

EMPIRICAL ARTICLE

# The Dutch moral foundations stimulus database: An adaptation and validation of moral vignettes and sociomoral images in a Dutch sample

Frederic R. Hopp<sup>1</sup>, Benjamin Jargow<sup>2</sup>, Esmee Kouwen<sup>3</sup>, and Bert N. Bakker<sup>1</sup>

<sup>1</sup>Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands; <sup>2</sup>Department of Psychology, University of Amsterdam, Amsterdam, Netherlands and <sup>3</sup>Institute for Interdisciplinary Studies, University of Amsterdam, Amsterdam, Netherlands

**Corresponding author:** Frederic R. Hopp; Email: [fhopp@uva.nl](mailto:fhopp@uva.nl)

**Received:** 18 September 2023; **Revised:** 2 February 2024; **Accepted:** 5 February 2024

**Keywords:** moral judgment; vignettes; images; cross-cultural; moral foundations theory

## Abstract

Moral judgments are shaped by socialization and cultural heritage. Understanding how moral considerations vary across the globe requires the systematic development of moral stimuli for use in different cultures and languages. Focusing on Dutch populations, we adapted and validated two recent instruments for examining moral judgments: (1) the Moral Foundations Vignettes (MFVs) and (2) the Socio-Moral Image Database (SMID). We translated all 120 MFVs from English into Dutch and selected 120 images from SMID that primarily display moral, immoral, or neutral content. A total of 586 crowd-workers from the Netherlands provided over 38,460 individual judgments for both stimuli sets on moral and affective dimensions. For both instruments, we find that moral judgments and relationships between the moral foundations and political orientation are similar to those reported in the US, Australia, and Brazil. We provide the validated MFV and SMID images, along with associated rating data, to enable a broader study of morality.

Moral intuitions—instant feelings of approval or disapproval that come with witnessing moral actions (Haidt, 2001)—vary within and between cultures (Graham et al., 2011, 2016; Haidt and Joseph, 2004). To investigate morality across the globe, we need valid and reliable instruments that adopt the language and cultural context of specific regions (Atari et al., 2023). With mounting studies tailoring their moral judgment tasks to cultural idiosyncrasies (e.g., Bobbio et al., 2011; Kim et al., 2012; Marques et al., 2020; van Leeuwen and Park, 2009), we can expand our understanding of how individuals' moral compass is guided by regional and sociopolitical pressures (Malik et al., 2021).

We contribute to this line of work by adapting and validating two existing, popular moral stimulus sets for studying moral judgment among Dutch populations: the Moral Foundations Vignettes (MFV; Clifford et al., 2015) and the Socio-Moral Image Database (SMID; Crone et al., 2018). Specifically, we adhere to the original validation procedures of the MFV and SMID as closely as possible by utilizing a crowd-sourced procedure based on a Dutch sample. Our focus on the Netherlands and these stimulus sets is motivated by three reasons. First, the Netherlands is a multiparty system that has recently witnessed an increase in affective polarization (Harteveld, 2021), and understanding how moral intuitions diverge across partisan lines can foster mutual understanding (Puryear et al., 2022).

Second, the text-based MFV have already successfully been adopted to the Portuguese language with a Brazilian sample (Marques et al., 2020), yet how well the MFV transfer to European populations is largely unknown (but see a pilot study by Wagemans et al., 2018, who used a small selection of 8–10 vignettes in Dutch samples, while we adopt and validate 120 vignettes). Third, although images can have diverse moral interpretations based on cultural differences, they do not require a literal translation and hence offer a more direct approach for probing cross-cultural differences in moral intuitions; however, the SMID's applicability to European contexts remains unclear.

## 1. Moral foundations theory

Moral Foundations Theory (MFT; Haidt and Joseph, 2004) provides a taxonomy of moral intuitions by postulating that a set of separate but interrelated moral foundations has developed over the course of cultural evolution. In its original conceptualization, MFT spanned a set of five moral foundations: care-harm, fairness-cheating, authority-subversion, loyalty-betrayal, and sanctity-degradation (Haidt and Joseph, 2004). These moral intuitions were further organized into two broader moral categories: care-harm and fairness-cheating as 'individualizing' foundations that primarily serve to protect the rights and freedoms of individuals; and loyalty-betrayal, authority-subversion and sanctity-degradation into 'binding' foundations that primarily operate at the group level (Haidt, 2008). Since its inception, additional candidate moral foundations were discussed, including liberty (Iyer et al., 2012), honor (Atari et al., 2020), and ownership (Atari and Haidt, 2023). Analogously, in a recent update of MFT, Atari et al. (2023) also split the fairness-cheating foundation into distinct and new foundations of equality and proportionality. This split aimed to capture the distinct moral concerns of fairness in procedure (proportionality) and equality of outcome (equality).

Extant studies show robust support for cultural and ideological differences in the endorsement of moral foundations (Graham et al., 2011; Kivikangas et al., 2021). This line of research has primarily relied on the Moral Foundations Questionnaire (MFQ; Graham et al., 2011). Briefly, the MFQ includes two sets of questions that either probe (a) the *relevance* of moral foundations for deciding whether something is morally right or wrong (e.g., 'Whether someone suffered emotionally') or (b) the *(dis)agreement* with statements concerning the upholding of moral foundations in society (e.g., 'It can never be right to kill a human being'). The MFQ has proven very useful for the evaluation of variations in the endorsement of moral values, particularly as they pertain to differences in political orientation (Graham et al., 2009, 2012) and sex (Graham et al., 2011). Moreover, a plethora of studies tapping into cross-cultural differences in moral concerns have relied on the MFQ (Graham et al., 2011) and its recent successor MFQ-2 (Atari et al., 2023), demonstrating that world region is a significant and reliable predictor for describing cultural variation in moral principles. Although the factor structure of the MFQ remains a topic of ongoing debate (Curry et al., 2019; De Buck and Pauwels, 2023; Harper and Rhodes, 2021; Zakhari and Bates, 2021), a more practical limitation of the MFQ is concerned with the types of questions that its design permits to address (Clifford et al., 2015).

In summary, the MFQ largely captures respondents' rating of *abstract* principles, rather than moral judgments of *concrete* scenarios. As Graham et al. (2009, p. 1031) have stated, moral relevance 'does not necessarily measure how people actually make moral judgments', but these ratings are 'best understood as self-theories about moral judgment'. Yet, individuals' theories of morality (i.e., endorsement of moral principles) might diverge from their specific moral judgments (Haidt, 2001). As Clifford et al. (2015, p. 1179) have argued, 'one might view harm or loyalty as highly relevant to morality, yet refrain from making harsh judgments about others' harmful or disloyal behavior'. Relatedly, Graham et al. (2013) argue that the existence of a moral foundation can be doubted if there is a lack of response to *third-party transgressions* of that foundation. Because the MFQ does not probe respondents' moral judgment of third-party transgressions, it may not be an ideal instrument for testing the existence of moral foundations.

Analogously, the Social Intuitionism Model (Haidt, 2001) postulates that people may not possess reliable introspective access to the causes of their moral judgments, and thus their reports about abstract

moral concerns may not perfectly reflect how they actually form moral judgments. Moreover, many of these items include an ‘unstated, and ambiguous referent’, such as an authority figure, yet people may ‘judge MFT issues differently depending on the referents’ (Frimer et al., 2013, p. 1053; see also Eriksson et al., 2019). Finally, the brevity of the MFQ—a total of 30 items—makes it unsuitable for use in neuroimaging studies of moral judgment, partially due to insufficient statistical power as well as the lack of control for question length and complexity that introduce neurological confounds (e.g., Baciú et al., 2002; Church et al., 2011).

As an alternative to probing individuals’ endorsement of moral foundations, some studies have used the Moral Foundations Sacredness Scale (MFSS), which was designed to examine respondents’ willingness to engage in taboo trade-offs (Tetlock et al., 2000), such as kicking a dog in the head (Care) or renouncing one’s citizenship (Loyalty) for money (Graham et al., 2009). However, Clifford et al. (2015, p. 1180) point out that ‘the MFSS is designed to measure an individual’s willingness to violate moral norms in exchange for money, as opposed to judgments of others’ behaviors’.

To better probe individuals’ moral judgment of concrete situations and behaviors, various studies have developed moral vignettes on an ad hoc basis, with some devising scenarios corresponding to the harm and purity moral foundations (Heekeren et al., 2005; Parkinson et al., 2011; Schaich Borg et al., 2008, 2011). Others used photographic images to depict certain moral violations but collapsed across all forms of violations in their analyses, making it difficult to examine potential differences between specific moral foundations (Harenski et al., 2008; Harenski and Hamann, 2006; Moll et al., 2002).

In view of these limitations, researchers have started to develop standardized and normed moral foundation vignettes and image databases for studying moral judgment. Popular databases for morally relevant scenarios include the Moral Foundations Vignettes (MFV; Clifford et al., 2015; cited 363 times to date on Google Scholar). The MFV span 120, one-sentence descriptions detailing the violation of one (and only one) of seven moral foundations: physical care, emotional care, fairness, liberty, loyalty, authority, and sanctity. The vignettes also contain nonmoral, social norm transgressions. Notably, the creators of the MFV split the care-harm foundation into three subcomponents to reflect the diversity of the original conception of care-harm: emotional harm to a human, physical harm to a human, and physical harm to a nonhuman animal. This division is also aligned with neural evidence showing that the introduction of bodily harm into either a moral or nonmoral scenario can influence the levels of observed neural activity in certain brain regions (Heekeren et al., 2005).

The MFV have been employed in both behavioral (Clifford, 2017; Dehghani et al., 2016; Wagemans et al., 2018) and functional magnetic resonance imaging (fMRI) studies (Hopp et al., 2023; Khoudary et al., 2022). While these studies have administered the MFV solely in US samples, recent work by Marques et al. (2020) introduced a Portuguese adaptation of the MFV. Using a Brazilian sample ( $N = 494$ ), they demonstrated that the Portuguese version of the MFV performed similarly to the original English version in terms of its factor structure. Aside from this Portuguese case study, there have been no attempts to adapt and validate the MFV to other contexts, although MFT’s theoretical postulations demand cross-cultural research.

Adapting the MFV for non-English countries necessitates translating and adjusting specific vignettes for cultural comprehension (Marques et al., 2020). A remedy for this issue may be offered by recent studies that have developed photographic and even audiovisual moral stimulus databases (Crone et al., 2018; McCurrie et al., 2018). The SMID (Crone et al., 2018; cited 49 times to date on Google Scholar) offers a large resource for examining differences in moral judgment, both within and across cultures. The SMID contains 2,941 images, each annotated for moral and affective qualities using crowd-sourced samples from the United States and Australia. Each image was rated on how much it depicts each moral foundation as well as for general valence, arousal, and (im)morality. Notably, images in the SMID also display morally good actions, extending previous stimulus sets which solely contain moral transgressions. Moreover, images may offer increased ecological validity over text-based vignettes, which have been criticized for creating an artificial moral psychology of ‘raceless, genderless strangers’ (Hester and Gray, 2020). Subsets of SMID images were already used in previous studies in Japan (Chunyu et al., 2021, 160 images; Sudo et al., 2021; 60 images) and China (Tao et al., 2022a, 66 images;

2022b, 192 images), but validations of SMID images in a European context are absent. Furthermore, prior research has predominantly utilized the SMID to gather general moral and immoral images, often relying on student samples for image evaluations. To advance MFT, it is essential to procure images that consistently elicit perceptions of distinct moral foundations in more diverse cultural populations. Given that SMID's moral foundation ratings originate from crowd-workers in the United States and Australia, validation is required to examine the applicability of visual representations of moral foundations in other cultures.

### **1.1. Current work**

In view of demands for culturally tailored moral stimulus sets, we adapt and validate the MFV and SMID for studying moral judgment among Dutch populations. We first translated and adapted the MFV into Dutch. Second, we selected images from the SMID that primarily display moral and immoral exemplars of each moral foundation as well as neutral images that do not display moral information. In turn, we validated these stimuli sets using a large crowd-sourced sample from the Netherlands. Crowd-sourced validations of moral stimuli are increasingly becoming the gold standard in moral psychology, particularly because they capture a more diverse moral signal and are less prone to introducing annotator biases (Crone et al., 2018; Hopp et al., 2021; Hopp and Weber, 2021; McCurrie et al., 2018).

## **2. Method**

We report how we determined our sample size and all data exclusions in the study. All materials, data, analysis code, and supplementary information (SI) are accessible at <https://osf.io/9gnza>. This study's design and its analysis were not preregistered. All procedures were approved by the ethics board of the host institution.

### **2.1. Moral foundations vignettes**

The full MFV database by Clifford et al. (2015) contains 132 moral transgressions. Previous work had adapted selections of 8–10 (Wagemans et al., 2018) and 90 (Marques et al., 2020) vignettes, but we aimed to adapt a larger selection of 120 vignettes that have been employed in past experimental research (Hopp et al., 2023; Khoudary et al., 2022). Each vignette consists of a one-sentence description (14–17 words) detailing the violation of one (and only one) of seven moral foundations (see Table 1 for examples): physical care, emotional care, fairness, liberty, loyalty, authority, and sanctity. The vignettes also contain a nonmoral, social norm transgression category. Each of the eight conditions featured 15 vignettes. One of the authors—a Dutch native—translated each vignette from English into Dutch. After translating all vignettes, the translator met with the remaining authors of the paper to ensure that minor adjustments of the vignettes fit the context of the Netherlands (all edits are reported in SI Table 1).

### **2.2. Socio-Moral Image Database**

The SMID (Crone et al., 2018) contains 2,941 images, all annotated for moral and affective qualities using crowd-sourced samples located in the United States and Australia. Each image was rated on how much it depicts Care, Fairness, Loyalty, Authority, and Sanctity, using a five-point Likert-type scale from 1 (not at all) to 5 (very much). Similarly, each image was also rated using five-point Likert-type scales for valence (1 = unpleasant or negative; 5 = pleasant or positive), arousal (1 = calming; 5 = exciting), and morality (1 = immoral/blameworthy; 5 = moral/praiseworthy). Because valence and morality ratings correlated at  $r = .87$  (Crone et al., 2018), we only retrieved the morality ratings. Based on these ratings, we organized all images into a circumplex model typically used for stimulus sampling in emotion research (Russell, 1980), with one axis describing morality and the other axis capturing

**Table 1.** Example moral foundations vignettes and Dutch translations.

MFV category	Language	Example vignettes
Physical care	ENG	You see a teacher hitting a student's hand with a ruler for falling asleep in class.
	NL	U ziet een leraar die een leerling op de hand slaat met een liniaal voor het in slaap vallen in de klas.
Emotional care	ENG	You see a boy making fun of his brother for getting dumped by his girlfriend.
	NL	U ziet een jongen die zijn broer uitlacht omdat hij gedumpt is door zijn vriendin.
Fairness	ENG	You see a referee intentionally making bad calls that help his favored team win.
	NL	U ziet een scheidsrechter opzettelijk slechte beslissingen nemen die zijn favoriete team helpen winnen.
Liberty	ENG	You see a teenager at a cafeteria forcing a younger student to pay for her lunch.
	NL	U ziet een tiener in de kantine een jongere student dwingen haar lunch te betalen.
Loyalty	ENG	You see the coach's wife sponsoring a bake sale for her husband's rival team.
	NL	U ziet de vrouw van de coach een koekjes verkoop sponsoren voor het rivaliserende team van haar man.
Authority	ENG	You see a boy spray-painting anarchy symbols on the side of the police station.
	NL	U ziet een jongen anarchie symbolen verfspuiten aan de zijkant van het politiebureau.
Sanctity	ENG	You see a man having sex with a frozen chicken before cooking it for dinner.
	NL	U ziet een man seks hebben met een bevroren kip voordat hij het voor het avondeten kookt.
Social norms	ENG	You see a woman answering a phone call with the word goodbye instead of hello.
	NL	U ziet een vrouw die een telefoontje opneemt met het woord 'tot ziens' inplaats van 'hallo'.

Note: All translated vignettes are available in SI Table 1.

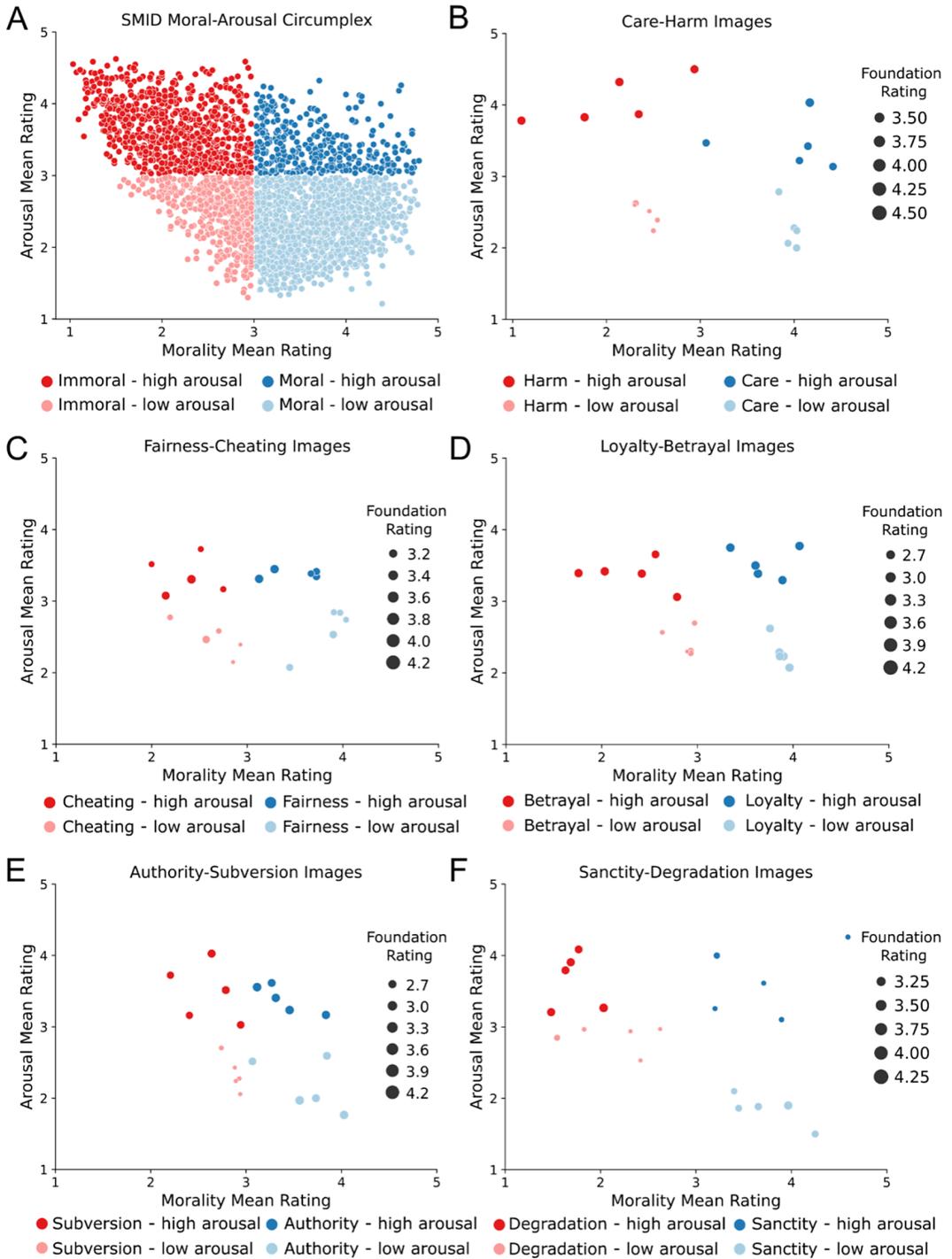
arousal, thereby creating four image quadrants (Figure 1A): moral-high arousal ( $N = 340$ ); moral-low arousal ( $N = 1,247$ ); immoral-high arousal ( $N = 767$ ); and immoral-low arousal ( $N = 500$ ). Next, within each quadrant, we selected 20 images rated highest on a single foundation and lowest on all other foundations (Figure 1B–F). In a similar fashion, we also sampled five ‘neutral’ images in each quadrant that received high and low arousal ratings, but clustered close to a morally neutral rating of ‘3’ (i.e.,  $\geq 2.9$ ;  $\leq 3.1$ ) and were rated lowest across all moral foundations. This resulted in a final sample of 120 images, with 10 moral and 10 immoral images per moral foundation category as well as 10 high arousal, morally neutral and 10 low-arousal, morally neutral images (for example images, please see Figure 2).

### 2.3. Participants

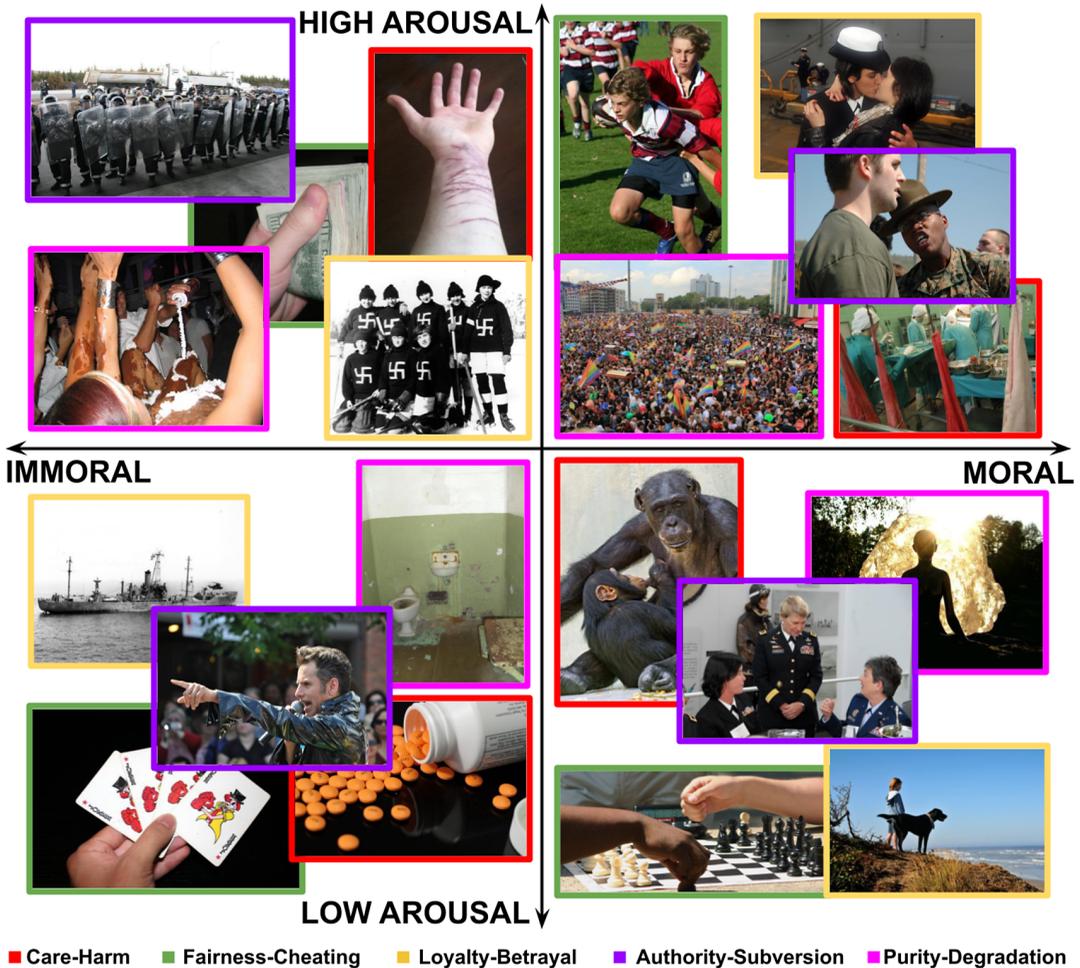
We used the Prolific academic (PA) platform (<https://www.prolific.co/>) for recruiting participants. Eligibility criteria included speaking Dutch as a first language, holding Dutch nationality, and being located in the Netherlands. Our sample size was determined by following previous moral stimulus validation studies (Clifford et al., 2015; Crone et al., 2018; McCurrie et al., 2018) and thus aimed to obtain at least 20 ratings for each stimulus on each dimension. In total, 648 survey responses were collected, of which 62 were excluded who provided incomplete responses or finished the survey in under 6 minutes (<5% quantile), leaving us with a total sample size of 586. Complete demographic information for 572 participants could be retrieved and indicated that we had a diverse sample of the Dutch population: participants had a mean age of 28.39 years ( $SD = 8.89$ ) of which 326 (57%) identified as male (244 female; 1 nondisclosed). We assessed political orientation using a slider ranging from ‘very left’ (0) to ‘very right’ (100) (Dodd et al., 2012). Our sample was politically diverse, with a slight skew toward the political left ( $M = 38.95$ ,  $SD = 22.12$ ). The majority reported a White ethnicity (497; 87%), followed by mixed (42; 8%), Asian (14; 3%), Black (8, 1%), and ‘other’ (8; 1%). 253 (44%) participants indicated to not hold a student status, 239 (41%) held a student status, and student status data for 80 participants had expired.

### 2.4. Procedure

Data were collected through an online survey using Qualtrics. After signing the informed consent, the survey started with a brief overview of MFT—adapted from Crone et al. (2018) and translated by us into Dutch—to familiarize participants with the basic contents of moral foundations. Next, participants provided ratings of vignettes, images, or news clips (not reported here), in which the order of stimuli blocks varied randomly across participants. Each participant was assigned to a random selection of five vignettes and five images, respectively. For each vignette, participants used a five-point Likert scale to rate the vignette’s *moral wrongness* (‘How morally wrong is the displayed behavior?’; 1: not at all wrong – 5: extremely wrong), *comprehensibility* (‘How easy is it for you to understand what is described in the scenario?’; 1: not at all easy to understand – 5: extremely easy to understand), *imaginability* (‘How easy is it for you to clearly imagine what is happening in the scenario?’; 1: not at all easy to imagine – 5: extremely easy to imagine), *frequency* (‘How often do you see or hear about actions like the one described in this scenario in the media or your daily life?’; 1: not at all often – 5: extremely often), and *emotional response* (‘How strong was your emotional response to the behavior depicted in this scenario?’; 1: not at all strong – 5: extremely strong). Participants were also asked why the action is morally wrong and could choose one out of seven response options reflecting each vignette category (all vignette-related item prompts and response options are provided in their original English and translated Dutch version in the SI). Every participant was prompted to indicate why a vignette was morally wrong, independent of their answer to the moral wrongness rating prompt. Yet, participants did have the option to indicate that ‘It is not morally wrong and does not apply to any of the provided choices’. Hence, if a participant judged a vignette ‘not at all wrong’ they could still indicate that this vignette did not violate any moral norms.



**Figure 1.** SMID image sampling procedure. (A) The 2,941 images were first organized into a circumplex model according to the midpoint (3) of the Arousal and Morality rating axes. (B–F) The selection of foundation-specific images proceeded as follows: From each quadrant of the original circumplex model, five images were selected that received the highest rating for a given foundation and the lowest ratings for all other foundations. Dot sizes in B–F reflect the average degree to which images in each category were perceived to display that moral foundation, with greater sizes indicating a higher average foundation-specific rating.



**Figure 2.** Examples of selected SMID images for each moral-arousal quadrant. Image border color denotes the moral foundation that received the highest rating in the original study (Crone et al., 2018). Neutral images rated low on all moral foundations are not shown.

Similarly, for each image, participants used a five-point Likert scale to rate the image’s general valence, arousal, and morality as well as the degree (henceforth: moral foundation relevance) to which the image makes them think about each moral foundation (all image-related item prompts and response options are provided in their original English and translated Dutch version in the SI).

### 3. Results

#### 3.1. Moral foundations vignettes

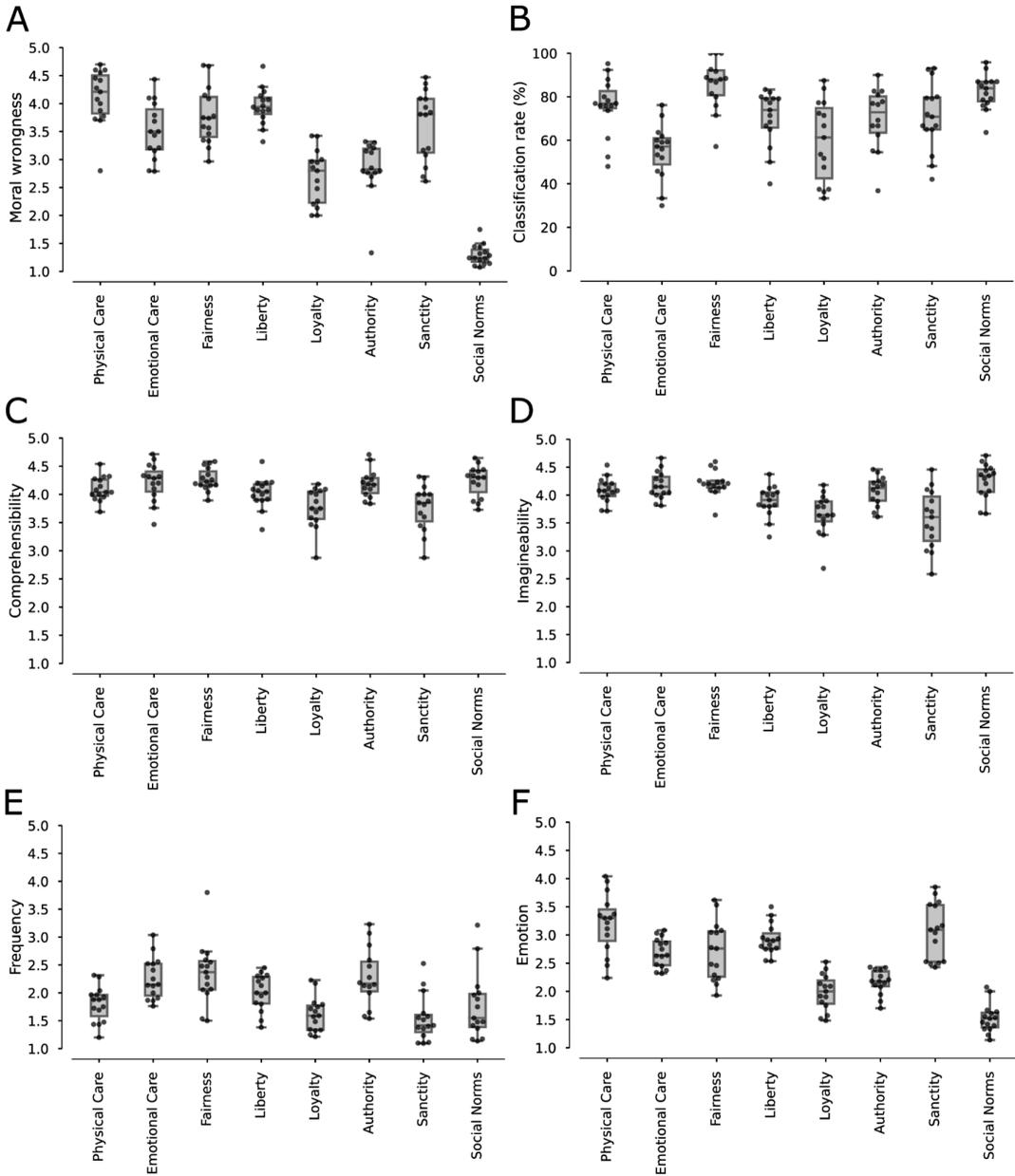
All vignettes were rated an average of 21.97 times (min: 14; max 30).<sup>1</sup> We first tested whether vignettes displaying a moral violation were rated as more morally wrong than vignettes describing a social norm transgression (Table 2). Indeed, every moral vignette item was rated as more morally wrong than every social norm vignette item (Figure 3A), except for one Authority item (MFV 61 ‘You see a teaching

<sup>1</sup>Due to a technical error, ratings for one authority vignette (MFV 80, ‘You see a boy turning up the TV as his father talks about his military service’) could not be retrieved and thus are not reported.

**Table 2.** Ratings across MFV categories.

Category	Moral wrongness		Classification rate (%)		Comprehensibility		Imaginability		Frequency		Emotion	
	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI
Physical care	4.09	0.95 [3.99, 4.19]	75.89	4.28 [71.48, 80.30]	4.09	1.10 [3.98, 4.20]	4.08	0.96 [3.99, 4.18]	1.79	0.84 [1.71, 1.88]	3.19	1.10 [3.08, 3.31]
Emotional care	3.53	0.94 [3.42, 3.63]	54.57	4.99 [49.33, 59.81]	4.19	0.88 [4.10, 4.29]	4.17	0.82 [4.08, 4.25]	2.27	0.99 [2.17, 2.38]	2.70	1.02 [2.59, 2.80]
Fairness	3.72	0.91 [3.62, 3.82]	85.63	3.51 [81.89, 89.37]	4.25	0.89 [4.16, 4.35]	4.19	0.83 [4.10, 4.28]	2.32	1.01 [2.21, 2.42]	2.68	1.02 [2.57, 2.79]
Liberty	3.94	0.89 [3.84, 4.04]	69.59	4.61 [64.69, 74.49]	4.02	1.01 [3.91, 4.13]	3.88	0.97 [3.78, 3.98]	2.01	0.87 [1.92, 2.10]	2.92	1.01 [2.81, 3.02]
Loyalty	2.68	1.09 [2.57, 2.80]	60.20	4.90 [55.34, 65.07]	3.79	1.11 [3.68, 3.90]	3.68	1.12 [3.57, 3.80]	1.62	0.80 [1.54, 1.70]	1.97	0.89 [1.88, 2.06]
Authority	2.89	0.92 [2.78, 3.00]	69.97	4.59 [64.86, 75.07]	4.19	0.79 [4.10, 4.28]	4.08	0.83 [3.99, 4.18]	2.30	1.05 [2.18, 2.41]	2.18	0.84 [2.08, 2.27]
Sanctity	3.64	1.15 [3.52, 3.76]	72.82	4.45 [68.32, 77.32]	3.74	1.24 [3.61, 3.86]	3.54	1.22 [3.42, 3.67]	1.52	0.82 [1.43, 1.60]	3.11	1.08 [3.00, 3.22]
Social Norms	1.29	0.67 [1.22, 1.36]	82.60	3.80 [78.79, 86.40]	4.25	0.98 [4.15, 4.35]	4.27	0.95 [4.18, 4.37]	1.78	1.01 [1.68, 1.89]	1.52	0.79 [1.44, 1.60]

Note: Classification rate reflects the percentage of categorization into the intended foundation.



**Figure 3.** Moral foundations vignettes ratings. (A) Moral wrongness. (B) Classification rate in percent. (C) Comprehensibility. (D) Imaginability. (E) Frequency. (F) Emotional response. Each dot reflects the mean response of all participants to a single vignette item. Box plots for each condition display median (center line), upper and lower quartiles (box limits), whiskers connote 1.5 × interquartile range (IQR) and points that fall outside the whiskers are outliers.

assistant talking back to the teacher in front of the classroom’; moral wrongness ratings for each vignette item are summarized in SI Table 1). Replicating previous work (Clifford et al., 2015; Hopp et al., 2023; Khoudary et al., 2022; Marques et al., 2020), moral vignettes violating physical care received the highest moral wrongness rating, whereas loyalty violations received the lowest moral wrongness ratings among moral vignettes (Table 2). We also tested whether each moral foundation category was rated more morally wrong than social norms. Using the Tukey–Kramer Method for multiple comparisons of

**Table 3.** Difference of moral wrongness ratings between each moral foundation and social norms.

Foundation	Difference	95% CI	<i>d</i>	<i>q</i>	<i>p</i>
Physical care	2.80	[2.58, 3.02]	3.43	54.62	<.001
Emotional care	2.24	[2.01, 2.46]	2.77	43.02	<.001
Fairness	2.43	[2.21, 2.66]	3.07	46.77	<.001
Liberty	2.65	[2.43, 2.88]	3.41	50.55	<.001
Loyalty	1.39	[1.18, 1.61]	1.53	27.60	<.001
Authority	1.60	[1.37, 1.83]	2.03	29.81	<.001
Sanctity	2.35	[2.13, 2.57]	2.52	46.04	<.001

Note: Results of Tukey's honest significance test on the difference between moral wrongness ratings of each moral foundation and social norms.

groups with unequal sample sizes (Kramer, 1956), we found that violations of each moral foundation were rated as significantly more morally wrong than social norm transgressions (Table 3).

Next, we tested whether each vignette was classified into its originally intended category (Table 2 and Figure 3B). To this end, we calculated the classification rate (%)—the percentage of times a vignette was classified into their intended category. We observed that the majority of vignettes (97%) were classified into their intended category, with average classification rates ranging from 85.63% for Fairness vignettes to 60.2% for Loyalty vignettes (classification rates for each vignette item are reported in SI Table 1). Only four vignettes were mostly classified into a nonintended category: 1) The above-mentioned Authority item MFV 61 (73.33% 'Not Wrong'); 2) Loyalty item MFV 1: 'You see a former Secretary of State publicly giving up his citizenship to the Netherlands' (53.33% 'Not Wrong'); 3) Loyalty item MFV 72: 'You see a Dutch swimmer cheering as a Chinese foe beats his teammate to win the gold' (56.52% 'Not Wrong'); and 4) Emotional Care item MFV 35: 'You see a man laughing at a disabled co-worker while at an office softball game' (45.00% 'Liberty'). Curiously, both Loyalty items received higher average moral wrongness than any social norm vignette, suggesting that participants may indeed have intuitively perceived them as moral violations. In addition, all vignettes were rated as highly comprehensible and imaginable, and ratings of frequency, as well as emotional response, were comparable to those reported in the original MFV (Clifford et al., 2015) study (Table 2 and Figure 3C–F).

Furthermore, we explored the correlation between moral wrongness ratings across vignette categories and participants' political orientation (Table 4). Consistent with MFT, authority and loyalty, which both belong to MFT's *binding* moral foundations, were significantly positively correlated ( $r = .25, p = .008$ ). Analogously, we found that a more right-leaning political attitude correlated significantly and positively with wrongness ratings of the binding moral foundations loyalty ( $r = .16, p = .011$ ), authority ( $r = .21, p = .001$ ), and sanctity ( $r = .14, p = .022$ ). Surprisingly, Fairness, which belongs to MFT's *individualizing* foundations, significantly positively correlated with Loyalty ( $r = .24, p = .009$ ) and Sanctity ( $r = .27, p = .004$ ); both belonging to the binding moral foundations. As demonstrated by Hopp et al. (2023), more right-leaning individuals also rated Fairness ( $r = .18, p = .004$ ) and Social Norms ( $r = .21, p < .001$ ) as more morally wrong. In view of these empirical results and how they compare to previous studies, we consider our translation and adaptation of the MFV to the Dutch context successful.

### 3.2. Socio-Moral Image Database

All images were rated an average of 21.88 times (min: 10; max 36). First, we examined whether images originally rated as moral (immoral) were also judged as moral (immoral) by our Dutch sample (Table 5). Collapsing all images across their foundation-specific categories, moral images were rated as more moral ( $M = 3.66, SD = 0.49$ ) and immoral images were judged to be more immoral ( $M = 2.64, SD = 0.56$ ). This difference was large in terms of effect size and statistically significant  $t(118) = 10.63$ ,

**Table 4.** Correlations of moral wrongness ratings between MFV categories and political orientation.

	Physical care	Emotional care	Fairness	Liberty	Loyalty	Authority	Sanctity	Social norms
Emotional care	.03							
Fairness	.12	.08						
Liberty	.07	-.03	.15					
Loyalty	.11	.09	<b>.24**</b>	.13				
Authority	-.08	.08	.15	<b>.24*</b>	<b>.25**</b>			
Sanctity	.10	.05	<b>.27**</b>	-.03	.11	.03		
Social norms	-.02	-.11	-.10	.13	-.09	.14	-.17	
Political orientation	-.08	.01	<b>.18**</b>	.05	<b>.16*</b>	<b>.21**</b>	<b>.14*</b>	<b>.21**</b>

Note: A positive correlation between Political Orientation and Moral Wrongness rating implies that more conservative participants made higher ratings. Bold cells indicate significant correlations at \* $p < .05$ . \*\* $p < .01$ .

$p < .001$ ,  $d = 1.94$ , 95% CI = [0.83, 1.21], indicating that moral images were indeed perceived to display something morally praiseworthy compared to immoral images judged to depict immoral and blameworthy content. Critically, these moral versus immoral differences were also statistically significant within each foundation-specific image category (Table 5 and Figure 4A). Likewise, images within the neutral category did not differ significantly in their moral valence ratings  $t(18) = 1.2$ ,  $p = .122$ ,  $d = 0.54$ , 95% CI = [-0.13, 0.48]. Yet, we also observed that eight images originally placed into the ‘immoral’ category and associated with a moral foundation were rated as moral ( $<3$ ), and four supposedly moral images were rated as immoral ( $>3$ ; moral valence ratings for each image item are summarized in SI Table 2).

Moreover, we investigated differences in arousal (Table 5 and Figure 4B). Similar to ratings on moral valence, high-arousal images received a higher arousal rating ( $M = 3.42$ ,  $SD = 0.39$ ) than low-arousal images ( $M = 2.70$ ,  $SD = 0.61$ ). This difference was again large in terms of effect size and statistically significant  $t(118) = 7.67$ ,  $p < .001$ ,  $d = 1.40$ , 95% CI = [0.53, 0.90]. Compellingly, these mean differences were statistically significant within each foundation-specific as well as neutral image category. Despite these averaged categorical differences, there were images whose arousal rating differed from the intended arousal category. Nine high-arousal items were rated with lower arousal ( $<3$ ), and 23 low-arousal items were rated with higher arousal ( $>3$ ; arousal ratings for each image item are summarized in SI Table 2). As morality ratings and valence ratings were again highly correlated ( $r = 0.73$ ,  $p < .001$ ), we provide no further analysis of valence ratings.

Thereafter, we tested whether participants rated the presence of moral foundations according to their intended foundation-specific image category (Tables 6 and 7 and Figure 4C). To this end, we conducted a series of independent, one-sided t-tests comparing the mean foundation rating for images of the intended foundation with the mean foundation rating of images across all other categories (e.g., mean rating of care-harm in images classified as care-harm compared to mean rating of care-harm for all other images). As expected, we observed that for all foundations, the corresponding images received significantly higher ratings on their foundation compared to images from all other categories (Table 7).

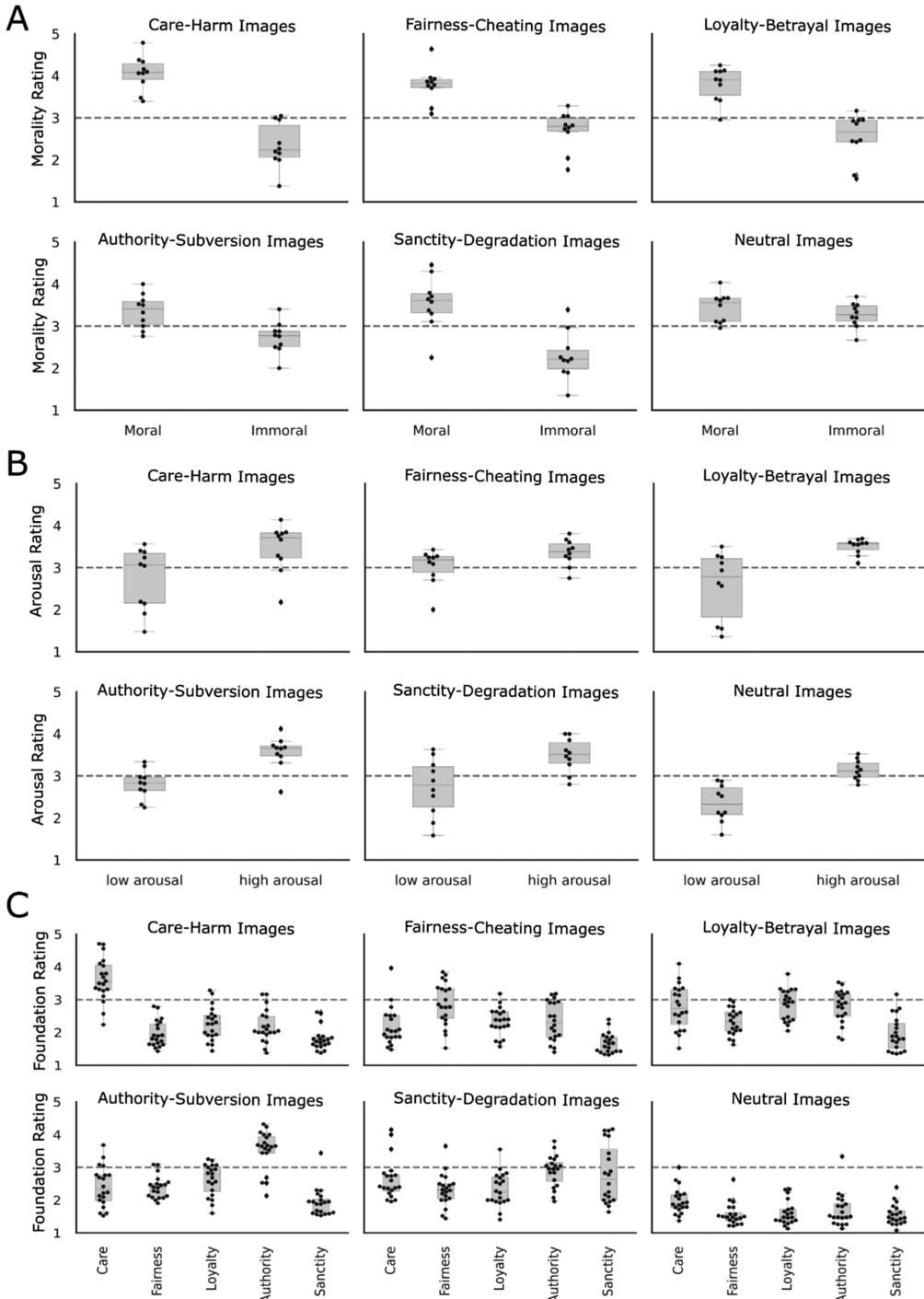
We also determined whether individual images received the intended foundation-specific ratings. To this end, we computed the mean foundation rating for each image and assigned each image to the foundation that received the highest mean rating. Likewise, the 20 images with the lowest mean foundation ratings were classified into the ‘neutral’ category. The resulting confusion matrix crossing intended and rated foundation is displayed in Table 8. Notably, 19 (95%) of the intended care images indeed received the highest care ratings across images, followed by 18<sup>2</sup> (90%) authority images and 16

<sup>2</sup>For three images, sanctity and authority were the highest mean ratings. We therefore added those images to both categories during calculations.

**Table 5.** Arousal and moral valence ratings across image categories.

Category	Moral		Immoral		Difference 95% CI				High arousal		Low arousal					
	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI					<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI				
Care	4.06	0.41 [3.76, 4.36]	2.34	0.53 [1.97, 2.72]	1.72 [1.27, 2.16]	3.62	8,09	<.001	3.47	0.58 [3.05,3.88]	2.74	0.74 [2.21,3.27]	0.73 [0.10,1.35]	1.09	2,45	.012
Fairness	3.78	0.42 [3.48, 4.08]	2.70	0.46 [2.37, 3.03]	1.08 [0.67, 1.50]	2.45	5,48	<.001	3.35	0.32 [3.13,3.58]	3.02	0.42 [2.72,3.33]	0.33 [-0.02,0.68]	0.89	1,99	.031
Loyalty	3.80	0.41 [3.51, 4.09]	2.54	0.56 [2.14, 2.94]	1.26 [0.80, 1.72]	2.58	5,78	<.001	3.50	0.19 [3.37,3.63]	2.58	0.80 [2.00,3.15]	0.92 [0.38,1.47]	1.59	3,56	.001
Authority	3.35	0.40 [3.06, 3.64]	2.73	0.38 [2.46, 2.99]	0.62 [0.26, 0.99]	1.60	3,57	.001	3.56	0.39 [3.28,3.84]	2.80	0.35 [2.55,3.05]	0.75 [0.40,1.10]	2.02	4,52	<.001
Sanctity	3.55	0.62 [3.11, 3.99]	2.28	0.57 [1.88, 2.69]	1.27 [0.71, 1.83]	2.13	4,77	<.001	3.49	0.41 [3.20,3.78]	2.72	0.69 [2.23,3.22]	0.77 [0.24,1.30]	1.36	3,03	.004
Neutral	3.44	0.35 [3.19, 3.69]	3.26	0.30 [3.05, 3.48]	0.18 [-0.13, 0.48]	0.54	1,2	.122	3.14	0.24 [2.96,3.31]	2.35	0.44 [2.03,2.66]	0.79 [0.46,1.12]	2.23	4,98	<.001

Note: *t* and *p* values are the results of independent, one-sided *t*-tests comparing moral > immoral and high arousal > low arousal for each image category separately.



**Figure 4.** Sociomoral image ratings. (A) Morality ratings for moral versus immoral images. (B) Arousal ratings for low versus high arousal images. (C) Foundation ratings for each moral foundation category. Each dot reflects the mean response of all participants to a single image. Box plots display median (center line), upper and lower quartiles (box limits), whiskers connote  $1.5 \times$  interquartile range (IQR) and points that fall outside the whiskers are outliers.

**Table 6.** Moral foundation ratings across intended image categories.

Intended image category	Moral foundation ratings									
	Care		Fairness		Loyalty		Authority		Sanctity	
	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI	<i>M</i>	SD 95% CI
Care	3.61	0.66 [3.31, 3.92]	1.96	0.40 [1.77, 2.15]	2.26	0.51 [2.02, 2.50]	2.19	0.50 [1.95, 2.42]	1.80	0.34 [1.64, 1.96]
Fairness	2.19	0.59 [1.91, 2.46]	2.87	0.65 [2.57, 3.18]	2.33	0.42 [2.13, 2.53]	2.34	0.59 [2.06, 2.62]	1.67	0.31 [1.53, 1.82]
Loyalty	2.78	0.68 [2.46, 3.09]	2.33	0.41 [2.14, 2.52]	2.86	0.47 [2.64, 3.08]	2.80	0.50 [2.57, 3.03]	1.97	0.53 [1.72, 2.21]
Authority	2.41	0.60 [2.13, 2.70]	2.39	0.34 [2.23, 2.55]	2.62	0.49 [2.39, 2.85]	3.52	0.60 [3.24, 3.80]	1.94	0.43 [1.73, 2.14]
Sanctity	2.64	0.62 [2.35, 2.93]	2.31	0.50 [2.08, 2.55]	2.33	0.51 [2.09, 2.57]	2.89	0.48 [2.66, 3.12]	2.79	0.89 [2.37, 3.21]
Neutral	1.98	0.40 [1.79, 2.16]	1.57	0.34 [1.41, 1.72]	1.60	0.36 [1.43, 1.77]	1.70	0.50 [1.46, 1.93]	1.55	0.31 [1.41, 1.70]

**Table 7.** Mean differences in foundation ratings across image categories.

Foundation	Difference			<i>d</i>	<i>t</i> (118)	<i>p</i>
	<i>M</i>	95% CI				
Care	1.21	[0.90, 1.53]		1.88	7.68	<.001
Fairness	0.76	[0.51, 1.02]		1.44	5.87	<.001
Loyalty	0.63	[0.36, 0.90]		1.14	4.67	<.001
Authority	1.14	[0.82, 1.46]		1.73	7.06	<.001
Sanctity	1.01	[0.75, 1.26]		1.92	7.85	<.001

*Note:* Results of five independent, one-sided *t*-tests. For each directional test, the average foundation rating of images within one foundation was compared against the average foundation rating across all other image categories (i.e., foundation ratings for foundation images > foundation ratings for all images not within the foundation).

**Table 8.** Confusion matrix comparing intended and rated image categories.

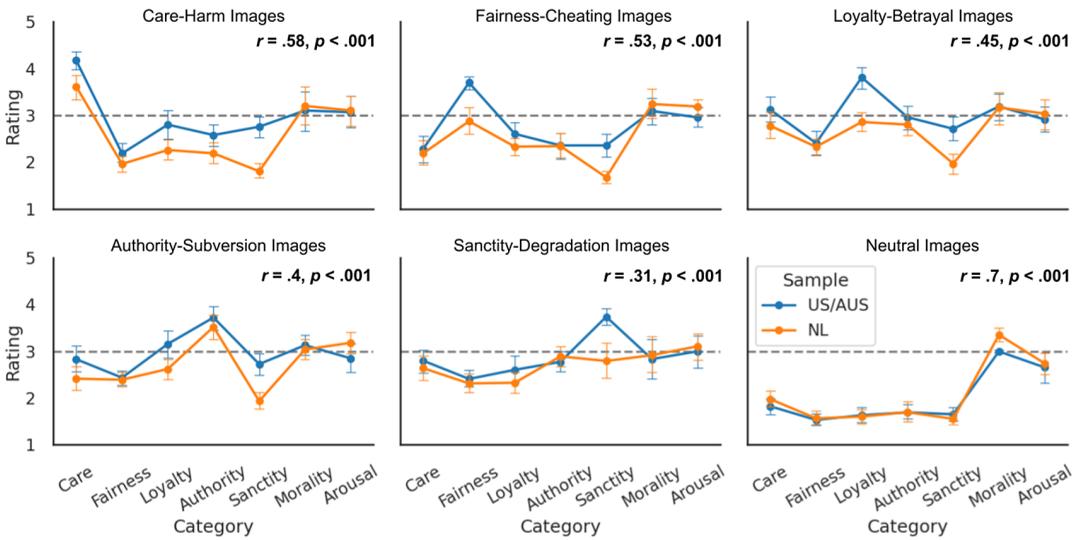
Category	Rated foundation							Sum	Recall	Accuracy	F1
	Care	Fairness	Loyalty	Authority	Sanctity	Neutral	Authority Sanctity				
Care	19	0	1	0	0	0	0	20	95,00	85,83	69,09
Fairness	2	10	1	5	0	2	0	20	50,00	90,83	64,52
Loyalty	7	0	7	4	0	1	1	20	35,00	86,67	46,67
Authority	0	0	1	17	0	1	1	20	90,00	82,50	63,16
Sanctity	4	1	0	7	7	0	1	20	40,00	88,33	53,33
Neutral	3	0	0	1	0	16	0	20	80,00	93,33	80,00
Sum	35	11	10	34	7	20	3	120	<b>65,00</b>	<b>87,92</b>	<b>62,79</b>
Precision	54,29	90,91	70,00	48,65	80,00	80,00			<b>63,37</b>		

*Note:* We assigned each image the foundation which received the maximum rating and assigned ‘Neutral’ to the 20 images with the lowest mean ratings. For three images, sanctity and authority both received the highest mean rating, therefore we added them to both, authority and sanctity, when we calculated our measures. The numbers in **bold** are weighted averages for the respective measure.

(80%) ‘neutral’ images. In contrast, discrepancies were larger for fairness images (10 images; 50%), sanctity (8<sup>2</sup> images; 35%), and loyalty (7 images, 35%). Across all images, 63% were rated according to their intended category, with an average accuracy of 87.92%, suggesting that even on the individual image level, the majority of images were correctly categorized into their intended foundation.

Next, we compared the image ratings from our Dutch respondents to those from US/Australian samples studied by Crone et al. (2018) during the creation of the SMID (Figure 5). We observed that image judgments were highly similar between the samples, with Pearson correlations ranging from .3 to .7 (all *p* < .001) across image categories. However, we did find that ratings of moral foundations showed nuanced differences between cultures. With the exception of authority-subversion images, ratings for the main moral foundation of each moral image class were higher in the US/Australian samples compared to the Netherlands. We return to an interpretation of these differences in the discussion section.

Lastly, we examined the correlation across all image rating categories and participants’ political orientation (Table 9). Higher ratings on each of the moral foundations correlated with distance of morality ratings from the midpoint of the scale, a metric that we termed ‘moral polarity’. In line with exemplification theory (Zillmann, 1999), this could imply that individuals who perceive an image to be more exemplary for a moral foundation also deem this image to be more moral or immoral. Interestingly, more morally polarized ratings did not correlate with arousal ratings (*r* = .03, *p* = .519).



**Figure 5.** Comparison of SMID ratings between US/Australian (US/AUS) samples (Crone et al., 2018) and respondents from the Netherlands (NL). The blue line denotes ratings from US/AUS samples, whereas the orange line reflects ratings from Dutch respondents. Error bars connotate 95% confidence intervals based on 1,000 bootstrap samples.

**Table 9.** Correlation table for image ratings.

	Care	Fairness	Loyalty	Authority	Sanctity	Moral valence	Moral polarity	Arousal
Fairness	.45**							
Loyalty	.49**	.71**						
Authority	.47**	.60**	.71**					
Sanctity	.41**	.52**	.51**	.47**				
Moral valence	-.10*	.01	.10*	.07	-.11*			
Moral polarity	.28**	.26**	.27**	.32**	.19**	.34**		
Arousal	.16**	.11*	.06	.10*	.08	-.31**	.03	
Political orientation	.02	.10*	.15**	.11*	.10*	.15**	.10*	-.03

Note: A positive correlation between political orientation and other ratings implies that more conservative participants made higher ratings. Moral polarity refers to how distant the rating was from the scale midpoint. \* $p < .05$ . \*\* $p < .01$ .

Rather, the more arousing an image, the less it was perceived to display something moral/praiseworthy ( $r = -.31, p < .001$ ). Replicating findings from Crone et al. (2018), all five foundation ratings were moderately correlated with each other (all  $r$ 's  $> .4, p < .001$ ), although all our pairwise foundation correlations were lower than those in the original study (Figure 4 in Crone et al., 2018). Again, fairness was strongly related to binding foundations. In particular, the highest foundation correlations were between fairness and loyalty ( $r = .71, p < .001$ ), fairness and authority ( $r = .60, p < .001$ ) and loyalty and authority ( $r = .71, p < .001$ ). Moreover, we found that ratings for all foundations were positively associated with more conservative political orientations—but the association between care and ideology was close to zero and not statistically significant. Note that in contrast to the MFV, foundation image ratings reflect how strongly participants perceived those foundations in the images and not how morally wrong they found those images to be. While conservatives tended to provide more polarized morality ratings overall ( $r = .10, p = .025$ ), this is likely driven by the fact that conservatives rated images as more moral compared to progressives ( $r = .15, p < .001$ ).

#### 4. Discussion

We adapted and validated two widely used moral stimulus sets for examining moral judgment in a Dutch sample. We translated the MFV (Clifford et al., 2015) into the Dutch language and selected a wide range of morally salient images from the SMID (Crone et al., 2018), which we then validated in a crowd-sourced sample from the Netherlands. These instruments offer advantages over alternatives by allowing participants to make moral judgments about specific situations (Crone et al., 2018; Marques et al., 2020; Schein, 2020).

The results of our MFV analysis suggest that we successfully adapted them to the Dutch context. Participants rated scenarios violating a moral foundation as more morally wrong than those describing social norm transgressions. Additionally, trends in moral wrongness ratings across MFV categories were similar to those reported in the original MFV study (Clifford et al., 2015). Furthermore, participants predominantly accurately identified the intended type of moral or social norm violation in the vignettes. These results suggest that we have successfully provided a valid and reliable MFV for the Dutch population.

Our demonstrated relationship between MFV moral wrongness ratings and political orientation only partially replicates prior findings. As Haidt and Graham (2007) argued and a meta-analysis by Kivikangas et al. (2021) confirmed, conservatives in the US usually judge the binding moral foundations as more morally relevant than progressives. Compellingly, this pattern also emerged in our study, and even social norms were rated more morally wrong by more right-leaning individuals. Yet, extant literature suggests that left-leaning (progressive) individuals in the US judge transgressions of individualizing foundations as more morally wrong than conservatives do (Graham et al., 2011; Kivikangas et al., 2021). However, we found no statistically significant associations between progressiveness and moral wrongness ratings of care violations, and wrongness ratings of fairness transgressions even showed a small to mid-sized association with conservatism. We reason that these discrepancies might be more driven by instruments than translational artifacts or genuine cultural differences.

On the one hand, the previously mentioned studies used the MFQ (Graham et al., 2011; Kivikangas et al., 2021), whereas we used the MFV. In the original MFV paper, fairness was unrelated to political orientation (Clifford et al., 2015). Analogously, Hopp et al. (2023), also using the MFV in a US college sample, found the same pattern as we did. On the other hand, Van Leeuwen and Park (2009) used the MFQ with Dutch participants and reported that fairness was associated with a more progressive political orientation. Why is fairness sometimes related to progressive political orientation and sometimes not? According to Janoff-Bulman (2023) and Atari et al. (2023), MFT omits the distinction between two kinds of distributional justice: equality and proportionality. Participants may interpret fairness items in the MFQ as questions of equality, which aligns with progressive concerns (Jost, 2017). In contrast, proportionality may be more associated with conservative political orientation (Lee et al., 2018). To clarify these relationships, future versions of the MFV may incorporate scenarios related to both equality and proportionality.

The SMID ratings analysis validated our image selection for studying Dutch moral judgment. We identified images that consistently evoked moral or immoral ratings across various moral foundations, while neutrally classified images were consistently rated as having neutral moral content and low appeal to all moral foundations. Furthermore, we offer evidence that foundation-specific images can be identified. Images primarily showcasing one moral foundation reliably elicited stronger perceptions of that foundation compared to images emphasizing other moral foundations. Mirroring our MFV results, when rating photographic images, Dutch conservatives also perceived a greater degree of loyalty, authority, sanctity, and fairness than progressives. This may indicate conservatives' greater recognition of these foundations and suggests that higher moral wrongness ratings for MFV and similar stimuli might result not only from greater relevance assigned to these foundations but also from more frequent recognition of morality in various contexts. Future studies should dissect these influences and interaction of perception and evaluation in moral judgment.

Importantly, although general trends in image ratings were highly similar between US/Australian and Dutch samples, notable differences in the perception of particular moral foundations did emerge between the samples. There are at least three explanations for this: First, these differences were to be expected as we sampled images from the SMID that received the highest ratings from US/Australian samples for each of the moral foundations. Second, they nevertheless demonstrate that moral judgments of photographic images are modulated by cultural differences, even across nations within the Western, Educated, Industrialized, Rich, and Democratic (WEIRD, Henrich et al., 2010) context. Third, this finding also invites future research to determine images that are more moralized within Dutch culture. For instance, images rated as high in Sanctity-Degradation by US/Australian respondents primarily displayed prostitution and recreational drug use. Yet, Dutch individuals are generally tolerant toward prostitution (Jonsson and Jakobsson, 2017) and illicit drug use (van der Sar et al., 2012). In turn, being confronted with these concepts may not have triggered moral emotions (e.g., disgust) in Dutch citizens that precede moralization (Clifford, 2019). Interestingly, although the Sanctity-Degradation ratings were lower in Dutch than in US/Australian samples for all image categories—with the exception of Neutral images—our study does not suggest that the moral foundation of Sanctity-Degradation is unknown to Dutch individuals: Respondents in the Netherlands did rate images that primarily display Sanctity-Degradation as higher in that foundation compared to images from other categories. Analogously, vignettes that described violations of Sanctity-Degradation were also accurately classified into the Sanctity-Degradation category by Dutch respondents.

#### 4.1. Limitations

This study has limitations. We used a crowd-sourced approach common for affective datasets (Crone et al., 2018; Hopp et al., 2021) and had each participant rate only a fraction of stimuli. This enabled us to simultaneously investigate two large stimuli sets, yet it also came at the cost of only around 20 ratings per stimulus. These subsamples are smaller than in classic scale development or validation studies (e.g., Clifford et al., 2015; Graham et al., 2011) and do not allow for factor analysis. Hence, we provide a fertile ground for future studies employing the full range of our adapted MFV and SMID within a repeated-measured design. Furthermore, we employed the same response options as the original MFV study in the United States (Clifford et al., 2015) and ensuing MFV validations in Brazil (Marques et al., 2020). However, behavioral data from these studies have not been made openly accessible, hindering a direct comparison. We therefore invite our fellow researchers to openly share their data in order to enable a broader comparison of MFV responses across the globe.

Moreover, both stimulus sets feature scenarios that were derived from MFT's original moral taxonomy. In view of MFT's recent advancements that split the fairness foundation into equality and proportionality (Atari et al., 2023), future studies should design and validate vignettes that distinctly feature these new foundations. For instance, the recently introduced Moral Foundations Questionnaire 2 (MFQ-2; Atari et al., 2023) features this updated set of moral foundations and may serve as a template to construct concrete scenarios pertaining to equality and proportionality. At the same time, discrepancies in the inclusion, exclusion, and adaptations of moral foundations across tasks and studies hinders reproducibility. For example, the authors of the SMID did not include the liberty foundation (Iyer et al., 2012) due to ongoing debates concerning its status as a distinct moral foundation (Crone et al., 2018). Because we aimed to reproduce the original MFV and SMID ratings as closely as possible, we likewise did not collect Liberty ratings for the SMID. Yet, moral psychology should strive for a more standardized and unified approach in its employment of moral judgment scales (Malle, 2021).

Furthermore, there is some general critique of MFT arguing that all moral transgressions boil down to a single essence, such as harm (Schein and Gray, 2017). Yet, mounting arguments from philosophy (Sinnott-Armstrong and Wheatley, 2012, 2014), behavioral (Sackris and Larsen, 2022), and neuroimaging studies (Hopp et al., 2023; Parkinson et al., 2011) refute morality's unity on any level of explanation (e.g., content, function, neurobiological, etc.). We agree with such arguments and a pluralistic approach to morality. That said, there exist other candidate frameworks that explain (cross-cultural) variability in

moral judgments, including the recently introduced Morality as Cooperation (MaC) theory (Curry et al., 2019). Exploring how moral judgments of scenarios derived from competing moral theories converge and diverge across the globe presents a promising avenue for future research.

Analogously, the relationship between stimuli and political orientation rather reflects a starting point for future research than a mere limitation. For both stimuli sets, our results do not exactly replicate the expected left-right pattern, and the stimuli as such—and not only culture—present themselves as a possible explanation for our findings. This may be because of the translations we made, the sample we used, or the context that we studied (the Netherlands). We can only speculate, but think the best answer to this question lies in future research. Future studies should determine whether the differences in fairness rating for the MFV is due to wording or actual difference between general principles and actions (as this is the purported difference between MFV and MFQ; Clifford et al., 2015). For the SMID images, there are two possible ways to investigate the origin of our results. If we assume that method drives our findings, the variance in morality ratings should be increased. Thus, future studies should use our or alternative sampling procedures (Crone et al., 2018) to choose pictures which best discriminate between political orientation. In addition, if we assume that modality modulates moral judgment, neuroimaging studies may dissociate which processes are independent from modality and which are shared.

#### **4.2. Future outlook and conclusion**

Notwithstanding these limitations, several promising options exist for researchers interested in using our adapted version of the MFV and SMID for studying moral decision-making in Dutch populations. First, we suggest that future studies employ the full range of our adapted MFV and SMID, either jointly or separately, in a repeated-measures, within-subject moral judgment paradigm. Doing so will not only prove beneficial for examining the factor structure of our adapted MFV and SMID stimuli, but can also reveal novel insights into the moral decision-making process of Dutch individuals. Moral wrongness ratings, aggregated by moral foundations in either the MFV or SMID, can be taken as proxies for respondents' sensitivity toward concrete moral transgressions. In turn, individual differences in these moral judgments may reveal which concrete moral scenarios contribute to Dutch citizen's affective polarization (Harteveld, 2021). Second, the highly controlled structure and norming of these datasets renders them an ideal measurement tool for probing the cognitive neuroscience of moral judgment (Hopp et al., 2023; Khoudary et al., 2022). Because our knowledge concerning the neuroscience of moral cognition remains heavily US-centered, presenting our adapted MFV and SMID to Dutch individuals undergoing neuroimaging may prove useful for studying cross-cultural variation found for judgments of moral actions (Graham et al., 2011). Given that cultural differences may influence both the location and levels of observed neural activity (for a review see Han and Northoff, 2008), cross-cultural investigations into the neuroscience of morality hold immense future potential. Lastly, future work should continue to probe the relative (dis)advantages of different moral measurement instruments for advancing the central tenets of MFT. We argue that MFQs remain cost-effective tools for researchers aiming to advance predictions concerning individual, sociopolitical, and cross-cultural variations in the abstract and general endorsement of moral foundations. At the same time, probing the existence of moral foundations via responses to vicarious (i.e., third-party) transgressions of moral foundations calls for moral judgment paradigms (e.g., MFV and SMID) that feature stimuli displaying concrete violations of moral foundations. Alternatively, we reason that combining both abstract and concrete moral paradigms may prepare us well to advance MFT by illuminating for whom and in which contexts general moral priorities converge and diverge with contextualized realizations of moral actions (Eriksson et al., 2019; Hull et al., 2024).

Taken together, our results show successful adaptation of the MFV and SMID for studying moral judgment in Dutch populations. Our herein developed work underscores the importance of exploring cultural differences, especially for nonverbal moral stimuli, and emphasizes the need for more stimulus validation in social psychology and personality science.

**Data availability statement.** The behavioural data that support the findings of this study, the analysis code, and the experimental stimuli are available at The Open Science Framework platform (<https://osf.io/9gnza/>).

**Author contributions.** Conceptualization: F.R.H., B.N.B.; Data curation: B.J.; Formal analysis: F.R.H., B.J.; Funding acquisition: F.R.H., B.N.B.; Investigation: F.R.H., E.K.; Methodology: F.R.H., E.K.; Project administration: F.R.H.; Writing—original draft: F.R.H., B.J., E.K.; Writing—review and editing: B.N.B.

**Funding statement.** F.R.H. and B.N.B. acquired funding from the Amsterdam School of Communication Research (Grant No. ASCoR-u-2022-Hopp). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

**Competing interest.** The authors declare no competing interests.

## References

- Atari, M., Graham, J., & Dehghani, M. (2020). Foundations of morality in Iran. *Evolution and Human Behavior*, *41*(5), 367–384.
- Atari, M., & Haidt, J. (2023). Ownership is (likely to be) a moral foundation. *Behavioral and Brain Sciences*, *46*, e326.
- Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, *125*(5), 1157–1188. <https://doi.org/10.1037/pspp0000470>
- Baciu, M., Ans, B., & Carbonnel, S. (2002). Length effect during word and pseudo-word reading. An event-related fMRI study. *Neuroscience Research and Communications*, *30*(3), 155–165.
- Bobbio, A., Nencini, A., & Sarrica, M. (2011). Il Moral Foundation Questionnaire: Analisi della struttura fattoriale della versione Italiana. [The Moral Foundation Questionnaire. Factorial structure of the Italian version.]. *Giornale Di Psicologia*, *5*(1–2), 7–18.
- Chunyu, M., Zommarà, N. M., Ounjai, K., Ju, X., & Lauwereyns, J. (2021). *Speed is associated with, but does not cause, polarization in the moral evaluation of real-world images* [Preprint]. In Review. <https://doi.org/10.21203/rs.3.rs-1122225/v1>
- Church, J. A., Balota, D. A., Petersen, S. E., & Schlaggar, B. L. (2011). Manipulation of length and lexicality localizes the functional neuroanatomy of phonological processing in adult readers. *Journal of Cognitive Neuroscience*, *23*(6), 1475–1493. <https://doi.org/10.1162/jocn.2010.21515>
- Clifford, S. (2017). Individual differences in group loyalty predict partisan strength. *Political Behavior*, *39*(3), 531–552. <https://doi.org/10.1007/s11109-016-9367-3>
- Clifford, S. (2019). How emotional frames moralize and polarize political attitudes. *Political Psychology*, *40*(1), 75–91.
- Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations vignettes: A standardized stimulus database of scenarios based on moral foundations theory. *Behavior Research Methods*, *47*(4), 1178–1198. <https://doi.org/10.3758/s13428-014-0551-2>
- Crone, D. L., Bode, S., Murawski, C., & Laham, S. M. (2018). The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes. *PLOS ONE*, *13*(1), e0190954. <https://doi.org/10.1371/journal.pone.0190954>
- Curry, O. S., Jones Chesters, M., & Van Lissa, C. J. (2019). Mapping morality with a compass: Testing the theory of ‘morality-as-cooperation’ with a new questionnaire. *Journal of Research in Personality*, *78*, 106–124. <https://doi.org/10.1016/j.jrp.2018.10.008>
- De Buck, A., & Pauwels, L. J. R. (2023). Moral Foundations Questionnaire and Moral Foundations Sacredness Scale: Assessing the Factorial Structure of the Dutch Translations. *Psychologica Belgica*, *63*(1), 92–104. <https://doi.org/10.5334/pb.1188>
- Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J., Vaisey, S., Iliev, R., & Graham, J. (2016). Purity homophily in social networks. *Journal of Experimental Psychology: General*, *145*(3), 366–375. <https://doi.org/10.1037/xge0000139>
- Dodd, M. D., Balzer, A., Jacobs, C. M., Gruszczynski, M. W., Smith, K. B., & Hibbing, J. R. (2012). The political left rolls with the good and the political right confronts the bad: Connecting physiology and cognition to preferences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1589), 640–649.
- Eriksson, K., Simpson, B., & Strimling, P. (2019). Political double standards in reliance on moral foundations. *Judgment and Decision Making*, *14*(4), 440–454.
- Frimer, J. A., Biesanz, J. C., Walker, L. J., & MacKinlay, C. W. (2013). Liberals and conservatives rely on common moral foundations when making moral judgments about influential people. *Journal of Personality and Social Psychology*, *104*(6), 1040–1059. doi: <https://doi.org/10.1037/a0032277>
- 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*(5), 10291046. <https://doi.org/10.1037/a0015141>
- 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*, 55130. <http://doi.org/10.1016/B978-0-12-407236-7.00002-4>

- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125–130. <https://doi.org/10.1016/j.copsyc.2015.09.007>
- Graham, J., Nosek, B. A., & Haidt, J. (2012). The moral stereotypes of liberals and conservatives: Exaggeration of differences across the political spectrum. *PLOS ONE*, 7(12), e50092.
- 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(2), 366–385. <https://doi.org/10.1037/a0021847>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science*, 3(1), 65–72. <https://doi.org/10.1111/j.1745-6916.2008.00063.x>
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1), 98–116. <https://doi.org/10.1007/s11211-007-0034-z>
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55–66. <https://doi.org/10.1162/0011526042365555>
- Han, S., & Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: A transcultural neuroimaging approach. *Nature Reviews Neuroscience*, 9, 646–654.
- Harenski, C. L., Antonenko, O., Shane, M. S., & Kiehl, K. A. (2008). Gender differences in neural mechanisms underlying moral sensitivity. *Social Cognitive and Affective Neuroscience*, 3(4), 313–321. <https://doi.org/10.1093/scan/nsn026>
- Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, 30(1), 313–324. [10.1016/j.neuroimage.2005.09.034](https://doi.org/10.1016/j.neuroimage.2005.09.034)
- Harper, C. A., & Rhodes, D. (2021). Reanalysing the factor structure of the Moral Foundations Questionnaire. *British Journal of Social Psychology*, 60(4), 1303–1329. <https://doi.org/10.1111/bjso.12452>
- Harteveld, E. (2021). Fragmented foes: Affective polarization in the multiparty context of the Netherlands. *Electoral Studies*, 71, 102332. <https://doi.org/10.1016/j.electstud.2021.102332>
- Heekeren, H. R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H.-P., & Villringer, A. (2005). Influence of bodily harm on neural correlates of semantic and moral decision-making. *NeuroImage*, 24(3), 887–897. <https://doi.org/10.1016/j.neuroimage.2004.09.026>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83.
- Hester, N., & Gray, K. (2020). The moral psychology of raceless, genderless strangers. *Perspectives on Psychological Science*, 15(2), 216–230. <https://doi.org/10.1177/1745691619885840>
- Hopp, F. R., Amir, O., Fisher, J., Grafton, S., Sinnott-Armstrong, W., & Weber, R. (2023). Moral foundations elicit shared and dissociable cortical activation modulated by political ideology. *Nature Human Behavior*, 7, 2182–2198. <https://doi.org/10.1038/s41562-023-01693-8>
- Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, 53(1), 232–246. <https://doi.org/10.3758/s13428-020-01433-0>
- Hopp, F. R., & Weber, R. (2021). Reflections on extracting moral foundations from media content. *Communication Monographs*, 88(3), 371–379. <https://doi.org/10.1080/03637751.2021.1963513>
- Hull, K., Warren, C., & Smith, K. (2024). Politics makes bastards of us all: Why moral judgment is politically situational. *Political Psychology*. <https://doi.org/10.1111/pops.12954>
- Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLOS ONE*, 7(8), e42366. <https://doi.org/10.1371/journal.pone.0042366>
- Janoff-Bulman, R. (2023). *The two moralities: Conservatives, liberals, and the roots of our political divide*. New Haven, CT: Yale University Press.
- Jonsson, S., & Jakobsson, N. (2017). Is buying sex morally wrong? Comparing attitudes toward prostitution using individual-level data across eight Western European countries. *Women's Studies International Forum*, 61, 58–69.
- Jost, J. T. (2017). Ideological asymmetries and the essence of political psychology. *Political Psychology*, 38(2), 167–208. <https://doi.org/10.1111/pops.12407>
- Khoudary, A., Hanna, E., O'Neill, K., Iyengar, V., Clifford, S., Cabeza, R., De Brigard, F., & Sinnott-Armstrong, W. (2022). A functional neuroimaging investigation of moral foundations theory. *Social Neuroscience*, 17(6), 491–507. <https://doi.org/10.1080/17470919.2022.2148737>
- Kim, K. R., Kang, J.-S., & Yun, S. (2012). Moral intuitions and political orientation: Similarities and differences between South Korea and the United States. *Psychological Reports*, 111(1), 173–185. <https://doi.org/10.2466/17.09.21.PR0.111.4.173-185>
- Kivikangas, J. M., Fernández-Castilla, B., Järvelä, S., Ravaja, N., & Lönnqvist, J.-E. (2021). Moral foundations and political orientation: Systematic review and meta-analysis. *Psychological Bulletin*, 147(1), 55–94. <https://doi.org/10.1037/bul0000308>
- Kramer, C. Y. (1956). Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 12(3), 307–310. <https://doi.org/10.2307/3001469>
- Lee, Y., Yoon, S., Lee, Y. W., & Royne, M. B. (2018). How liberals and conservatives respond to equality-based and proportionality-based rewards in charity advertising. *Journal of Public Policy & Marketing*, 37(1), 108–118. <https://doi.org/10.1509/jppm.16.180>

- Malik, M., Hopp, F. R., Chen, Y., & Weber, R. (2021). Does regional variation in pathogen prevalence predict the moralization of language in COVID-19 news? *Journal of Language and Social Psychology, 40*(5–6), 653–676. <https://doi.org/10.1177/0261927X211044>
- Malle, B. F. (2021). Moral judgments. *Annual Review of Psychology, 72*, 293–318.
- Marques, L. M., Clifford, S., Iyengar, V., Bonato, G. V., Cabral, P. M., Dos Santos, R. B., Cabeza, R., Sinnott-Armstrong, W., & Boggio, P. S. (2020). Translation and validation of the moral foundations vignettes (MFVs) for the Portuguese language in a Brazilian sample. *Judgment and Decision Making, 15*(1), 149–158. <https://doi.org/10.1017/S1930297500006963>
- McCurrie, C. H., Crone, D. L., Bigelow, F., & Laham, S. M. (2018). Moral and Affective Film Set (MAAFS): A normed moral video database. *PLOS ONE, 13*(11), e0206604. <https://doi.org/10.1371/journal.pone.0206604>
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience, 22*(7), 2730–2736.
- Parkinson, C., Sinnott-Armstrong, W., Korulus, 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*(10), 3162–3180. [https://doi.org/10.1162/jocn\\_a\\_00017](https://doi.org/10.1162/jocn_a_00017)
- Puryear, C., Kubin, E., Schein, C., Bigman, Y., & Gray, K. (2022). Bridging political divides by correcting the basic morality bias [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/fk8g6>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.
- Sackris, D., & Larsen, R. R. (2022). The disunity of moral judgment: Evidence and implications. *Philosophical Psychology, 37*(2), 351–370.
- Schaich Borg, J., Lieberman, D., & Kiehl, K. A. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience, 20*(9), 1529–1546. <https://doi.org/10.1162/jocn.2008.20109>
- Schaich Borg, J., Sinnott-Armstrong, W., Calhoun, V. D., & Kiehl, K. A. (2011). Neural basis of moral verdict and moral deliberation. *Social Neuroscience, 6*(4), 398–413. <https://doi.org/10.1080/17470919.2011.559363>
- Schein, C. (2020). The importance of context in moral judgments. *Perspectives on Psychological Science, 15*(2), 207–215. <https://doi.org/10.1177/1745691620904083>
- Schein, C., & Gray, K. (2017). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review. Advance online publication.* <http://dx.doi.org/10.1177/1088868317698288>
- Sinnott-Armstrong, W., & Wheatley, T. (2012). The disunity of morality and why it matters to philosophy. *The Monist, 95*(3), 355–377.
- Sinnott-Armstrong, W., & Wheatley, T. (2014). Are moral judgments unified? *Philosophical Psychology, 27*(4), 451–474.
- Sudo, R., Nakashima, S. F., Ukezono, M., Takano, Y., & Lauwereyns, J. (2021). The role of temperature in moral decision-making: Limited reproducibility. *Frontiers in Psychology, 12*, 681527. <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.681527>
- Tao, D., Leng, Y., Huo, J., Peng, S., Xu, J., & Deng, H. (2022a). Effects of core disgust and moral disgust on moral judgment: An event-related potential study. *Frontiers in Psychology, 13*, 806784. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.806784>
- Tao, D., Leng, Y., Peng, S., Xu, J., Ge, S., & Deng, H. (2022b). Temporal dynamics of explicit and implicit moral evaluations. *International Journal of Psychophysiology, 172*, 1–9. <https://doi.org/10.1016/j.ijpsycho.2021.12.006>
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*(5), 853–870. <https://doi.org/10.1037/0022-3514.78.5.853>
- Van der Sar, R., Ødegård, E., Rise, J., Brouwers, E. P. M., Van de Goor, L. A. M., & Garretsen, H. F. L. (2012). Acceptance of illicit drug use in the Netherlands and Norway: A cross-national survey. *Drugs: Education, Prevention and Policy, 19*(5), 397–405.
- van Leeuwen, F., & Park, J. H. (2009). Perceptions of social dangers, moral foundations, and political orientation. *Personality and Individual Differences, 47*(3), 169–173. <https://doi.org/10.1016/j.paid.2009.02.017>
- Wagemans, F. M. A., Brandt, M. J., & Zeelenberg, M. (2018). Disgust sensitivity is primarily associated with purity-based moral judgments. *Emotion, 18*(2), 277–289. <https://doi.org/10.1037/emo0000359>
- Zakharin, M., & Bates, T. C. (2021). Remapping the foundations of morality: Well-fitting structural model of the Moral Foundations Questionnaire. *PLOS ONE, 16*(10), e0258910. <https://doi.org/10.1371/journal.pone.0258910>
- Zillmann, D. (1999). Exemplification theory: Judging the whole by some of its parts. *Media Psychology, 1*(1), 69–94. [https://doi.org/10.1207/s1532785xmep0101\\_5](https://doi.org/10.1207/s1532785xmep0101_5)

**Cite this article:** Hopp, F. R., Jargow, B., Kouwen, E., and Bakker, B. N. (2024). The Dutch moral foundations stimulus database: An adaptation and validation of moral vignettes and sociomoral images in a Dutch sample. *Judgment and Decision Making, e10*. <https://doi.org/10.1017/jdm.2024.5>