**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 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 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.





*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.

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 SD** | 16.4(7.6) 9.3(0.2) | 2.4(7.6) 1.3(0.8) | 6.4(6.5) 3.6(0.6) | 0.6(5) 0.9 (0) | 2.2(4.9) 2.8(1.3) | 0.6(6.7) 0.5(4) | 4(4.6) 7.8(0.5) | 1.4(5.5) 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 SD** |  18(7) 3.3(0.5) | 2.8(7.6) 1.3(1) | 9(7.1) 1.4(0.7) | 0(0) 0(0) | 0.8(6.3) 0.4(1.9) | 1.2(6) 1.8(0.7) | 0.6(2.3) 0.5(0.6) | 1.6(6.7) 2.6(0.5) |  |

|  |  |
| --- | --- |
|  | **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 SD** | 0.6(7.8) 0.9(1.8) | 9.2(6.5) 6.5(1.2) | 3.2(6.3) 1.3(0.1) | 5(6.5) 5.4(0.5) | 2.4(6.5) 2.7(1.5) | 0.8(8.5) 1.3(0.7) | 5.4(5.5) 6.6(0.5) | 7.4(6.7) 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 |  |
| P5 | 0 | 1(3) | 1(5) | 6(6.3) | 13(7.3) | 4(7) | 2(6.5) | 7(6.7) |  |
| **Mean SD** | 4.2(8.2) 8.8(0.3) | 4.2(6.4) 2(1.9) | 6.4(6.6) 4.5(1.4) | 2.2(5.6) 2.4(1.3) | 5.6(5.8) 4.7(1.1) | 3.2(7.4) 1.9(1) | 6(5.2) 5.2(1.3) | 2.2(5.1)2.8(1.4) |