the 20 scenarios of the main study, but instead of focusing on facial expressions of emotions, we gathered self-ratings. Following each scenario, we asked participants to select the first emotion they felt while reading it ("While you were reading about this situation, what was your first emotion?") and to rate the intensity of that emotion ("Please express a judgment on the intensity of this emotion"). Participants could respond to the first question by selecting one of the following emotions: enjoyment, grin-and-bear-it, embarrassment, shame, disgust, anger, contempt, surprise. The reason we included enjoyment, grin-and-bear-it, embarrassment and shame was to help disambiguate the meaning of smiles in our main study (for a more detailed discussion, see Supplementary materials).

Notwithstanding the limitations of measuring emotions through self-reports (see Introduction), with respect to the MFT hypotheses, the results are largely convergent with those of the main study (see Figure S1).[4] For both MFT and naturalistic harm scenarios anger was the most frequent response, whereas for purity scenarios there was no clear prevalent emotion. In the MFT purity scenarios the most frequent response was disgust, but it was closely followed by surprise, enjoyment, and grin-and-bear-it. In the naturalistic purity scenarios the most frequent response was contempt, closely followed by grin-and-bear-it and embarrassment. Thus, the results do not support either MFT hypotheses. Strong MFT can explain the results for harm scenarios (anger was the predominant response) but not for purity scenarios (no emotion prevailed). Weak MFT can explain the responses to the MFT scenarios (harm scenarios elicited more anger than disgust, whereas purity scenarios elicited more disgust than anger), but not to

[4]In the Introduction, we raised two issues with respect to measuring emotion through self-reports: (a) certain emotional words might be ambiguous and (b) the results could be influenced by post hoc rationalizations. With regards to (a) we found that five participants expressed doubts about the meaning of certain emotional words (e.g., '*sdegno'*, the Italian word for contempt). With regards to (b) since we asked directly about emotional judgments—these judgments were not preceded by moral judgments—the impact of post hoc rationalizations should be limited.

For Peer Review Only

naturalistic scenarios (harm and purity scenarios elicited equal amounts of disgust). The low frequency of disgust expression in the new study further questions the possibility that participants in the main study experienced disgust, but they suppressed the facial expression of it because they were video-recorded.

Returning to the main study on facial expressions, the present pattern of results is also consistent with a semi-exclusive correspondences model, in which harm violations elicit anger, and purity violations elicit not only disgust but also other emotions (see Cameron et al., 2015, Figure 4 for an adaptation to our results of the schema proposed by Cameron et al., 2015). The results of the follow up study provide further support for this assertion (see Figure S1). This happens probably because harm violations instantiate clear and direct cases of personal harm, while purity violations are more complex, which, in turn, is reflected in the different emotions they elicit. For example, the purity items showing rape and adultery are instances of violations not only in the purity domain, but also in other domains such as harm and loyalty (Graham, 2015). Thus, it is of no surprise that, by using such 'mixed' items, one does not find the specific association effects predicted by MFT.

--INSERT FIGURE 4 ABOUT HERE--

This argument rests on the assumption that pure violations of purity exist, which is exactly what has been highly controversial from the beginning of this debate (Turiel, 1989). Domain general views deny the existence of harmless moral violations of purity and claim that all purity violations are perceived as potentially harmful (Royzman et al., 2016). The data reported by Gray and Keeney (2015) show that people do associate purity items, both naturalistic and MFT ones, to the harm domain, as claimed by the domain general view (on the lack of coherence and clarity in defining the domain of purity, see Russell & Giner-Sorolla, 2013). So, defenders of pluralist models should not assume, by simply relying on intuition, that some instances of purity violations are fine while others are not, but should instead provide an objective way of identifying such cases.

For Peer Review Only

Meanwhile, the use of alternative sets of scenarios that allow control over relevant factors, such as typicality, seems a fruitful research strategy.

In the present study we analyzed the emotional reactions that were immediately and spontaneously expressed while listening to moral violations, but we cannot specify at what stage of the information processing such reactions occurred. They could have occurred as a response to a covert judgment, or they could have been the basis upon which the judgment took place, along the lines of classic works in the empiricist tradition (Haidt et al., 1993; Hume, 1751/1957). This ambiguity does not allow us to determine which particular aspect of domain specificity is challenged by the present results. The results of the follow up study similarly do not shed light on this matter. Future research may wish to tackle this issue.

In conclusion, the present results go against MFT theory. Although anger was the most frequent expression in response to harm violations, disgust was not the most frequent response to purity violations. The present results also go against a weaker version of MFT theory. In the main study, the naturalistic harm and purity scenarios—which were matched in weirdness—were virtually identical in the amounts of anger, disgust, and contempt expressions they triggered. This result fits particularly well with monist claims that eventual differences between harm and purity scenarios are due to lack of control of several important factors such as weirdness. In the follow up study with self-ratings a slightly different pattern emerged: naturalistic harm and purity scenarios elicited equal amounts of disgust and contempt, but harm scenarios elicited more anger. However, one result in the present research keeps alive the possibility of distinct moral foundations: Smiles were selectively associated with violations of purity. Therefore, the general theory may be right—different violations have distinct emotional footprints—but the original claims about violation-emotion pairs may be wrong or incomplete. Violations of purity, like benign norm violations, may be characterized by the elicitation of amusement, but also by other emotions that may underlie the smiling expression. This opens exciting avenues for new research.

For Peer Review Only

References

Alvarado, N., & Jameson, K. A. (2002). Varieties of anger: the relation between emotion terms and components of anger expressions. *Motivation and Emotion, 26*, 153-182. doi:10.1023/A:1019815402873

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic.

Cannon, P. R., Schnall, S., & White, M. (2011). Transgressions and expressions: Affective facial muscle activity predicts moral judgments. *Social Psychological and Personality Science, 2*, 325-331. doi:10.1177/1948550610390525

Cameron C. D., Lindquist K. A., & Gray K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review, 9*, 371–394. doi:10.1177/1088868314566683

Cushman, F. (2013). Action, outcome, and value: a dual-system framework for morality. *Personality and Social Psychology Review, 17*, 273-92. doi:10.1177/1088868313495594.

Durán, J. I., Reisenzein, R., & Fernández-Dols, J.-M. (2017). Coherence between emotions and facial expressions. In J.-M. Fernández-Dols & J. A. Russell (Eds.), *The science of facial expression* (pp.107-129). Oxford University Press. doi:10.1093/acprof:oso/9780190613501.001.0001

Ekman, P. (1985). *Telling lies: Clues to deceit in the marketplace, marriage, and politics*. New York: W.W. Norton.

Ekman, P. (1992). Facial expressions of emotion: New findings, new questions. *Psychological science, 3*, 34-38. doi:10.1111/j.1467-9280.1992.tb00253.x

Ekman, P. (1994). Antecedent events and emotion metaphors. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 146-149). New York: Oxford University Press.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression

For Peer Review Only

and brain physiology: II. *Journal of Personality and Social Psychology, 58*, 2, 342-353.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6*, 238-252. doi:10.1007/BF00987191

Ekman, P., & Friesen, W. V. (1986). A new pancultural facial expression of emotion. *Motivation and Emotion, 10*, 159-168. doi:10.1007/BF00992253

Ekman, P., & Friesen, W. V. (2003). *Unmasking the face. A guide to recognizing emotions from facial expressions*. Cambridge, MA: Malor Books.

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*, 1125-1134. doi:10.1037/h0077722

Ekman, P., Friesen, W. V., & Hager, J. C. (2002a). *Facial Action Coding System (FACS)*. Salt Lake City, UT: Research Nexus division of Network Information Research Corporation.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002b). *Facial Action Coding System. Investigator's guide*. Salt Lake City, UT: Research Nexus division of Network Information Research Corporation.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. doi:10.3758/BRM.41.4.1149

Giner-Sorolla, R., & Chapman, H. A. (2017). Beyond purity: Moral disgust toward bad character. *Psychological Science*, *28*, 80–91. doi:10.1177/0956797616673193

Graham, J. (2015). Explaining away differences in moral judgment: Comment on Gray and Keeney (2015). *Social Psychological and Personality Science, 6*, 869-873. doi:10.1177/1948550615592242

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. H. (2013). Moral Foundations Theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology, 47*, 55-130. doi:10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of

For Peer Review Only

moral foundations. *Journal of Personality and Social Psychology, 96*, 1029–1046. doi:10.1037/a0015141

Gray, K., & Keeney, J. E. (2015). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science, 6*, 859-868. M doi:10.1177/1948550615592241

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. doi:10.1037/a0036149

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition and Emotion, 9*, 87-108. doi:10.1080/02699939508408966

Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion, 7*, 853-868. doi:10.1037/1528-3542.7.4.853

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834. doi:10.1037/0033-295X.108.4.814

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. doi:10.1037//0022-3514.65.4.613

Hume, D. (1957) *An inquiry concerning the principle of morals* (Vol. 4). New York: Liberal Arts Press. (Original work published 1751).

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology, 68*, 441-454. doi:10.1037/0022-3514.68.3.441

Keltner, D., & Buswell, B. N. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition and Emotion, 10*, 155-171. doi:10.1080/026999396380312

For Peer Review Only

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139*, 917-940. doi:10.1037/a0030811

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion, 24*, 1377-1388. doi:10.1080/02699930903485076

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314-324. doi:10.3758/s13428-011-0168-7

Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion, 16*, 363-368. doi: 10.1007/BF00992972

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones, & L. Feldman Barrett (Eds.), *Handbook of emotions* (pp. 211-234). New York: The Guilford Press.

McGraw, A. P., & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science, 21*, 1141-1149. doi:10.1177/0956797610376073.

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences, 11*, 143–152. doi:10.1016/j.tics.2006.12.007

Nabi, R. (2002). Anger, fear, uncertainty, and attitudes: a test of the cognitive-functional model. *Communication Monographs, 69*, 204-216. doi:10.1080/03637750216541

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience, 23*, 3162–3180. doi:10.1162/jocn_a_00017

Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition and Emotion, 14*, 1-38.

For Peer Review Only

doi:10.1080/026999300378978

Rosenberg, E. L., Ekman, P., & Blumenthal, J. A. (1998). Facial expression and the affective component of cynical hostility in male coronary heart disease patients. *Health Psychology, 17*, 376-380. doi:10.1037/0278-6133.17.4.376

Rottman, J., Kelemen, D., & Young, L. (2014). Tainting the soul: Purity concerns predict moral judgments of suicide. *Cognition, 130*, 217–226. doi:10.1016/j.cognition.2013.11.007

Royzman, E., Atanasov, P., Landy, J. F., Parks, A., & Gepty, A. (2016). CAD or MAD? Anger (not disgust) as the predominant response to pathogen-free violations of the Divinity code. *Emotion, 14*, 892-907. doi:10.1037/a0036829

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology, 66*, 870-881. doi:10.1037/0022-3514.66.5.870

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76,* 574-586. doi:10.1037/0022-3514.76.4.574

Russell, P. S., & Giner-Sorolla, R. (2013). Bodily moral disgust: what it is, how it is different from anger, and why it is an unreasoned emotion. *Psychological Bulletin, 139*, 328-351. doi:10.1037/a0029319

Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional smiles: Tools for love, sympathy, and war. *Psychological Science, 28*, 1259-1270. doi: 10.1177/0956797617706082

Sayette, M. A., & Hufford, M. R. (1995). Urge and Affect: A Facial Coding Analysis of Smokers. *Experimental and Clinical Psychopharmacology, 3*, 417-423. doi:1064-1297/95/$3.00

Schaich Borg, J., Kahn, R. E., Sinnott-Armstrong, W., Kurzban, R., Robinson, P. H., & Kiehl, K. A. (2013). Subcomponents of psychopathy have opposing correlations with punishment

For Peer Review Only

judgments. *Journal of Personality and Social Psychology, 105*, 667-687. doi:10.1037/a0033485

Smith, C. A. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology, 56*, 339-353. doi:0022-35I4/89/S00.75

Turiel, E. (1989). Domain-specific social judgments and domain ambiguities. Merrill- Palmer Quarterly, 35, 89-114.

Turiel, E., Killen, M., & Helwig, C. C. (1987). Morality: Its structure, functions, and vagaries. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 155-243). Chicago: University of Chicago Press.

Uhlmann, E. L., & Zhu, L. (2014). Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science, 5*, 279-285. doi:10.1177/1948550613497238

Vrana, S. R. (1993). The psychophysiology of disgust: Differentiating negative emotional contexts with facial EMG. *Psychophysiology, 30*, 279-286. doi:10.1111/j.1469-8986.1993.tb03354.x

Wiggers, M. (1982). Judgments of facial expressions of emotion predicted from facial behavior. *Journal of Nonverbal Behavior, 7*, 101–116. doi:10.1007/BF00986872

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition, 2*, 202–214. doi:10.1016/j.cognition.2011.04.005

For Peer Review Only

Table 1

Moral Foundations Theory (MFT) and naturalistic scenarios (respectively, see Graham et al., 2009, and Gray & Keeney, 2015).

	HARM	PURITY
MFT scenarios	HM1. Someone kicks a dog in the head, hard	PM1. Someone signs a piece of paper that says: " I hereby sell my soul after my death, to whoever has this piece of paper"
	HM2. Someone shoots and kills an animal that is a member of an endangered species	PM2. Someone cooks and eats their dog, after it dies of natural causes
	HM3. Someone makes cruel remarks to an overweight person about his or her appearence	PM3. Someone gets plastic surgery that adds a 2 inch tail onto the end of their spine
	HM4. Someone steps on an anthill, killing thousand of ants	PM4. Someone gets a blood transfusion of 1 liter of disease-free, compatible blood from a convicted child molester
	HM5. Someone sticks a pin into the palm of a child they don't know	PM5. Someone attends a performance art piece in which all participants (including that person) have to act like animals for 30 minutes, including crawling around naked and urinating on stage
Naturalistic scenarios	H1. Someone physically strikes another person	P1. Someone has an affair with another person, while they are married to someone else
	H2. Someone speaks to another person in a cruel and offensive manner	P2. Someone hires a prostitute for an evening of sex
	H3. Someone harasses and intimidates another person	P3. Someone has a sex on camera, making a pornographic film that will be distributed for profit
	H4. Someone steals a valuable item from another person	P4. Someone forces another person to have sexual intercourse with them, without that person's consent
	H5. Someone intentionally kills another person	P5. Someone strips nude on stage in a room full of strangers

For Peer Review Only

Table 2

Frequencies of the Action Units codified in correspondence with the Moral Foundations Theory (MFT) and the naturalistic scenarios.

MFT Harm scenarios																			
Item	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU18	AU20	AU23	AU24	AU25	AU26	AU28
HM1	1	1	12	1	1	7	0	0	4	1	1	1	0	0	3	0	1	0	0
HM2	3	3	10	0	0	6	0	0	0	5	0	4	1	1	3	1	0	0	0
HM3	1	1	6	1	0	2	0	0	1	9	0	0	0	0	0	0	0	2	2
HM4	4	4	7	1	3	3	1	0	6	0	1	2	0	0	2	0	0	0	0
HM5	2	2	23	1	0	18	1	2	0	2	1	2	0	1	3	1	0	0	0
Tot.	11	11	58	5	4	36	2	2	11	17	3	9	1	2	11	2	1	2	2
MFT Purity scenarios																			
Item	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU18	AU20	AU23	AU24	AU25	AU26	AU28
PM1	4	3	7	0	7	3	0	0	11	0	0	1	0	0	0	0	1	0	0
PM2	3	3	7	4	1	7	2	12	2	1	0	0	0	0	0	0	0	0	0
PM3	4	3	10	1	8	5	2	2	13	0	1	2	0	0	0	0	2	0	0
PM4	2	2	12	5	0	9	1	2	0	3	1	2	0	0	0	0	0	0	0
PM5	3	3	3	0	15	2	1	2	23	0	0	1	0	0	0	0	6	0	0
Tot.	16	14	39	10	31	26	6	18	49	4	2	6	0	0	0	0	9	0	0

For Peer Review Only

Naturalistic Harm scenarios																			
Item	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU18	AU20	AU23	AU24	AU25	AU26	AU28
H1	1	1	4	0	2	2	0	0	3	0	0	0	0	1	1	0	0	0	0
H2	2	1	5	0	0	3	0	0	0	5	0	0	1	0	1	0	0	2	2
H3	1	1	10	1	0	3	0	2	0	2	1	0	0	1	2	0	0	0	0
H4	3	3	6	1	0	4	1	1	0	3	0	3	0	0	4	0	1	0	0
H5	3	3	5	3	0	3	0	0	0	5	0	1	0	0	1	1	0	0	1
Tot.	10	9	30	5	2	15	1	3	3	15	1	4	1	2	9	1	1	2	3

Naturalistic Purity scenarios																			
Item	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU10	AU12	AU14	AU15	AU17	AU18	AU20	AU23	AU24	AU25	AU26	AU28
P1	5	4	3	0	2	4	0	0	4	5	0	3	0	0	1	1	0	1	1
P2	4	3	3	0	9	0	0	1	11	3	2	1	0	1	1	0	1	1	0
P3	3	3	1	2	8	1	0	0	12	2	1	1	0	0	0	1	3	2	1
P4	1	1	13	1	0	6	0	0	0	0	0	1	2	0	3	1	0	0	0
P5	5	3	7	0	5	4	0	0	11	1	1	1	0	0	0	0	1	1	0
Tot.	18	14	27	3	24	15	0	1	38	11	4	7	2	1	5	3	4	5	2

Note. Each AU was coded as present if it was visible in the slightest degree and bilaterally (except for AU10 and AU14). All AU10s reported in the table were bilateral. Although AU10 is indicative of both disgust and anger, we followed Rozin et al. (1999) and coded it as disgust. In several cases some AUs appeared in isolation or in an ambiguous combination with other AUs. Because the

For Peer Review Only

classification of these cases is problematic, we did not count them as indicative of an emotional expression, but we report them in the table. Here is a list of these cases: AU1 appeared alone twice; AU2 appeared in combination with only AU1 eighteen times; AU5 appeared in combination with only AU1+AU2 four times; AU15 appeared alone six times, and twice in combination with only AU1 or AU1+AU2+AU17; AU20 appeared alone once; AU18 appeared alone four times; AU26 appeared with only AU28 six times.

For Peer Review Only

Table 3

Frequencies of the facial expressions of interest for the Moral Foundations Theory (MFT) and the naturalistic scenarios.

	Harm						Purity					
	Item	Angry	Disgusted	Contemptuous	Smiling	Surprised	Item	Angry	Disgusted	Contemptuous	Smiling	Surprised
MFT scenarios	HM1	14	0	1	4	0	PM1	8	0	0	11	0
	HM2	12	0	5	0	0	PM2	9	12	1	2	0
	HM3	6	0	9	1	1	PM3	10	2	0	13	0
	HM4	7	1	0	6	1	PM4	13	2	3	0	0
	HM5	24	3	2	0	1	PM5	3	2	0	23	0
	Tot.	63	4	17	11	3	**Tot.**	43	18	4	49	0
Naturalistic scenarios	H1	5	0	0	3	0	P1	7	0	5	4	0
	H2	6	0	5	0	0	P2	3	1	3	11	0
	H3	10	2	2	0	0	P3	3	0	2	12	1
	H4	10	1	3	0	1	P4	16	0	0	0	0
	H5	6	0	5	0	2	P5	7	0	1	11	0
	Tot.	37	3	15	3	3	**Tot.**	36	1	11	38	1

For Peer Review Only

Note. Considering the high frequency of smiles elicited by purity scenarios, we reanalyzed the data distinguishing between two types of smile: the simple smile (non-Duchenne), consisting of pulling the lip corners back (AU12, without any other AU activity), and the Duchenne smile, defined by the combination of AU6 + AU12 (cheek raise with lip corner pull) (e.g., Ekman, 1992b). The Duchenne smile is thought to be associated with enjoyment, while non-Duchenne smiles are not (Ekman et al., 1990). In MFT purity scenarios there were 18 simple and 31 Duchenne smiles, while in naturalistic purity scenarios there were 14 simple and 24 Duchenne smiles. In regards to the emotional expressions which were not of our immediate interest, we counted 9 partial expressions of sadness (7 of these were based on the isolated presence of AU15) and 1 of fear (based on the isolated presence of AU20).

For Peer Review Only

Table 4

Additional information about the frequencies of the facial expression of anger. We differentiate between expressions of anger identified by a combination of AUs (less ambiguous) from those identified by the presence of a single AU (more ambiguous).

	Harm			Purity		
	Item	Combination	Single	Item	Combination	Single
MFT scenarios	HM1	7	7	PM1	4	4
	HM2	8$^{\Delta}$	4	PM2	7	2
	HM3	2	4	PM3	6*	4
	HM4	5*	2	PM4	12*	1
	HM5	19$^{*\Delta\dagger}$	5	PM5	3	0
	Tot.	41	22	**Tot.**	32	11
Naturalistic scenarios	H1	2	3	P1	5	2
	H2	3	3	P2	1$^{\Delta}$	2
	H3	4$^{\Delta}$	6	P3	2†	1
	H4	7	3*	P4	8*	8
	H5	5	1*	P5	4	3*
	Tot.	21	16	**Tot.**	20	16

Note. We observed the following AU combinations for anger: AU4+AU7+AU17/23/AU24; AU4+AU5+AU7/AU17; AU4+AU17+AU23/24; AU4+AU5/AU7; AU7+AU5/AU17/AU23/AU24. The single AUs we observed were: AU4; AU7; AU23; AU24. *Indicates the existence of a single ambiguous case where, in addition to the AUs that are characteristic of anger, we also observed the presence of AU1/AU2, which could be associated with fear (e.g., AU1+AU2+ AU4 and/or AU7). $^{\Delta}$Indicates the existence of a single ambiguous case where we also observed the presence of AU20, which could be associated with fear (HM2: AU4+AU17+AU20+AU24; HM5: AU4+AU7+AU20+AU24; H3: AU4+AU20+AU23; P2:

For Peer Review Only

AU4+AU17+AU20+AU23). †Indicates the existence of a single ambiguous case where we also observed the presence of AU15, which could be associated with sadness (HM5: AU4+AU7+AU15; P3: AU4+AU7+AU15+AU17).

For Peer Review Only

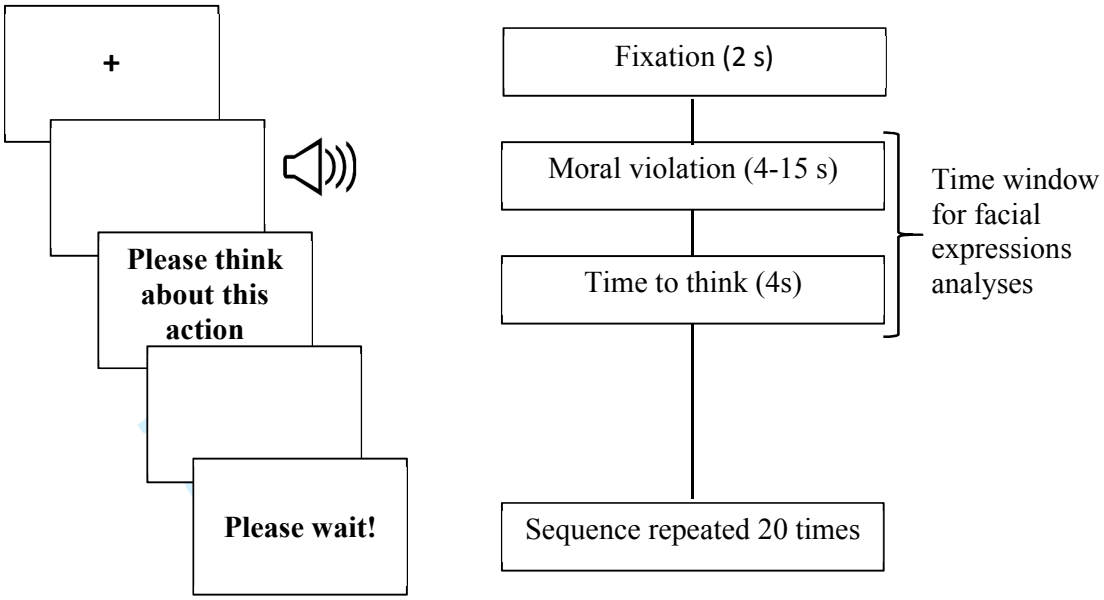


Figure 1. Schema of the experimental procedure of Task 1.

Peer Review Only

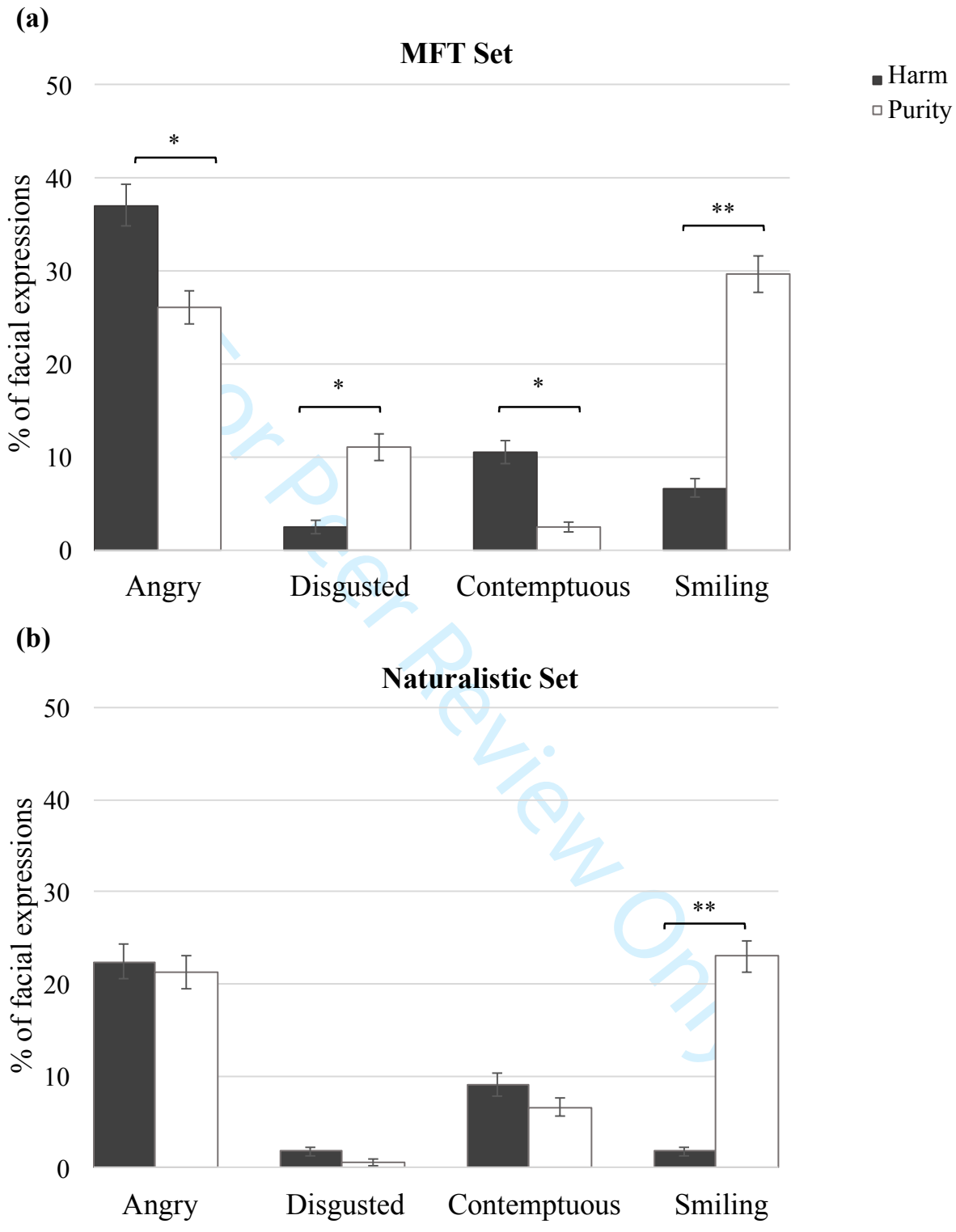


Figure 2. Mean percentage of facial expressions in response to the harm and purity scenarios, separately for the MFT set (a) and the naturalistic set (b). The surprise expression was not included due to its low frequency of exhibition. Error bars indicate standard error. *p<.005; **p≤.001

For Peer Review Only

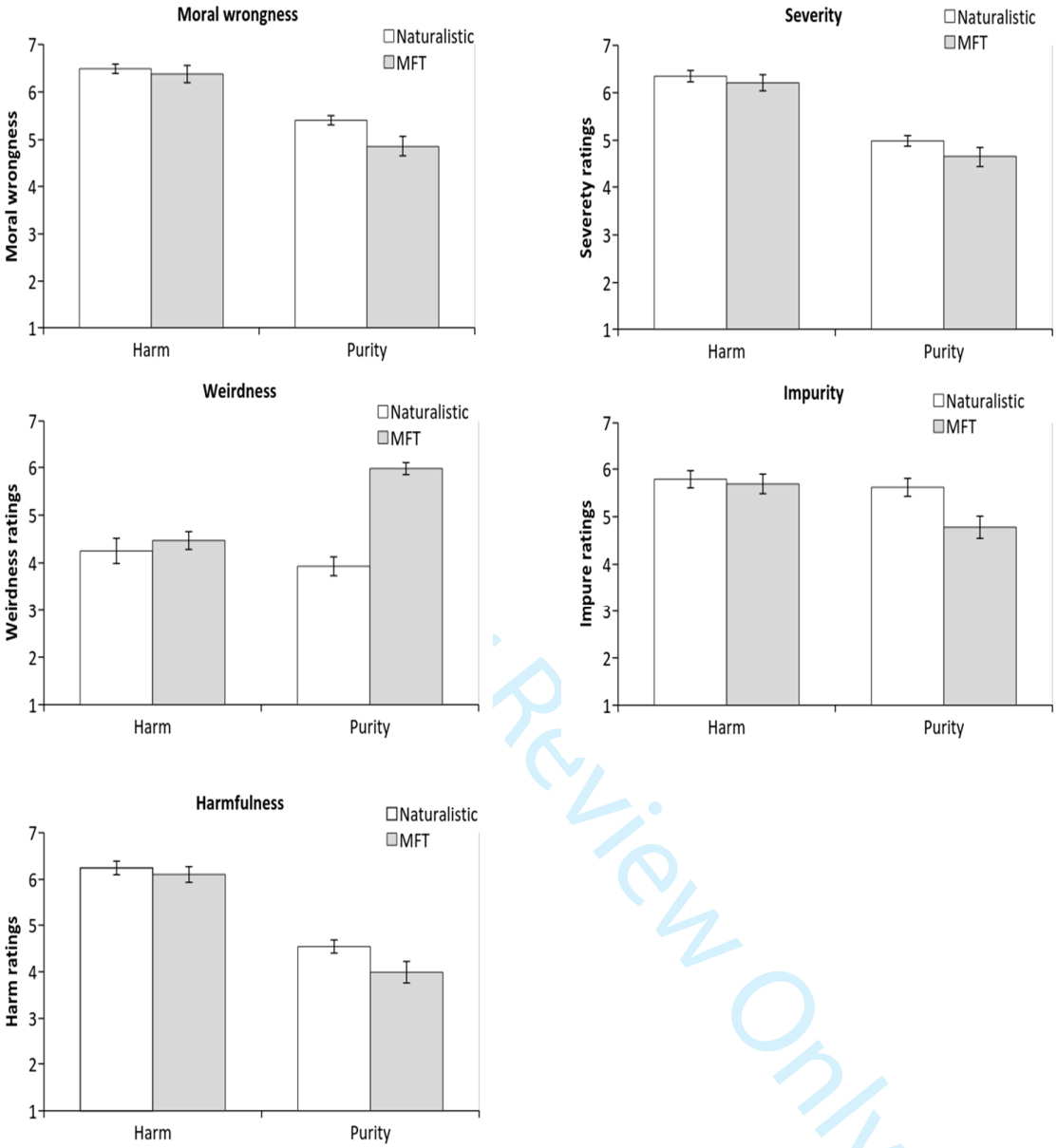


Figure 3. Moral wrongness, severity, weirdness, impurity, and harmfulness ratings by content (harm vs. purity) and source (MFT vs. naturalistic). Error bars indicate standard error of the means.

For Review Only

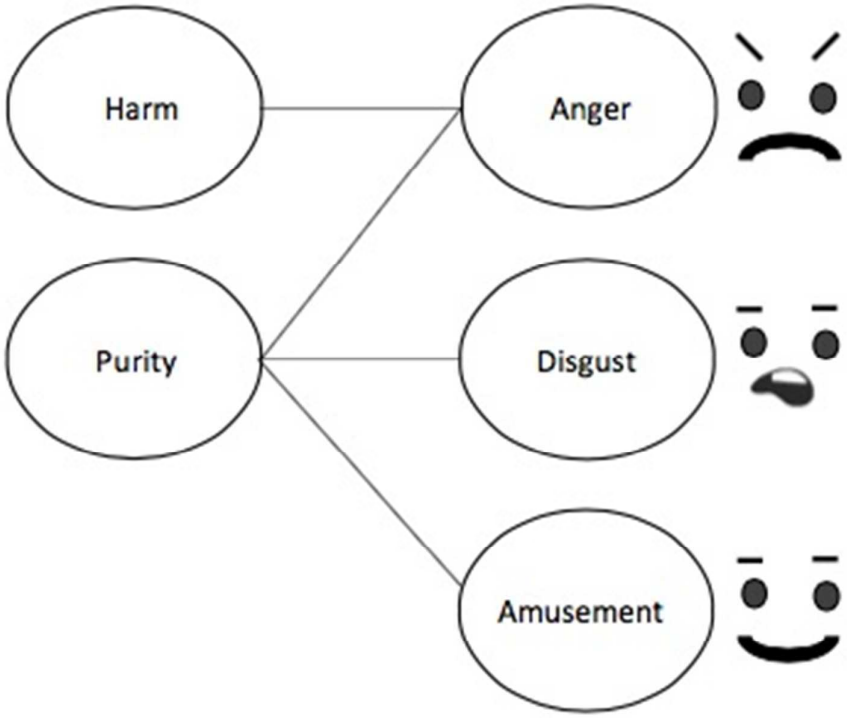


Figure 4. Semi-exclusive morality-emotion correspondences model (an adaptation of the model proposed by Cameron et al. 2015, p.8). The lack of lines between harm and disgust or amusement means non-significant relationships.

review Only

Supplementary Information

Study on self-reported emotions

This study used the same materials as the main study, but here we focused on self-reported emotions rather than on facial expressions. This study aimed to address two issues related to the main study. First, facial expressions are not a perfect proxy for emotions because emotions are not always accompanied by facial expressions. As stated in the paper, this is particularly true for the two emotions of interest: anger and disgust (e.g., Durán et al., 2017). Therefore, we wanted to investigate whether we would obtain similar results with regards to the strong and weak MFT hypotheses with self-reported emotions. We expected that neither of these hypotheses would be fully supported. Second, we wanted to disambiguate the meaning of smiles (enjoyment versus non-enjoyment), which were observed in response to scenarios depicting purity violations. To this end, we included various emotions that could underlie a smile: enjoyment, embarrassment, shame, and grin-and-bear-it.

Participants

Thirty-four native-Italian speakers (22 females, 12 males, M_{age} = 22, age range = 19–28) participated in the study. They were recruited via an announcement through the University of Trento mailing list.

Materials and Procedure

Participants were asked to complete a paper-and-pencil questionnaire. The questionnaire contained the 20 statements which were used in the main study: five involved MFT harm violations, five MFT purity violations, five naturalistic harm violations, and five naturalistic purity violations (for the full list of statements, see Table 1). The statements were presented in a pseudorandomized order. Following each statement, participants were asked to respond to two questions: (1) “While you were reading about the situation, what was your first emotion?” (*enjoyment, grin-and-bear-it, embarrassment, shame, disgust, anger, contempt, surprise*), and

For Peer Review Only

then (2) "Please express a judgment on the intensity of this emotion" (1 = *very weak*, to 9 = *very strong*). The entire experiment lasted about 10 minutes.

Results and Discussion

Figure S1 illustrates the percentage of times participants selected a specific emotion in response to each class of items (harm versus purity) and set of scenarios (MFT versus naturalistic). As it can be seen in Figure S1, for harm items, across both MFT and naturalistic scenarios, anger was the most frequent response, followed by contempt. For purity items, no single emotion prevailed. Specifically, for MFT purity items the most frequent response was disgust, but it was closely followed by surprise, enjoyment, and grin-and-bear-it. For naturalistic purity scenarios, the most frequent response was contempt, which was closely followed by grin-and-bear-it and embarrassment. Turning to emotions that could underlie smiles, MFT purity scenarios elicited an equal amount of positive and negative emotions (enjoyment and grin-and-bear-it), whereas naturalistic purity scenarios predominantly elicited negative emotions (grin-and-bear-it and embarrassment).

Table S1 shows the frequency and intensity of emotional responses to each of the 20 items. With respect to harm items, for both MFT and naturalistic scenarios, the predominant response was anger. There was a single exception, item HM4, for which the prevalent response was grin-and-bear-it. With respect to purity items, disgust was the prevalent response for just two MFT items, PM2 and PM5. Other prevalent responses included enjoyment and surprise (PM1, PM3), and grin-and-bear-it (PM4). For naturalistic items, prevalent responses were grin-and-bear-it (P2, P3), contempt (P1), anger (P4), and embarrassment (P5).

In sum, just like in the main study, the results from the MFT scenarios provide support for the weak MFT hypothesis. Specifically, anger was more prevalent than disgust in response to harm scenarios, whereas disgust was more prevalent than anger in response to purity scenarios. This result also replicates previous findings (Rozin et al., 1999). However, just like in the main study, the results from the naturalistic scenarios conflict with weak MFT

For Peer Review Only

hypothesis. Although anger was more prevalent than disgust in response to harm scenarios, disgust was equally infrequent in response to both harm and purity scenarios. In sum, the results do not offer support for the MFT hypothesis.

However, the results of the two studies also show some differences (compare Figure S1 with Figure 2). For example, the differences in the frequency of anger between harm and purity scenarios is much more pronounced in the follow up than in the main study. More generally, the results of the purity scenarios are more evenly distributed among the eight emotions. This finding is consistent with the hypothesis that purity violations elicit more than one emotion. Specifically, it could be a consequence of people's difficulty to remember which emotion came first or to correctly label that emotion. Indeed, three participants mentioned that they felt a mix of emotions in response to certain scenarios, while two others mentioned that they were uncertain about the difference between shame and contempt. In our opinion, the results of the main study are 'cleaner' because participants were unaware about the purpose of the study, and because the measure used was more direct (self-reports are more subject to meta-judgments).

It is noteworthy to mention that the present analyses lend support to Gray and Keeney's (2015) claim that the MFT purity items are weird (weirder than the MFT harm items, and weirder than the naturalistic harm and purity items). First, surprise was a frequent response to MFT purity items, almost as frequent as disgust. Indeed, for the naturalistic purity items in which surprise responses were rare, disgust responses were also rare, and the results no longer supported weak MFT. Second, MFT purity items elicited more enjoyment than their naturalistic counterparts. Together, the increased surprise and enjoyment associated with MFT purity items are consistent with the idea that these items are weird.

For Peer Review Only

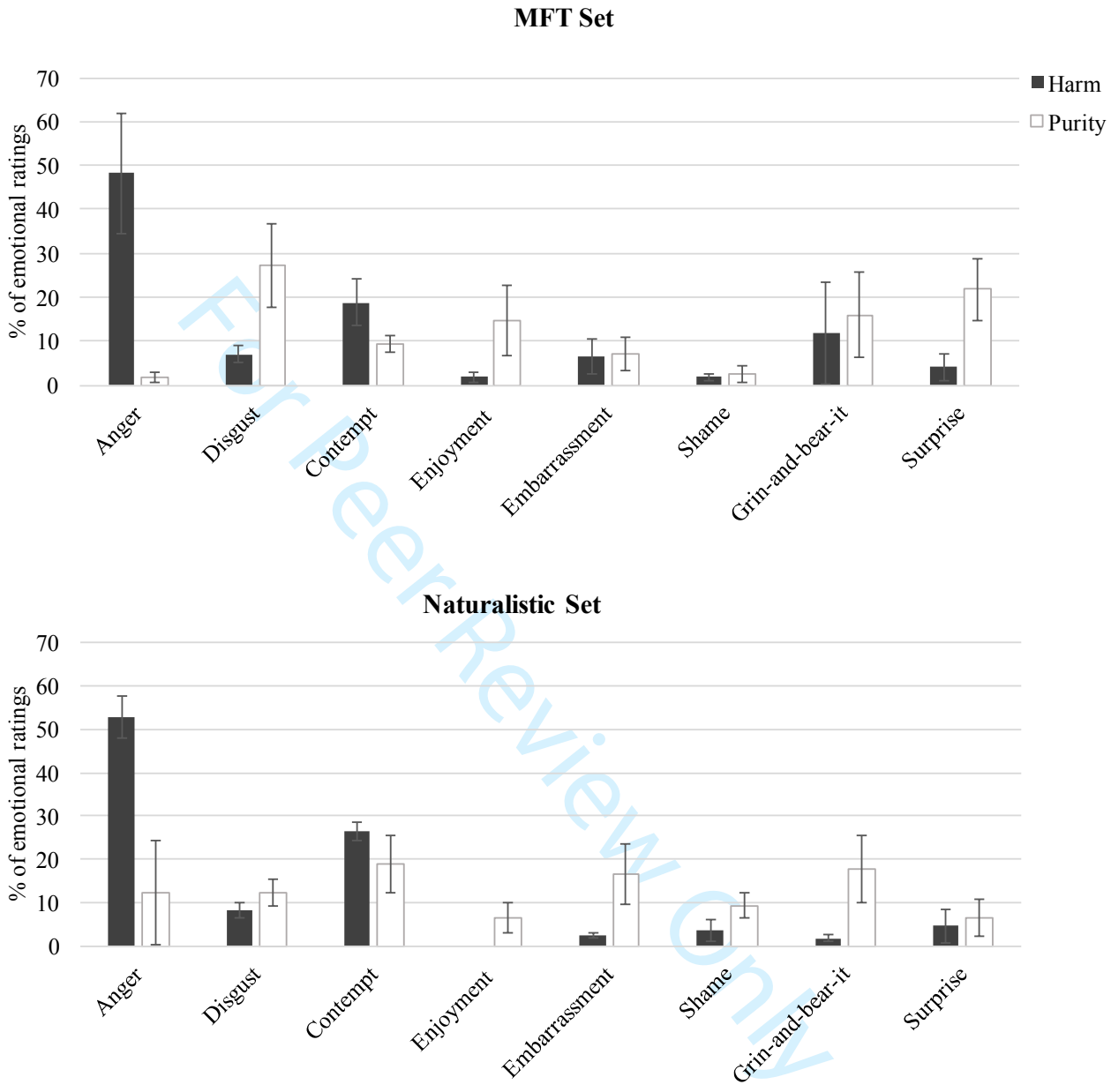


Figure S1. Mean percentage of times an emotion was chosen as a response response to the harm and purity scenarios of the MFT set (a) and the naturalistic set (b). Error bars indicate standard deviation.

For PeerReview Only

Table S1

Frequency of selection of each emotion (in brackets, mean intensity rating) in correspondence with the Moral Foundations Theory (MFT) and the naturalistic scenarios.

		Harm							
	Item	Anger	Disgust	Contempt	Enjoyment	Embarrassment	Shame	Grin-and-bear-it	Surprise
MFT scenarios	HM1	25(7.8)	4(7.5)	4(6.5)	0	0	0	0	1(9)
	HM2	26(7.8)	1(8)	4(6.5)	0	1(4)	1(9)	1(4)	0
	HM3	16(7.4)	1(8)	8(7.5)	0	7(6.7)	1(9)	1(5)	0
	HM4	4(7.5)	3(6.3)	4(5.8)	2(5)	1(5)	1(2)	18(4.8)	1(1)
	HM5	11(7.5)	3(8.3)	12(6.4)	1(5)	2(4)	0	0	5(6.6)
	Mean	16.4(7.6)	2.4(7.6)	6.4(6.5)	0.6(5)	2.2(4.9)	0.6(6.7)	4(4.6)	1.4(5.5)

For Peer Review Only

	SD	9.3(0.2)	1.3(0.8)	3.6(0.6)	0.9 (0)	2.8(1.3)	0.5(4)	7.8(0.5)	2.1(4.1)
Naturalistic scenarios	H1	17(7.1)	2(6.5)	8(7.1)	0	1(5)	0	0	6(6.3)
	H2	15(6.9)	3(7.7)	10(6.6)	0	1(9)	4(6.5)	1(3)	0
	H3	21(7.3)	4(8.3)	8(7.1)	0	1(6)	0	0	0
	H4	15(6.6)	4(6.8)	11(6.4)	0	1(5)	2(5.5)	1(2)	0
	H5	22(7.8)	1(9)	8(8.1)	0	0	0	1(2)	2(7)
	Mean	18(7)	2.8(7.6)	9(7.1)	0(0)	0.8(6.3)	1.2(6)	0.6(2.3)	1.6(6.7)
	SD	3.3(0.5)	1.3(1)	1.4(0.7)	0(0)	0.4(1.9)	1.8(0.7)	0.5(0.6)	2.6(0.5)

For Peer Review Only

Purity

	Item	Anger	Disgust	Contempt	Enjoyment	Embarrassment	Shame	Grin-and-bear-it	Surprise
MFT scenarios	PM1	0	4(5.7)	2(6.5)	12(5.9)	2(5.5)	0	1(5)	13(6.2)
	PM2	1(9)	20(7.6)	4(6.3)	0	1(8)	1(9)	3(6.3)	4(7.8)
	PM3	0	5(4.8)	2(7.5)	9(6.6)	2(5)	0	4(5.3)	12(6.1)
	PM4	2(6.5)	7(6.6)	3(6.3)	0	0	0	17(5.5)	5(6.6)
	PM5	0	10(7.6)	5(6.2)	4(7)	7(7.4)	3(8)	2(5.5)	3(7)
	Mean	0.6(7.8)	9.2(6.5)	3.2(6.3)	5(6.5)	2.4(6.5)	0.8(8.5)	5.4(5.5)	7.4(6.7)
	SD	0.9(1.8)	6.5(1.2)	1.3(0.1)	5.4(0.5)	2.7(1.5)	1.3(0.7)	6.6(0.5)	4.7(0.7)
Naturalistic scenarios	P1	0	4(7.2)	13(7.7)	3(5)	5(5.2)	3(7.7)	4(3.5)	2(4.5)
	P2	0	6(6.7)	8(6.4)	1(7)	4(5)	2(8.5)	13(4)	0
	P3	1(8)	4(7)	4(5.8)	1(4)	6(5.5)	6(6)	10(5.9)	2(4)
	P4	20(8.5)	6(7.8)	6(8.3)	0	0	1(8)	1(6)	0

For Peer Review Only

P5	0	1(3)	1(5)	6(6.3)	13(7.3)	4(7)	2(6.5)	7(6.7)
Mean	4.2(8.2)	4.2(6.4)	6.4(6.6)	2.2(5.6)	5.6(5.8)	3.2(7.4)	6(5.2)	2.2(5.1)
SD	8.8(0.3)	2(1.9)	4.5(1.4)	2.4(1.3)	4.7(1.1)	1.9(1)	5.2(1.3)	2.8(1.4)

For Peer Review Only