The Socio-Moral Image Database (SMID): A novel stimulus set for the study of social, moral and affective processes
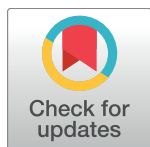
Damien L. Crone[1] *, Stefan Bode[1], Carsten Murawski[2], Simon M. Laham[1]

1 Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia,
2 Department of Finance, University of Melbourne, Melbourne, Australia

* croned@unimelb.edu.au

Abstract

A major obstacle for the design of rigorous, reproducible studies in moral psychology is the lack of suitable stimulus sets. Here, we present the Socio-Moral Image Database (SMID), the largest standardized moral stimulus set assembled to date, containing 2,941 freely available photographic images, representing a wide range of morally (and affectively) positive, negative and neutral content. The SMID was validated with over 820,525 individual judgments from 2,716 participants, with normative ratings currently available for all images on affective valence and arousal, moral wrongness, and relevance to each of the five moral values posited by Moral Foundations Theory. We present a thorough analysis of the SMID regarding (1) inter-rater consensus, (2) rating precision, and (3) breadth and variability of moral content. Additionally, we provide recommendations for use aimed at efficient study design and reproducibility, and outline planned extensions to the database. We anticipate that the SMID will serve as a useful resource for psychological, neuroscientific and computational (e.g., natural language processing or computer vision) investigations of social, moral and affective processes. The SMID images, along with associated normative data and additional resources are available at https://osf.io/2rqad/.

Introduction

In fields such as affective science, large, diverse and systematically validated stimulus sets (e.g., [1–10]) have facilitated substantial scientific progress [11,12]. Such stimulus sets enable rigorous studies that can be easily compared, replicated and aggregated, and moreover, obviate the need for individual research groups to duplicate each other's efforts in conducting labor-intensive stimulus validation (for discussion, see [13,14]). In the field of moral psychology, however, there is a clear shortage of standardized stimulus sets. This shortage places substantial constraints both on the kinds of paradigms that can feasibly be implemented, and on the reliability and validity of the conclusions researchers reach.

Motivated by this gap in the literature, we present the development and validation of a new picture set, the Socio-Moral Image Database (SMID). The SMID is the largest systematically

validated moral stimulus set assembled to date, containing images with diverse moral content. This image set will facilitate a wide range of novel research, including basic inquiries into the cognitive and neural underpinnings of moral perception, evaluation and judgment [15–17], applied research on moral communication and persuasion [18–20], and large-scale computational investigations of moral content [21].

Existing moral psychology stimulus sets

Currently, much moral psychology research relies on small, *ad hoc* stimulus sets, constructed with little (if any) prior validation. Such practices have several undesirable consequences. Specifically, *ad hoc* stimulus sets present obstacles to comparing and integrating findings across studies, and are relatively more vulnerable to unwanted confounds (e.g., see [22–25]). An obvious remedy is for moral psychology to employ standardized, well-validated materials, however only a small number of systematically validated moral stimulus sets exist (e.g., [26–29]; for a historically-oriented perspective, see [30,31]). Unfortunately, each suffers from important limitations.

Specifically, existing sets are often (1) brief, textual representations of moral content, (2) representative of only a narrow segment of the moral domain (e.g., sacrificial dilemmas) and normed on only a limited number of variables, (3) based on an assumption that each stimulus specifically represents one and only one class of moral content, or (4) restricted to one pole (typically the negative pole) of the moral spectrum. A summary of these features as they apply to prominent existing stimulus sets is presented in Table 1, along with a comparison to the SMID. In the following paragraphs, we elaborate on methodological challenges relating to each of these features, and how the SMID can contribute to overcoming them.

Challenge 1: Text-bound morality is inherently limited. In the real world, morality is not limited to a single modality; a wide range of mediums can elicit moral evaluations, from spoken and written words, to images, videos, and actual social interactions. To the extent that the moral psychology literature is built predominantly upon any single medium, it can provide only a partial understanding of morality. Unfortunately, nearly all existing moral stimulus sets

Table 1. Comparison of contemporary moral stimulus sets with the SMID.

Stimulus Set	Medium	*N* Stimuli	Breadth of Normative Data		Moral Valence		
			Moral Content	Non-Moral Content	Pos.	Neut.	Neg.
Chadwick et al. [27][a]	Text	500	Each stimulus rated on *one and only one* of the following dimensions: Charitability, Uncharitability, Cooperativeness, Uncooperativeness, Honesty, Dishonesty, Loyalty, Disloyalty, Friendliness, Unfriendliness	–	×		×
Clifford et al. [26][b]	Text	132	All stimuli rated on: Wrongness and *categorized as one of* a Care, Fairness, Ingroup, Authority, Purity, Liberty, or Other violation	All stimuli rated on: Comprehensibility, Imageability, Frequency, Strength of emotional response			×
Knutson et al. [29]	Text	312	All stimuli rated on: Wrongness, Harm, Benefit to actor, Benefit to others, Premeditation, Legality, Social rule violation	All stimuli rated on: Valence, Arousal, Sociality, Frequency, Personal familiarity, General familiarity	×	×	×
Lotto et al. [28][c]	Text	75	All stimuli rated on: Whether rater would perform sacrifice, Acceptability of performing sacrifice	All stimuli rated on: Valence, Arousal			×
SMID	Images	2,941	All stimuli rated on Wrongness, Care, Fairness, Ingroup, Authority, Purity	All stimuli rated on: Valence, Arousal	×	×	×

[a]Each stimulus in the Chadwick et al. set was rated only on the single dimension it was assumed to be relevant to.

[b]The Clifford et al. set also includes eight factor-analytically-derived dimensions based on wrongness judgments, which map onto the six moral foundations (including Liberty, and with three variants of harm).

[c]The Lotto et al. set consists of sacrificial dilemmas which we tentatively classify as having negative moral valence.

are text-based (the lone potential exception being the Geneva Affective Picture Database [2], discussed in further detail in S1 Text).

Unsurprisingly, most studies rely on text stimuli (e.g., see Table 1 of [32] and Table 1 of [33]). The near exclusive reliance on text imposes two constraints on the range of possible study designs. The first constraint concerns the time and effort required for participants to read a large set of vignettes. To present rich stimuli resembling everyday moral phenomena, researchers have few alternatives to detailed vignettes that require sustained concentration to process. The use of such vignettes raises a dilemma for researchers, in which they must choose between either protracted testing durations if the stimulus set is large, or reduced validity and statistical power if the stimulus set is small [34–38] (regarding the latter, statistical power can asymptote well below 1 when stimulus variability is high and the number of stimuli is low [36]). The time constraints imposed by text-based stimuli are especially pertinent in neuroimaging studies in which, for example, presenting a single sacrificial dilemma can require over half a minute of scanning time (e.g., [39,40]), compared to the mere seconds required to present an image stimulus ([15,41]).

A second constraint imposed by text-based stimuli is that such stimuli can be difficult to transport into paradigms that target rapid affective or intuitive psychological processes that are central to many current theories of moral cognition [40,42–47]. Although researchers can potentially employ such paradigms with the presentation of one or a few words [48–51], this comes at the expense of the richness and realism that can be achieved with other mediums (consider the vastly different psychological experiences elicited by seeing the word "assault" as opposed to seeing an image or video of assault, let alone witnessing assault in person).

Moreover, studies using many single-word stimuli may require matching on several potentially confounding factors (e.g., length, frequency). Given the finite number of words in a language however, a suitable set of word stimuli simply may not be identifiable [52]. This is far less of a concern for images. Whereas the English lexicon is estimated to contain around one million words [53], a Flickr search for the term "house" currently returns around 3.5 million unique images. Furthermore, irrelevant features of image stimuli can often be modified to reduce confounds (for example by cropping, altering color composition, etc.), whereas analogous modifications are far less feasible for word stimuli [52].

Beyond these practical considerations, researchers must also consider the possibility that text-based stimuli may be processed differently to non-textual portrayals of the same content (e.g., [54,55]). According to Construal Level Theory (CLT [56,57]), presenting words (vs. images) promotes a high-level construal [58,59]. In the context of moral judgments, recent studies suggest that high-level construal may be associated with greater attention to ends vs. means [60], greater value-behavior consistency [61], and emphasis on different moral values [62–64]. Thus, the CLT literature suggests that presentation medium may systematically influence multiple aspects of moral cognition and behavior. Accordingly, overreliance on any single medium (here, text) may inadvertently distort theoretical accounts of moral psychological phenomena.

Developing a stimulus set that permits rapid presentation of rich, realistic moral content (and differs from the currently dominant medium) would thus help address significant challenges facing much existing moral psychology research. Of the forms such a stimulus set could take, arguably the most versatile medium is images. Image stimuli have a long history in moral psychology [65–68] (apparently even predating sliced bread [69]), offering a set of unique benefits over text-based stimuli. Images can be used in populations where linguistic stimuli are problematic [66,70]. Moreover, images can often be used similarly to text-based stimuli in explicit (or active) paradigms in which participants have their attention directed towards, and are required to deliberate upon, the moral content of the stimuli (e.g., sacrificial dilemmas).

Unlike text-based stimuli however, images can also readily be used in implicit (or passive) paradigms [15–17,50,71–73], without sacrificing richness. For these reasons, we chose images as our preferred modality.

Challenge 2: Morality is diverse. At present, there exists a large, historically rich literature characterizing the various ways in which people differ in their moral concerns, and in which situations differ in their moral content. Within existing moral psychology stimulus sets however, far less attention has been paid to systematic variation in moral content. Accounting for diversity in the content of moral stimuli is critical, given recent findings that the processing of different kinds of moral content has been shown to (1) recruit, require, or be moderated by, different psychological processes (e.g., [63,74–76]), (2) result in different patterns of inferences in the context of character judgments [77,78], and (3) be differentially affected by psychopathology [79–82]. This implies that if researchers restrict stimuli to a single domain of moral content (e.g., instances of harm), findings cannot be generalized beyond that moral domain, and certainly cannot be considered representative of morality in its entirety. Likewise, if researchers ignore variation in moral content by treating morality as a homogenous entity, findings will either be substantially noisier, or skewed by whatever unaccounted-for moral content happens to dominate the selected stimuli [83].

To address diversity of moral content, we used Moral Foundations Theory (MFT) as an organizing framework for the initial development and description of the image set [84–87]. MFT posits five purportedly innate foundational moral values: (1) *Care*, concerned with prevention or alleviation of suffering, (2) *Fairness*, concerning identification of cheating and exploitation, (3) *Ingroup*, concerned with self-sacrifice for group benefit and preventing betrayal, (4) *Authority*, concerned with respecting and obeying superiors, and (5) *Purity*, concerned with avoiding pathogens through, for example, regulation of sexual and eating behaviors. A tentative sixth foundation, *Liberty*, has been proposed [88,89], but not yet fully incorporated into the MFT research program (thus, for simplicity, we omitted Liberty from the initial validation process). While we remain agnostic about MFT's claims regarding innateness or modularity [90–94], the theory nonetheless provides a broad, useful, and widely-used description of the way in which people's moral values differ, as well as of the situational factors that are likely to reveal those differences.

Challenge 3: Morality is complex. Another desirable feature of any stimulus set is that it adequately reflects the complexity of its subject matter. One pertinent limitation of many moral stimulus sets in this regard is that they are constructed on an assumption of discreteness, whereby each stimulus is assumed to represent just one moral construct (e.g., grouping stimuli into separate "harm" and "fairness" categories). As a particularly striking example, the Chadwick et al. stimulus set, which sorts 500 stimuli into 10 discrete categories, classifies the act of "Helping build a home for the needy" as charitable, but neither cooperative nor friendly, whereas "Helping someone find a lost dog" is classified as friendly, but neither charitable nor cooperative (see Table 2 in [27]).

Although assuming discreteness in the moral domain may occasionally be desirable, such an assumption may be problematic for two reasons in particular: (1) different kinds of moral content often covary, and (2) different kinds of moral content may interact when people form moral judgments. In both cases, existing research highlights how assuming discreteness may produce misleading findings.

Regarding the first problem, covariation of moral content, stimuli that fit into discrete moral categories (e.g., impure but harmless) may be the exception rather than the rule [24,49,95]. Instead, stimuli that are judged as violations of one moral norm (e.g., Care) are likely to also be judged as violations of other moral norms (e.g., Purity) [24,26,96] (although see [97,98]). To appreciate the implications of this, it is instructive to consider analogous

findings in the emotion literature. It is widely accepted–even by basic emotion theorists [99,100]–that people can, and regularly do, experience mixed emotions [101–103]. Likewise, ratings of basic emotion content (e.g., anger, disgust) are often highly correlated in normative ratings for pictorial [104–107], auditory [108] and word-based [109] affective stimulus sets. If researchers fail to account for mixed emotional experiences, they risk mistakenly attributing findings to one emotion (e.g., disgust) which may in fact be driven by another emotion that reliably co-occurs with the target emotion (e.g., anger; for discussion, see [110,111]). Alternatively, if researchers successfully narrow their focus to situations in which *just one* emotion is elicited in isolation, their findings may only be informative about a narrow subset of emotional experience (for reasons discussed next). The point here is not that different emotions or different kinds of moral content are hopelessly confounded, but rather that content overlap is a challenge that must be tackled head-on to gain a deeper understanding of moral phenomena (e.g., [92,95,97,111]).

Regarding the second problem, many recent studies underscore the importance of considering interactions between different kinds of moral content. For example, evaluations of harmful actions can differ markedly, depending on the perceived justness of the action or the relational context in which it takes place: many people judge self-defence differently than they do unprovoked assault, just as they judge a teacher hitting a student differently than the converse, even if each example entails similar amounts of suffering [112–114]. Additionally, people's *relative* concern for specific values (e.g., fairness as opposed to loyalty) influence moral judgments (e.g., of whistleblowing [115]). In person perception, people's valuation of a particular virtue (e.g., dedication) depends on the target's standing on other virtues (e.g., kindness [78]). Once again, one can look to the emotion literature where responses to emotion-eliciting stimuli are contingent on *combinations* of emotional responses (for example when experiencing sympathy either in isolation or in conjunction with other emotions [116]).

Assuming discreteness of moral content (e.g., that a purported harm violation contains only harm-related content) potentially misrepresents the structure of the moral domain or, at the very least, restricts research to a potentially narrow subset of it. In either case, the discreteness assumption discourages exploration of interactions between kinds of moral content (e.g., [78]), and likely overlooks real-world cues that are frequently relied upon for moral judgment, resulting in misleading conclusions and hampering efforts to decompose the moral domain into its constituent parts (for more general discussion, see [117,118]).

Constructing stimulus sets that overcome this challenge necessitates collecting a broad range of normative data without artificially constraining stimuli to be purportedly discrete instances of different moral content domains (e.g., by having stimuli only rated on the content domain to which they are assumed to belong [27], or by systematically excluding stimuli that map onto multiple content domains [26]). Equally importantly, the complex, multidimensional nature of moral content necessitates the construction of a *large* stimulus set that can accommodate both *systematic* designs that can be used to avoid confounds (to the extent possible [52,119]), and more *representative* designs that broadly sample the moral domain in a way that resembles everyday moral experiences in terms of the prominence and co-occurrence of different moral factors [117].

Challenge 4: There is more to morality than immorality. A final challenge that limits most existing moral stimulus sets is that they often reflect the broader trend in moral psychology to segregate moral goods and moral bads into separate research programs. While some theories posit processes or content domains that describe one or the other end of the moral spectrum [113,120,121], others posit processes or content domains that operate across the entire spectrum (e.g., [85,122]). However, the data supporting any one of these theories are typically limited to one pole (often the negative pole). Moreover, when both ends of the spectrum are covered within the same theoretical framework, this is often explored in separate

studies with different methods, posing challenges for the integration of findings (one notable exception being [123]). Perhaps the primary reason for this is the scarcity of stimulus sets covering both ends of the moral spectrum.

This is unfortunate for two reasons. First, although various theories of moral psychology employ concepts with substantial overlap (e.g., Humanity [120], Harm/Care [85], and Not harming [122]), the dearth of stimulus sets covering both ends of the moral spectrum presents major obstacles to integrating across theories whose primary concerns and/or evidence bases lie at opposite ends of the spectrum. Second, for theories seeking to describe both ends of the moral spectrum (e.g., [85,122]), the lack of studies *simultaneously* covering both ends hinders efforts to assess whether these theories achieve this descriptive goal. This is particularly important given widespread evidence pointing to differential processing of positive and negative content both in general [124–126], and in the moral domain in particular [126–128]. This suggests that concepts and processes used to describe one end of the moral spectrum may not translate so well to the other. The obvious solution to this challenge is to develop a stimulus set covering the entire moral spectrum.

Summary and overview of the present studies

The primary aim of the current studies is to develop a large, versatile stimulus set that addresses the critical challenges described above. Specifically, we (1) offer a departure from standard text-based moral stimuli by assembling a large image database containing rich, concrete, realistic stimuli suitable for a wide range of paradigms, (2) provide a diverse stimulus set sampling as much of the moral domain as possible, (3) avoid assuming (or requiring) that each stimulus can be placed in a discrete category (e.g., "fairness violation") by collecting normative data across the five moral foundations, and (4) provide stimuli that span the entire moral spectrum, from negative to positive, to facilitate comparison and theoretical integration across theories that have thus far been restricted to one end of the moral spectrum.

To do so, we chose an approach rarely used in stimulus set construction: we build our stimulus set from the bottom up, through crowdsourcing, to mitigate potential researcher-selection biases (although the crowdsourced images are supplemented with researcher-selected content). Additionally, we restricted the database to include only images that are Creative Commons licensed (or have similarly permissive licenses), which has the benefits of (1) allowing researchers the freedom to present the materials in both online and offline settings without concern about copyright restrictions (an issue receiving increasing attention with the development of new research materials across various fields [4,14,129,130]), and (2) enabling a range of novel research applications that leverage the wealth of text (and other) data linked to many of the images, as discussed towards the end of the paper.

The remainder of the paper is structured as follows. In Study 1, we report the generation of the image set, covering the sourcing and screening of images. In Study 2, we report the norming of the image set, and provide a detailed description of important features of the image set that we expect to be of interest to moral psychology researchers.

Study 1

To avoid biasing the content of different moral dimensions, we opted for a bottom-up approach for image collection, crowdsourcing most images rather than collecting them ourselves. To this end, Study 1 reports the image collection and screening procedure.

Method and results

All participants for Study 1 were recruited via Amazon's Mechanical Turk (AMT), an online crowdsourcing platform where people perform tasks referred to as Human Intelligence Tasks

(HITs) in exchange for monetary compensation [131–133]. We restricted eligibility to participants with approval ratings ≥ 95% and ≥ 1,000 previous tasks completed.

Ethics statement. This study received ethical clearance from the University of Melbourne (ethics ID 1341310). Prospective participants were directed from AMT to an online information sheet describing the study, after which participants provided informed consent if they wished to participate.

Image collection. To generate a pool of candidate images, we recruited 476 participants into an image collection task. Exactly half of the participants were female, and a substantial majority (around 90%) reported having completed or commenced at least some university or college education. Participant ages ranged from 18 to 67 (M = 33.26, SD = 9.85). No other demographic information was recorded. To increase participant diversity, around 10% of participants were recruited from India, with the rest from the United States.

Each participant was asked to search for and provide URLs for 20 images, that were available from the Wikimedia Commons or Flickr, and that they believed were representative of two randomly assigned moral concepts (i.e., 10 images per concept). Concepts were both positively and negatively valenced, and spanned a wide range of moral content. While some of the concepts were directly related to those used in moral foundations research, we also included a number of non-MFT concepts (e.g., "Deception," "Self-Control" and generic morality / immorality concepts). To increase the diversity of the images, we adopted different strategies across iterations of the task, such as altering the concreteness of the moral concepts (e.g., "People behaving immorally" vs. "Unfairness" vs. "Theft"), and having participants generate their own search prompts related to each moral concept (a full list of moral concepts is provided in Table 2, and full search instructions provided in S2 Text).

After excluding duplicate URLs, corrupted or irretrievable images, and images that were smaller than 640 by 480 pixels, this process yielded 4,092 images. An additional 362 researcher-contributed images were added to the pool after reaching a saturation point where later participants frequently returned images that had already been submitted by previous participants. This increased the total to 4,454 images.

Image screening. Next, we programmatically collected metadata (including licensing information, image author, title etc.) for all retrievable images using the Wikimedia and Flickr

Table 2. Image search concepts used in image set generation.

Animals behaving immorally	Hedonism	Reminders that humans are animals too
Animals behaving morally	Hierarchical Relationships	Respect / Subversion
Benefitting one's own group at the expense of others	Hierarchy in the animal kingdom	Rudeness
Care / Harm	Inequality	Sacrifice for the good of one's group
Compassion in the animal kingdom	Injustice	Sanctity / Degradation
Conformity	Lack of Self Control	Self-Control
Contamination	Liberty / Oppression	Selfishness
Cooperation / group behavior in the animal kingdom	Loyalty / Betrayal	Social Hierarchy
Cruelty	Maintaining boundaries between groups	Teamwork
Cruelty in the animal kingdom	Naturalness	Theft
Deception	Obedience	Tradition
Discrimination	People behaving immorally	Treachery
Equality	People behaving morally	Un-naturalness
Exploitation	Politeness	Unfairness
Fairness in the animal kingdom	Proportionality (i.e. getting what one deserves)	Untrustworthiness
Fairness/Cheating	Proportionality (i.e. getting what one has earned)	Wholesomeness
Generosity	Reciprocity (i.e. paying back in kind)	

application programming interfaces (APIs). To ensure the final image set could be used as widely as possible, we retained only Creative Commons (or similarly permissively) licensed images. Filtering out images with more restrictive licensing left a pool of 3,726 images.

An independent AMT sample of 285 AMT workers (48% male; age $M = 35.72$, $SD = 10.75$) screened the remaining images for various features so that we could select a subset of images most useful for a wide range of research applications. Images were excluded if they (1) contained famous people, (2) prominent text (such that extracting the meaning of the image required reading), (3) watermarks or commercial logos, or (4) were non-photographic (e.g., cartoons). Each participant screened around 60–70 images, and each image was screened by at least five participants. Features were coded as present if a majority rated them as such. Through this process, we retained 2,941 eligible images to be subjected to further rating. A summary of this process is presented in Fig 1. Additionally, participants coded images for the presence of people (appearing in 63% of images in the final pool), animals (17%), and landscapes (15%) so that such features can be incorporated into stimulus selection procedures (coding of additional features is underway).

Study 2

Having generated a pool of images in Study 1, Study 2 involved collecting normative ratings for the images on a set of moral and affective dimensions.

Method

Ethics statement. This study received ethical clearance from the University of Melbourne (ethics ID 1341310). As in Study 1, prospective participants were directed to an online information sheet describing the study procedure, after which they provided informed consent if they wished to participate.

Participants. We recruited a large sample from two sources: AMT (as in Study 1), and the University of Melbourne undergraduate psychology research participation pool. For AMT participants, eligibility was restricted to workers located in the United States with approval rates $\geq$ 90%, and $\geq$ 100 previously approved HITs. After excluding participants using an extensive set of criteria to detect inattentiveness (detailed in S3 Text), our final sample comprised 1,812 AMT participants (49% male; $M_{age} = 36.75$, $SD_{age} = 11.20$), and 904 undergraduate participants (24% male; $M_{age} = 19.31$, $SD_{age} = 3.48$) who, combined, provided a total of 820,565 ratings.

Sample size was determined based on a target of obtaining at least twenty ratings for each image on each dimension, although the average number of ratings was considerably higher ($M = 34.88$). Such a number of ratings per image is comparable to existing affective image sets, especially considering the comparatively larger number of images and dimensions that were rated (see S1 Text for comparisons of the SMID with existing affective image sets regarding rating frequencies). To ensure that ratings were not skewed by a lack of moral or political diversity, we ensured that each image was rated on each dimension by a minimum of five
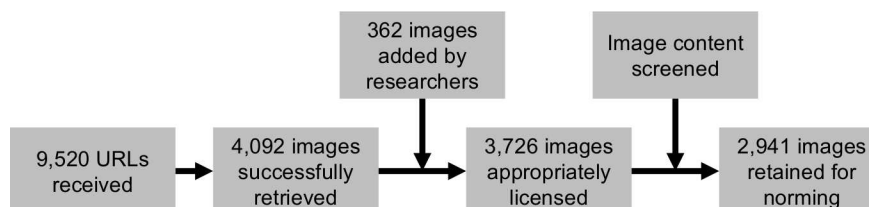

Fig 1. Summary of database construction process.

AMT participants each self-identifying as liberal, conservative, or moderate/other (and at least five Australian undergraduate participants).

Materials and procedure. All 2,941 eligible images were rescaled to a height of 400 pixels (maintaining their original aspect ratios), and then randomly split into 99 largely non-overlapping batches of 30–40 images. Two images (one of a thunderstorm, and one of the scales of justice) were deliberately included in all batches to provide a common context and reduce the likelihood of some batches (by chance) including highly idiosyncratic moral content. Additionally, we discovered a small proportion of images that appeared in multiple batches because, during Study 1, participants occasionally submitted different URLs indexing the same image. Because these images were identified *after* commencing the study, ratings for these duplicate images were combined *post hoc*.

In later stages of data collection (i.e., for most of the Australian undergraduate sample), we implemented a strategy to obtain additional ratings for images eliciting highly variable responses. Specifically, we constructed a pool of images whose normative ratings had the largest standard errors. For participants rating batches containing < 40 images, additional images were drawn from this pool until that participant had been assigned a total of 40 images.

After providing informed consent, participants were randomly assigned to rate one batch of images via their web browser in a custom-coded JavaScript task (N = 23–34 participants per batch). Images were rated on each of the following eight dimensions (each on a 1–5 scale, using the keyboard): valence ("unpleasant or negative" to "pleasant or positive"), arousal ("calming" to "exciting"), morality ("immoral/blameworthy" to "moral/praiseworthy"), and the five moral foundations, Care, Fairness, Ingroup, Authority and Purity. When rating images with respect to moral foundations, participants rated the extent to which the images made them think about that specific foundation ("not at all" to "very much"). Rating dimension labels are summarized in Table 3. Before rating images on a dimension, participants read a detailed description of that dimension (provided in full in S4 Text). Participants rated all images in the assigned batch on one dimension before proceeding to the next dimension, until all dimensions had been rated. Image and dimension order were randomized within participants to prevent order effects.

After completing the rating task, participants were redirected to a questionnaire in which they provided basic demographic information (including political orientation), and completed the 30-item Moral Foundations Questionnaire (MFQ [134]). Analyses of these self-report data are to be reported elsewhere.

Results and discussion

Here, we present a multifaceted assessment of data quality, followed by a high-level summary of image-level variability within and across dimensions. (Note that we defer discussion of general recommendations for use until the General Discussion.)

Table 3. Image rating dimension summary.

Dimension	Label	Lower Anchor	Upper Anchor
Valence	This image is . . .	UNPLEASANT or NEGATIVE	PLEASANT or POSITIVE
Arousal	This image is . . .	CALMING	EXCITING
Morality	This image portrays something . . .	IMMORAL / BLAMEWORTHY	MORAL / PRAISEWORTHY
Harm	This image makes me think about the concept of CARE / HARM	NOT AT ALL	VERY MUCH
Fairness	This image makes me think about the concept of FAIRNESS / CHEATING	NOT AT ALL	VERY MUCH
Ingroup	This image makes me think about the concept of LOYALTY / BETRAYAL	NOT AT ALL	VERY MUCH
Authority	This image makes me think about the concept of RESPECT / SUBVERSION	NOT AT ALL	VERY MUCH
Purity	This image makes me think about the concept of SANCTITY / DEGRADATION	NOT AT ALL	VERY MUCH

Inter-rater consensus

First, we sought to quantify the degree of consensus in the ratings for each dimension. One important motivation for such analyses is that it is unclear what degree of consistency to expect when eliciting single-item ratings of broad, abstract moral content dimensions (given the lack of previous research addressing the question). To this end, we computed two variants of the intra-class correlation coefficient (ICC) separately for each dimension in each batch of images. ICCs are commonly interpreted as reflecting the proportion of variance in ratings attributable to the target (here, image stimuli) [135–137]. As such, higher values indicate greater consensus arising from such factors as common (1) interpretations of the rating scale, (2) interpretations of the images, or (3) scale use.

Using the *irr* package for *R* [138], we first computed ICC(A,1) for each batch, where (1) target (image) and rater (participant) were both treated as random effects (see [136]), (2) coefficients were calculated based on absolute agreement (rather than consistency), and (3) the coefficient reflects the reliability of a *single* rating. The distribution of ICCs across batches for each dimension, across the entire sample, is presented in Fig 2. Note that ICCs calculated based on absolute agreement (i.e., ICC(A,1)) will tend to be lower than ICCs calculated based on consistency (i.e., ICC(C,1)), as indeed was the case here: across all dimensions and batches, the average ICC(A,1) was about .04 less than the average ICC(C,1).

As shown in Fig 2, there was substantial variability in consensus across rating dimensions, with valence (and to a lesser extent, morality) eliciting the greatest amount of agreement, and the five moral foundations eliciting relatively less. Importantly, moral foundation content had ICCs comparable to that of arousal. Amongst the five moral foundations, Care was the most agreed upon dimension, and Fairness the least.

Compared to frequently cited rules-of-thumb [139], these reliabilities range from "fair" ($.40 \leq \text{ICC} < .60$, for valence and morality) to "poor" ($\text{ICC} < .40$, for all other dimensions). However, it should be noted that these guidelines were intended for the evaluation of clinical assessment instruments (which often comprise multiple items). Moreover, to our knowledge, ICCs of any kind are neither reported for validation studies of existing affective image sets nor
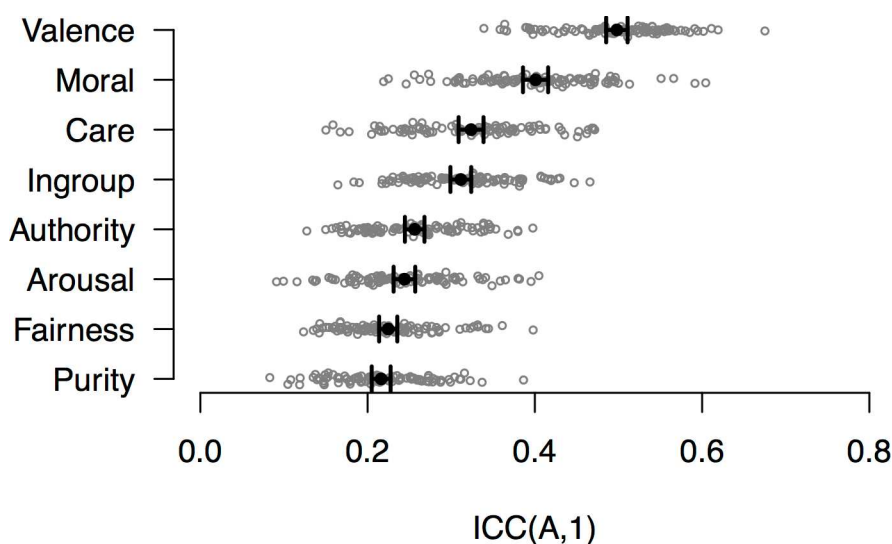

Fig 2. Distributions of intra-class correlation coefficients for each rating dimension across image batches. Grey points represent individual observations (i.e., each of the 99 batches). Black points represent the average ICC(A,1) across all batches, with error bars representing 95% CIs.

https://doi.org/10.1371/journal.pone.0190954.g002

textual moral stimulus sets, making it difficult to provide a sufficiently similar reference point for comparison. Finally, we note that ICC(A,1) is insensitive to the number of ratings obtained per image, and thus does not reflect the reliability of the norms, but rather of a single rating (however, the second variant of the ICC reported below does take rating frequency into account). Nonetheless, the fact that a large proportion of the variance in image ratings is explained by sources *other than* the image itself suggests, perhaps unsurprisingly, that factors such as people's idiosyncratic interpretations of moral concepts (and the stimuli themselves) exert substantial influence on ratings (we return to this point below).

Precision of measurement.   Next, we examined the degree of *precision* in the image norms which, unlike the analyses above, is not just a function of inter-rater consensus, but also of the amount of data collected (given that noisy but unbiased measures will give accurate estimates with enough observations). To measure precision, for each dimension, we used two metrics. First, we computed the expected width of the 95% confidence interval for an "average" image as a function of (1) rating frequency, and (2) the average of the standard deviations (SD) of ratings for all 2,941 images on that dimension. Expected CI widths for each dimension at various rating frequencies (measured in scale points for a five-point scale), are shown in Fig 3, along with observed CI widths for each image. Fig 3 thus shows (1) the accuracy of norms across the image set, as well as (2) the expected gain in precision if more data were to be collected. Additionally, we computed ICC(A,k) (also displayed in Fig 3), providing a measure of reliability of image norms created by averaging across raters.

As shown in Fig 3, we achieved a 95% CI width of less than one scale point (i.e., plus or minus half a scale point) for most images on most dimensions. More concretely, this means that if one wished to sample images that were typically perceived in a specific way (e.g., as highly immoral), the amount of data available allows researchers to do so with a reasonably high degree of confidence. Additionally, we note that averaged ratings on all eight dimensions achieved "excellent" reliability (ICC $\geq$ .75, based on the guidelines proposed in [139]). Although additional data would further enhance precision, such would only be achieved with diminishing returns for every additional participant: halving the average CI width (particularly for the moral content dimensions) would effectively require increasing the rating frequency by a factor of around four or five (requiring around 10,000 to 13,000 raters, given the current task parameters).

Moral content distributions.   Having explored the reliability and precision of the ratings, we next describe the distribution of moral content ratings, beginning with univariate and bivariate distributions of each moral content dimension (and pairs thereof), depicted below in Fig 4. Depending on researchers' goals or assumptions, an ideal image set might contain images spanning all possible values for each dimension (and combinations of dimensions), such that researchers could easily select images meeting arbitrary criteria (e.g., high on dimension A, low on dimension B, etc.). However, as has been repeatedly demonstrated for affective stimulus sets, this combinatorial goal is difficult to achieve in practice (e.g., finding negatively valenced, low-arousal stimuli [1,2,4,12]). As shown in Fig 4, a similar pattern obtains for moral content (and for valence and arousal).

Based on a strict "modular" view of moral foundations as discrete domains, one might *intuitively* expect relevance ratings for the five dimensions to be at most weakly or moderately correlated (even if a modular view does not strictly require this). Fig 4, however, shows that all five foundations were strongly positively correlated (all $rs > .5$, all $ps < .001$), suggesting that relatively "pure" instances of individual foundations (i.e., scoring highly on one, but low on all others) may be somewhat rare, as suggested in the Introduction and by previous research [26,141] (and mirroring findings in the basic emotion literature).

More broadly, we caution that these correlations ought not to be taken as refuting the existence of discrete moral "modules" for two reasons (although similar correlations have been
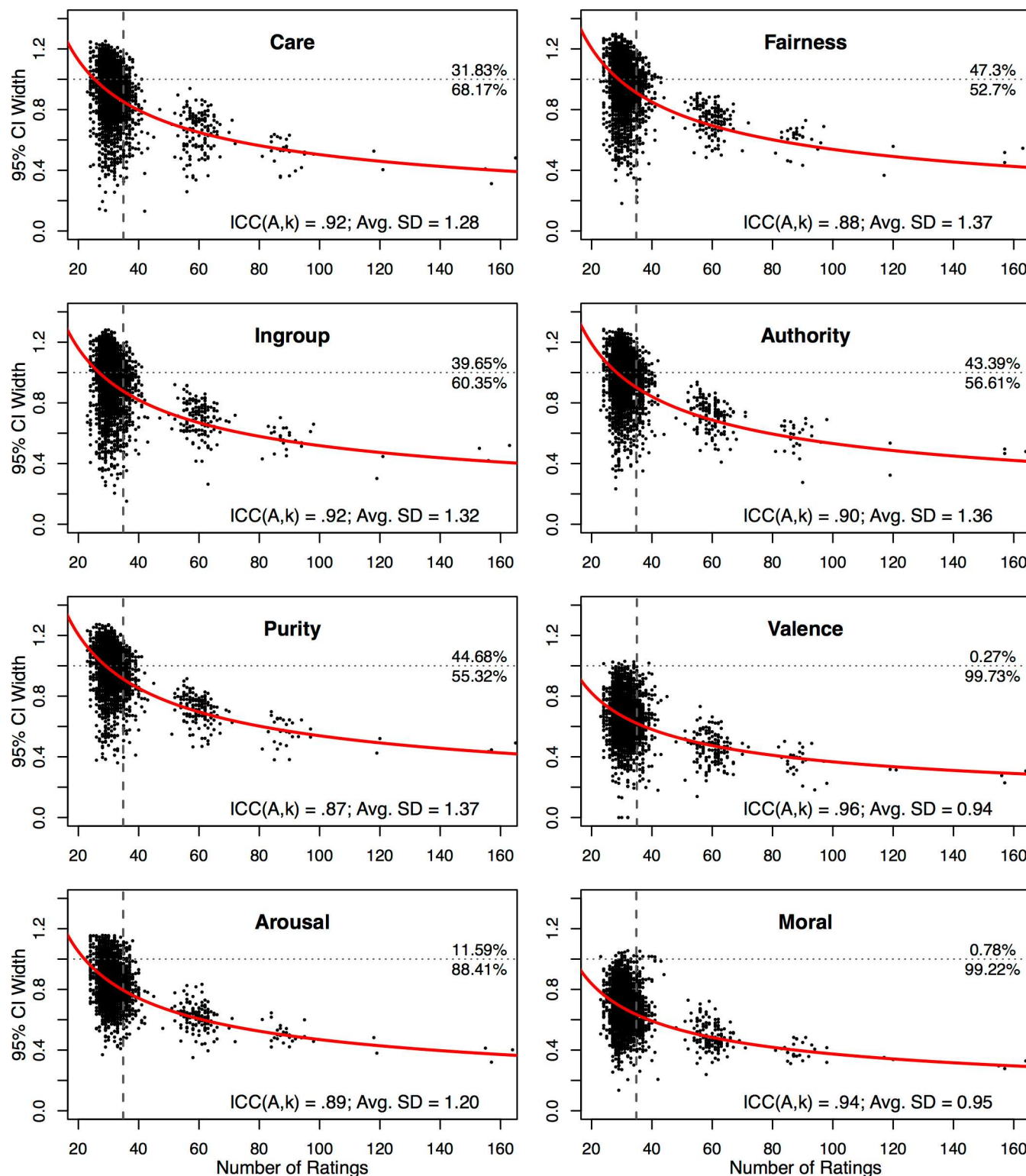
Fig 3. Scatterplots of image norm 95% confidence interval (CI) widths as a function of the number of ratings, by dimension. Each point represents an individual image. Vertical axis represents 95% CI width (in scale points) for each image, with images lower on the axis having more precise measurement. Horizontal axis represents the number of times each image has been rated. Red curve represents expected 95% CI width given the average rating SD (inset) for that dimension and number of ratings. Vertical dashed grey line represents average number of ratings per image for that dimension. Horizontal dotted grey line marks a 95% CI width of 1, with the percentage of images falling above or below this threshold presented at the right end of the line. ICC(A,k) (inset) represents the average ICC(A,k) across batches.
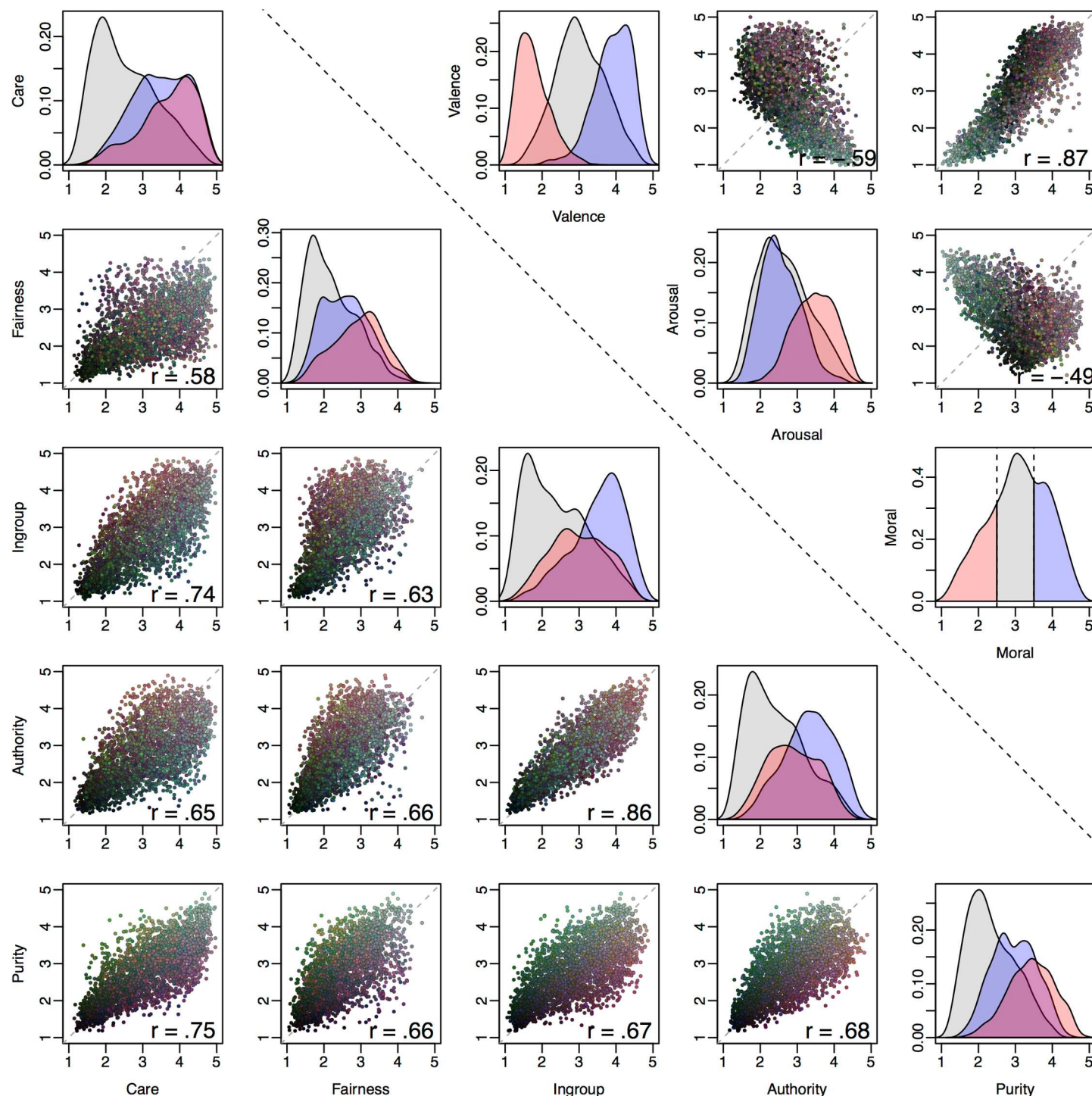
Fig 4. Correlations and rating distributions across images for (1) moral content dimensions (lower section below black dashed line), and (2) valence, arousal and morality (upper section above black dashed line). On diagonal: density plots of relevance ratings for each moral foundation (lower section), and valence, arousal and morality (upper section), with each plot divided into morally good (blue; mean moral rating > 3.5), bad (red; mean moral rating < 2.5), and neutral (grey; all other images). Off diagonal: scatterplots of average ratings for all images with Pearson correlation coefficients inset. To aid interpretation, point color represents moral content ratings collapsed into the CAD triad [140] with each of the three dimensions mapping onto a different color (Community [Ingroup + Authority] = red; Autonomy [Care + Fairness] = blue; Divinity [Purity] = green).

https://doi.org/10.1371/journal.pone.0190954.g004

interpreted as such elsewhere [24,142]). First, the fact that two variables are strongly correlated does not necessarily imply that they are the same thing (e.g., consider height and weight in humans which, in two large datasets available from http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data, exhibit correlations > .5). Second, the correlations reported in Fig 4 were observed at the group level (aggregated by image). It is entirely possible for analogous correlations within individuals to differ substantially [143–145]. For example, although the image-level correlation between Care and Fairness relevance was .58, when one computes the Care-Fairness correlation within each individual, the average correlation is .32, and the correlation is in fact negative for 11% of participants.

Can foundation-specific images be identified?. Although images that *exclusively* represented specific foundations were rare, it is possible to identify images that relate more strongly to one foundation than others. To identify such images, we devised a set of *uniqueness scores* for each image on each foundation (included in the normative data available at https://osf.io/2rqad/). Uniqueness scores were computed by taking an image's score on a given foundation, and subtracting from this value the maximum score the image received on the other four foundations (alternative methods, included in the norms, but omitted here for brevity, are described in [6,109]). For example, consider an image for which the average relevance to Care = 5. If that image's highest average score on the other foundations is Purity = 3, we assign a Care uniqueness score of $5 - 3 = 2$. A positive uniqueness score of x for a given foundation thus indicates that an image is on average rated at least x scale points higher on that foundation than all foundations. Uniqueness score distributions for each foundation are summarized in Fig 5.

As shown in Fig 5, uniqueness scores tended to cluster around or below zero (unsurprisingly, given that by definition, uniqueness scores for each image will be ≤ 0 on four of five dimensions). While the maximum possible uniqueness score was 4 (5–1), few images scored above 2 for any dimension. Importantly, however, the image set included at least 46 morally good and 21 morally bad images with positive uniqueness scores for each individual foundation (i.e., $\geq$ 46 morally good Care images, $\geq$ 46 morally good Fairness images, and so-on), indicating the presence of images predominantly (if not exclusively) depicting each foundation. Moreover, when one visually inspects the images with high uniqueness scores there is a high degree of face validity for the images representing each moral foundation.
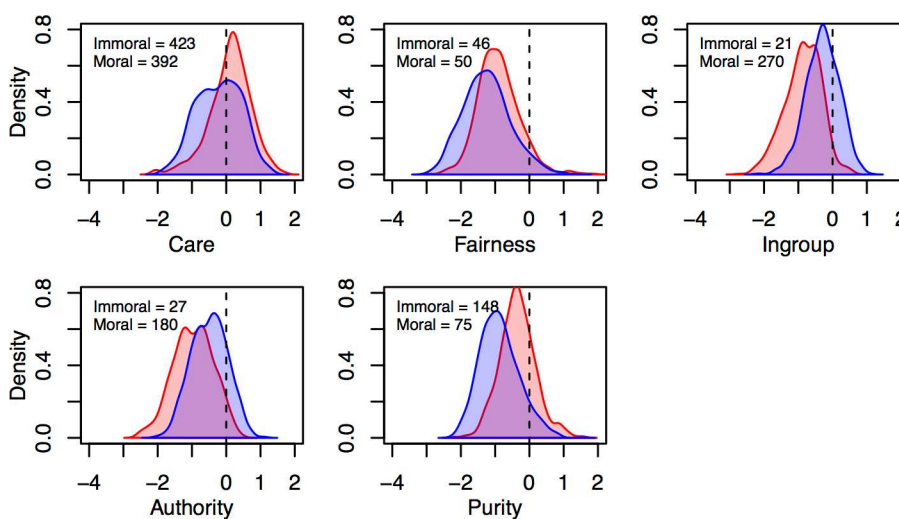

Fig 5. Density plots of image uniqueness score distributions for each moral foundation by moral valence. Morally good images (blue) are defined as having mean moral ratings > 3.5, and morally bad images (red) as having mean moral ratings < 2.5. Number of morally good and bad images per foundation with uniqueness scores > 0 inset.

https://doi.org/10.1371/journal.pone.0190954.g005

Mapping moral content onto valence, arousal and moral judgments. Two broad questions that have motivated much research in moral psychology concern (1) the relative importance of different moral content domains for explaining moral judgments [49,121,141,146] and (2) links between moral cognition and the core affective dimensions of valence and arousal [45,147–152]. Here, we describe image-level correlations between moral content ratings on the one hand (i.e., relevance to each of the five moral foundations), and moral judgments, valence and arousal on the other. It should be noted however that, as image-level correlations, one should not assume that analogous correlations hold within (or between) individuals. Rather, these correlations reflect the content of images as they tend to be perceived by groups (which may nonetheless serve as plausible *hypotheses* regarding within- or between-person correlations). These correlations are presented in Fig 6.

As can be seen for judgments of morality (Fig 6, right column), the bivariate distributions resembled a clear v-shaped relationship, such that images receiving extreme moral judgments (either positive or negative) were rarely rated as irrelevant to any of the moral foundations. This pattern was most pronounced for Care and Purity.

Recall that the moral content variables were coded on a non-valenced scale, ranging from irrelevant to highly relevant to a specific content domain (e.g., with Harm/Care *both* anchoring the upper-most response), rather than a valenced scale (with Harm and Care on opposite poles). Thus, the v-shaped pattern emerged as a predictable consequence of images with extreme moral content on a specific dimension tending to elicit positive or negative moral judgments depending on whether the image portrayed the positive or negative pole of that content domain.

The pattern of findings was similar for valence (Fig 6, left column), though with a less prominent (and less symmetric) but still noticeable v-shaped pattern for Care and Purity. The somewhat asymmetric pattern suggests that whereas negatively valenced images tended to be loaded with Care and/or Purity content, this was less often the case for positively valenced images (at least for the images included in the database).

No such v-shaped relationship was apparent for arousal. Instead, all five content dimensions were positively correlated with arousal (especially Care, Fairness and Purity), suggesting that low-arousal images were relatively devoid of moral content, whereas highly arousing images were more likely to be rated as containing various kinds of moral content.

Exploring within-dimension variability. Regardless of one's research goals, an important (but neglected) consideration in stimulus selection concerns stimulus-level variance (e.g., whether an image tends to elicit uniform responses or strong disagreement with regards to some feature). Users of the SMID may benefit from looking beyond simple averaged ratings by purposefully selecting images with levels of variability that match one's aims. As shown in Fig 7, using moral ratings as an example, images elicited a wide range of variability in their ratings, with some images eliciting nearly uniform judgments (with SDs at or around 0), and others eliciting substantial variability (SDs of around 1.5 scale points). Moreover, this was the case for images across the moral spectrum.

Rather than simply reflecting random noise in the ratings, we suggest that rating variability can be meaningfully accounted for by at least three separate substantive sources (all of which would be expected to produce higher SDs, and could in principle be clearly separated with additional measurements).

The first and most straightforward source is ambiguity, whereby images invite multiple interpretations (e.g., an image that could plausibly be construed as either play-fighting or assault), or may simply be difficult for viewers to interpret (e.g., because of high visual complexity). Rather than being an altogether undesirable quality, ambiguous stimuli may be ideally suited for many kinds of paradigms (e.g., [48,153–157]).
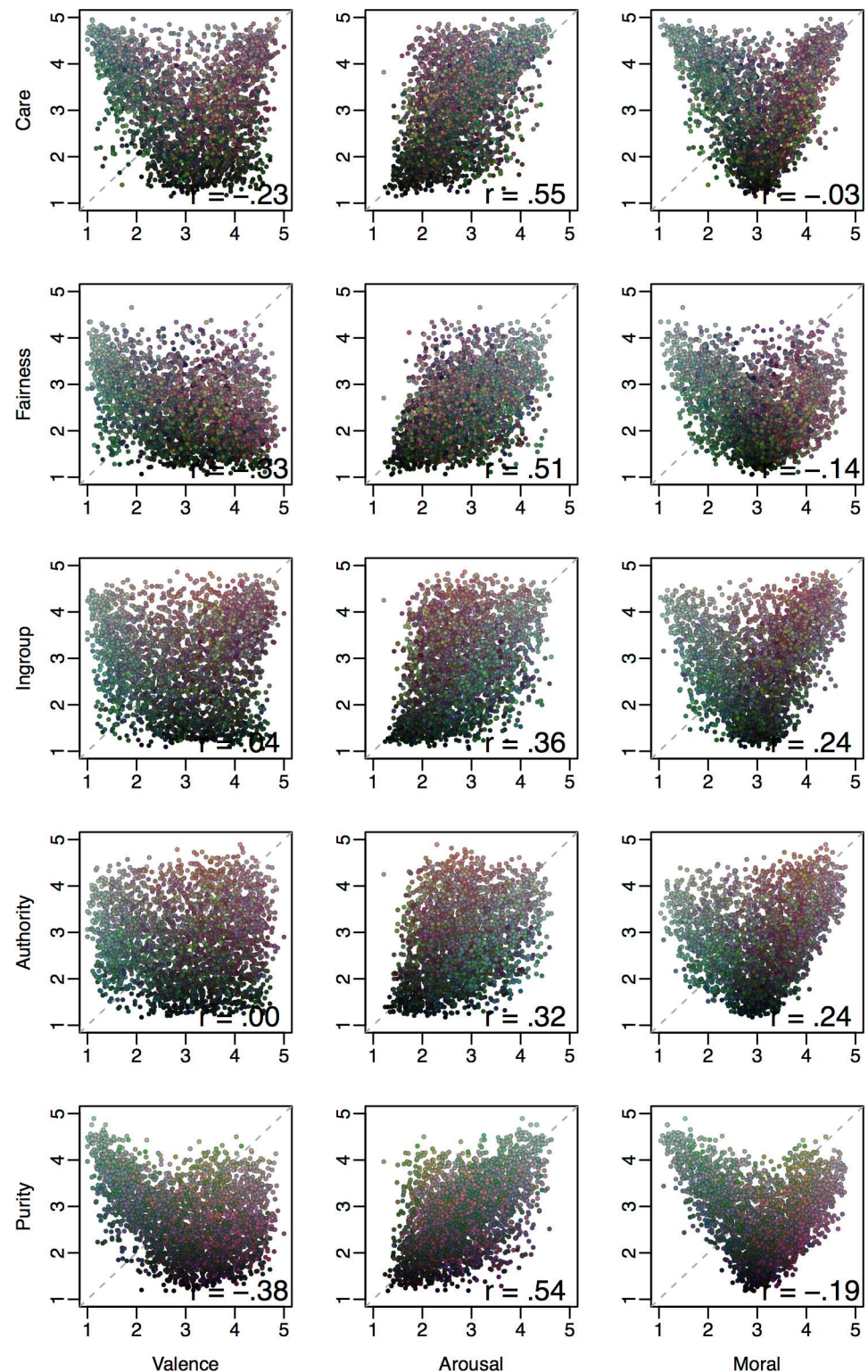
Fig 6. Correlations between moral content dimensions and valence, arousal and morality ratings across images.
Each cell contains a scatterplot of average ratings for all images for all moral foundations with valence, arousal and
morality; To aid interpretation, point color represents moral content ratings collapsed into the CAD triad [140] with
each of the three dimensions mapping onto a different color (Community [Ingroup + Authority] = red; Autonomy
[Care + Fairness] = blue; Divinity [Purity] = green).

https://doi.org/10.1371/journal.pone.0190954.g006

Fig 7. Scatterplot of moral rating standard deviation against moral rating mean for each image. To aid interpretation, point color represents moral content ratings collapsed into the CAD triad [140] with each of the three dimensions mapping onto a different color (Community [Ingroup + Authority] = red; Autonomy [Care + Fairness] = blue; Divinity [Purity] = green).
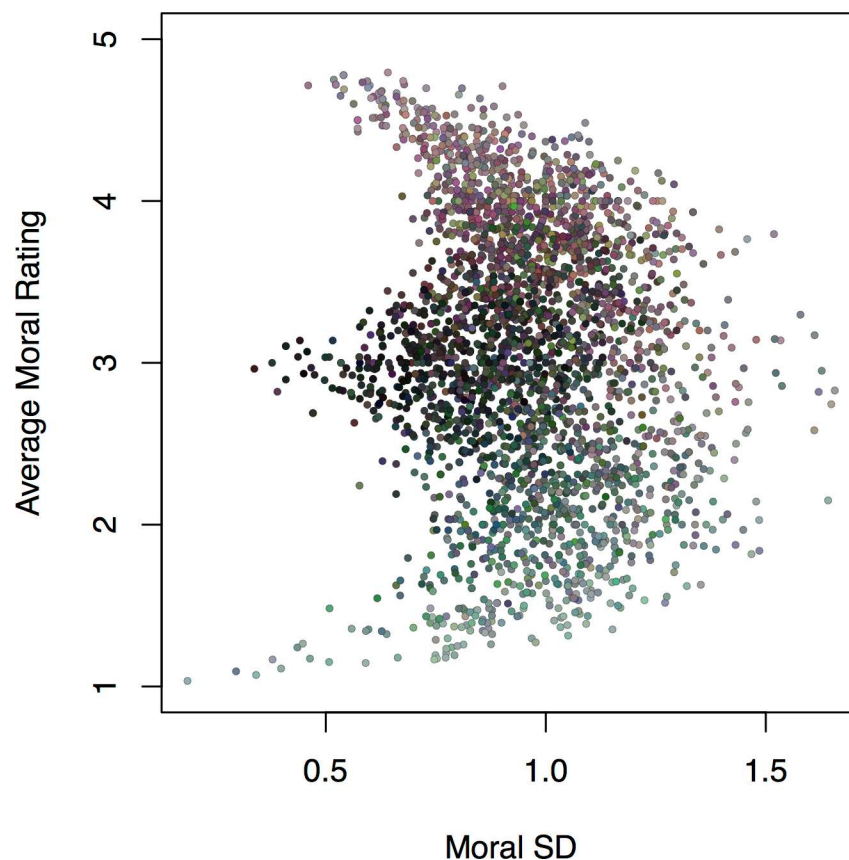
The second source of variability, reflecting intrapersonal processes, is ambivalence. People can simultaneously hold positive and negative evaluations of both moral [158] and non-moral [159–161] stimuli, which would result in greater variability in judgments (see [159]). Again, rather than reflecting an altogether undesirable feature of the image set, ambivalence-inducing images may prove useful for (among other applications) probing the integration of conflicting moral information [158]. Although the data presented here cannot speak to the presence of ambivalence, such can easily be measured by adapting measures such as the evaluative space grid [162].

Finally, we consider a third source of variability operating at the interpersonal level: divisiveness. Given differing moral concerns, people may simply disagree with each other regarding their moral evaluation of an image, absent any disagreement about what is portrayed in it (i.e., absent any ambiguity). Once again, divisiveness may prove highly useful for specific research goals such as eliciting psychological or physiological responses that are diagnostic of one's political or moral preferences (e.g., [41,72,163]), or developing pictorial measures of individual differences [41,65,66,130,156,164–166]. In S5 Text, we present preliminary analyses that attempt to empirically identify images that are divisive with respect to political orientation and gender.

General discussion

Methods and materials are a driving force behind scientific progress [167]. The primary aim of this project was to expand the range of tools available to moral psychology researchers by developing and validating a novel image database. In this final section, we briefly discuss (1) important gaps in the literature that the SMID can address, (2) potential applications of the image set, accompanied by some general guidelines for use, (3) potential extensions and finally, (4) limitations of the database.

Improvements over existing stimulus sets

One of the more obvious advantageous features of the SMID is its size and scope: it is the largest freely available moral stimulus database assembled to date, and one of the few that covers both the morally good and bad poles of a range of content dimensions. Moreover, unlike existing stimulus sets, the database is not limited to the portrayal of moral actions, but also contains images of objects and symbols that can also be the target of moral evaluations, and are worthy topics of study in themselves [168–170].

The size of the SMID gives rise to two particularly important benefits. First, in populations in which non-naïveté may be a concern, such as AMT [171,172], the size of the stimulus set reduces the likelihood of participants repeatedly encountering the same stimuli over multiple studies by different labs. Furthermore, in light of salient concerns within the field of psychology regarding reproducibility [173–175], the size of the image set enables researchers to address the underemphasized issues of (1) stimulus (re)sampling in replication efforts [176–178], and (2) sampling a sufficiently large number of stimuli to achieve sufficient statistical power to account for stimulus sampling variance (relevant in psychology as a whole [36], but especially pertinent in resource-intensive neuroimaging research [38,179]).

Beyond its size and scope, there are many ways in which the SMID is qualitatively different to currently available stimulus sets. Among the most prominent of these differences is the SMID's reliance on images (rather than text) which allows researchers to run studies that would be impractical with text-based stimuli (e.g., [163]). Perhaps the greatest benefit of the SMID over existing stimulus sets, however, is that of greater ecological validity. While one of the most commonly used tool in moral psychology (sacrificial dilemmas) has been criticised for a severe lack of ecological validity [180], the SMID contains detailed depictions of real-world actions, objects, scenes and situations.

The SMID is also among the only databases in which mixed moral content is explicitly modelled. We have departed from the common practice of constraining each stimulus to map onto one and only one content domain [26,27], instead favouring an approach that embraces the complexity of moral phenomena. Such an approach minimizes the impact of theoretical assumptions on the composition of the stimulus set, making it ideally suited to analytic approaches designed to accommodate complex, multidimensional stimuli [181–183], and allowing researchers to explore relatively neglected research topics concerning, for example, interactions between moral content domains [78,112].

Finally, although the SMID is first and foremost an image set, it is also unique in that it spans multiple mediums, with linked text data available for a substantial proportion of the images in the form of Wikipedia pages and Flickr tags or comments (and in many cases, other web pages that also use the same images). Thus, one intriguing application of the image set entails leveraging the enormous quantities of text and metadata in webpages containing these images using the many available methods (e.g., [184–190]). Much as previous research has attempted such feats as using linguistic data to estimate people's values [191,192], documents' value content [21,193–197], or the affective connotations of words [198,199], so too could

researchers attempt to estimate the moral content of images based on linked text data (for an impressive demonstration of this approach, in combination with computer vision techniques and applied to emotion recognition, see [200]).

Applications and recommendations for use

In addition to the recommendations provided throughout the paper, here we offer some additional guidelines for researchers intending to use the SMID in experimental research.

Selecting optimal subsets. Selecting an optimal set of stimuli can be a highly complex challenge [52,201,202]. In particular, sampling an optimal set from a larger pool becomes increasingly labor-intensive as either the size of the pool, the size of the sample, or the number of variables to be controlled, increases [52]. Thus, if one considers (1) the size of the SMID, (2) the number of stimuli required to run a well-powered study (see [36]), (3) the large number of variables for which normative ratings are available in the database, and (4) debates within moral psychology regarding the effects of various confounds on existing findings (e.g., [24,203–207]), manually selecting subsets of stimuli will produce far-from-optimal solutions for all but the simplest stimulus selection problems. Thus, systematic approaches to stimulus selection are of great importance. Fortunately systematic methods and software packages for stimulus selection are readily available [201,208–210]. To facilitate the adoption of systematic approaches to stimulus selection within the SMID, we have written a generic stimulus selection script for use with the SOS toolbox for MATLAB [201], along with a generic image rating task script programmed using the Python library PsychoPy [211]. Both are scripts available at https://osf.io/2rqad/, and can be easily modified to accommodate researchers' own research needs (for users without access to MATLAB, a standalone executable version of SOS is also available).

Similarly, in many paradigms, low-level visual features (e.g., luminance or contrast) or high-level visual features (e.g., the presence or absence of human faces) may produce unwanted confounds that undermine the validity of findings (e.g., [12,212–214]). Indeed, as shown in S1 Table, we observe weak though nonetheless significant associations between some visual features and content dimensions. Consideration of such factors is especially pertinent in neuroscientific investigations of moral cognition where researchers generally wish to avoid inducing differences in brain activity with confounded, non-moral features. Fortunately, such features can often be quantified and incorporated into the stimulus selection procedures described above. When control via selection proves difficult, low-level features can be manipulated using readily available software [215]. Additionally, manipulations of high-level image features (e.g., object transfiguration) are becoming increasingly feasible [216,217], presenting fascinating directions for future experimental research. However, for researchers manipulating any aspect of images to achieve statistical control, we urge caution given that modifying seemingly irrelevant perceptual features may influence affective [218,219] and moral processes [219–225].

Maximising reproducibility. The primary motivation for developing the SMID was to facilitate rigorous, efficient, and cumulative moral psychology research. Here, we briefly discuss how the database can best serve these goals. To meet the minimum standards for reproducibility, we first recommend that researchers list the unique identifiers of all images selected for their studies. Second, where stimuli are programmatically sampled, sharing code used to sample images will enable replications with different stimuli selected under exactly the same sampling regime [176]. Finally, an emphasis on data sharing offers perhaps the most productive step researchers can take when using the SMID. Given that each image has been normed on multiple dimensions (and that this set will continue to expand), data generated using the SMID has great potential for reuse beyond the original question(s) motivating a given study.

Given the potential to aggregate person-, stimulus- and trial-level data across studies, the benefits of data sharing for the SMID are arguably even greater than for the typical moral psychology study (especially for costly and often underpowered neuroimaging studies [38,179,226]).

Uses outside of moral psychology. Although the database is primarily intended for use in moral psychology, it is also worth highlighting its potential usefulness further afield. Although not primarily intended as an *affective* image set, the SMID could be used as such, given that (1) all images are normed on valence and arousal, (2) the database is more than twice the size of the largest available affective image database for which such norms are available (the NAPS [1]), and (3) given its diverse content, the database may be less vulnerable to confounds (compared for example to the GAPED [2]; see S1 Text for discussion). The SMID could therefore serve as a valuable resource for psychological and physiological investigations of emotion, and further afield, as a benchmarking or validation dataset in affective computing studies [200] (and possibly also in the fields of social computing and machine ethics [227–231]).

Extensions

Given the large quantity of image data available on the internet, and exponential growth in Creative Commons licensed material [14,232], there is great potential to expand the image set. Extending the number of images in the dataset will become a priority as under-represented content domains (or over-used stimuli) are identified. In particular, we plan on extending the SMID to include a larger number of emotionally loaded, morally neutral stimuli, to facilitate studies that, for example, contrast emotional and moral valence.

Beyond increasing the size of the image set, there is also the prospect of collecting additional data for the images currently in the set. To this end, we are currently planning extensions to the image set including extending the set of variables for which data is available with the aim of bridging gaps between theories of moral psychology.

Limitations

Having highlighted how the SMID could be deployed in future research, we must also acknowledge that there are applications for which the database is less well suited (e.g., contrasting first and third-person judgments [32]). Perhaps more importantly though, there are more subtle ways in which the SMID is limited, particularly regarding its representativeness of the moral domain. Whereas some of these limitations may be intrinsic to image sets in general, others may be overcome with future developments.

Rater representativeness. Perhaps the most obvious limitation with regards to representativeness concerns the specific population used to develop and validate the image set (i.e., AMT Workers and Australian undergraduates). Although, for the AMT sample, we recruited a politically balanced sample, there are a number of ways in which AMT samples may differ from the general population [233], which were not explicitly balanced across images. Undoubtedly, demographic factors (e.g., gender, religiosity, vegetarianism) would affect the way in which specific images are evaluated, raising interesting questions for future research.

Image content and representativeness. A further limitation concerning the representativeness of the SMID is that some content domains are (at least currently) less comprehensively covered than others. Taking immoral images as an example (considering, the number of images with positive uniqueness scores for each foundation, shown in Fig 5), the Fairness, Ingroup, and Authority foundations were substantially less well represented compared to Care and Purity. This is likely attributable to at least two sources: ease of portrayal and ease of retrieval.

Regarding ease of portrayal, some content domains may be more difficult to represent in the form of static images than others (much as some basic emotions are difficult to elicit with

specific methods [6]). For example, a prototypical Care violation (assault) can be easily portrayed in image-form (e.g., one person punching another), whereas portraying a prototypical Ingroup violation (e.g., marital infidelity) in image-form requires communicating the presence of multiple interlinked relationships (in other words, a metarelational model [234]) such that Person A is married to Person B, who is sleeping with Person C. Thus, to the extent that specific moral content domains revolve around complex role or relationship configurations, or other abstract features that are difficult to communicate in static images (e.g., morally laden mental states, or the *simultaneous* depiction of intention, action and consequences), (1) those content domains may be represented by fewer stimuli, and (2) the stimuli that do represent those domains may do so less effectively than those representing other domains.

Regarding ease of retrieval, to the extent that suitable images for a given content domain do exist, our ability to locate them will be limited by the current ability of search engines to handle the complexity of search queries containing moral content. Such highly abstract or semantically complex queries are currently a major challenge for current image retrieval methods (e.g., [235]).

More broadly, it is worth considering the extent to which photographs are representative of everyday life. Rather than being entirely representative of everyday moral phenomena, the database is at least subtly biased towards whatever happens to attract the attention of people taking (and sharing) photographs (a phenomenon referred to as capture bias or photographer bias) [236,237]. In this particular database, there is the added potential bias introduced by any differences between people who release their photos under a Creative Commons license, and those who do not. Use of Creative Commons licensing is more common in the United States than elsewhere [14], meaning that the specific photos included in the database (and accompanying metadata) may reflect a relatively WEIRD (Western, Educated, Industrialized, Rich, Democratic) perspective [238].

An important implication of the above considerations is that, even if the SMID were representative of morally laden, permissively-licensed photographs, it would not necessarily be representative of moral behavior in general. The extent of each of these potential sources of bias, and their influence on specific uses of the database, is an important topic for future research, and an important consideration for studies using the SMID.

Reliance on moral foundations theory. Finally, we must consider the implications of our use of MFT in constructing and validating the SMID. Our decision to draw on MFT was pragmatic, motivated by the theory's breadth and popularity. Although we use MFT as an organising framework, we do not argue (nor does the SMID require) that MFT provides a complete, final description of the moral universe. Although there exists research highlighting important aspects of morality potentially neglected by MFT (e.g., [122,123,239,240]), and critiquing aspects of the MFT taxonomy more generally (e.g., [90–92,241–243]), the SMID can still be used in various lines of research that are largely unrelated to MFT (e.g., [244–247]).

Most importantly, although much of the normative data refers to MFT, the *content* of the images themselves are not strictly limited to MFT-relevant content. Unlike previous stimulus sets in which materials are constrained to represent a specific theory (e.g., by excluding "ill-fitting" stimuli [26]), we made efforts to include stimuli that represent all kinds of moral content, even those that may be poorly described by MFT. As such, we anticipate the SMID to be of use to researchers working both inside and outside of MFT.

Conclusion

Motivated by the scarcity of large, diverse, systematically validated stimulus sets available to moral psychology researchers, we developed the SMID. It is our hope that the SMID will allow

researchers in the field of moral psychology to perform novel, robust and rigorous research that will ultimately make important contributions to unravelling the complexity of human moral psychology.

Availability

The SMID images, current normative data and additional resources, can be found at https://osf.io/2rqad/. Data is currently available for all images for the following variables:

- Means, standard deviations, standard errors and rating frequencies for all five moral foundations, and valence, arousal and morality

- Uniqueness scores for all five moral foundations (as defined in the Results section for Study 2)

- An alternative Euclidean distance-based measure of uniqueness for all five moral foundations, based on [109]

- The proportion of morality ratings that were either morally good, bad, or neutral (i.e., above, below, or on the midpoint of the scale)

- Image properties including average RGB values, luminance and height-width ratio

- Image metadata including URL, title, author and license

Supporting information

S1 Text. The SMID compared to existing picture sets.
(DOCX)

S2 Text. Materials for image set generation.
(DOCX)

S3 Text. Image rating participant exclusions.
(DOCX)

S4 Text. Image rating instructions.
(DOCX)

S5 Text. Quantifying image divisiveness.
(DOCX)

S1 Table. OLS regressions predicting normative ratings from physical image properties.
(HTML)

Acknowledgments

Author Contributions

Conceptualization: Damien L. Crone, Stefan Bode, Simon M. Laham.

Formal analysis: Damien L. Crone.

Methodology: Damien L. Crone, Carsten Murawski, Simon M. Laham.

Supervision: Stefan Bode, Carsten Murawski, Simon M. Laham.

Writing – original draft: Damien L. Crone.

Writing – review & editing: Damien L. Crone, Stefan Bode, Carsten Murawski, Simon M. Laham.

References

1. Marchewka A, Zurawski Ł, Jednoróg K, Grabowska A. The Nencki Affective Picture System (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behav Res Methods. 2014; 46: 596–610. https://doi.org/10.3758/s13428-013-0379-1 PMID: 23996831

2. Dan-Glauser ES, Scherer KR. The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. Behav Res Methods. 2011; 43: 468–77. https://doi.org/10.3758/s13428-011-0064-1 PMID: 21431997

3. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. Gainesville, FL; 2008.

4. Kurdi B, Lozano S, Banaji MR. Introducing the Open Affective Standardized Image Set (OASIS). Behav Res Methods. 2016; https://doi.org/10.3758/s13428-016-0715-3 PMID: 26907748

5. Bradley MM, Lang PJ. Affective Norms for English Words (ANEW): Instruction manual and affective ratings. 1999.

6. Gross JJ, Levenson RW. Emotion elicitation using films. Cogn Emot. 1995; 9: 87–108. https://doi.org/10.1080/02699939508408966

7. Hewig J, Hagemann D, Seifert J, Gollwitzer M, Naumann E, Bartussek D. A revised film set for the induction of basic emotions. Cogn Emot. 2005; 19: 1095–1109. https://doi.org/10.1080/02699930541000084

8. Gilman TL, Shaheen R, Nylocks KM, Halachoff D, Chapman J, Flynn JJ, et al. A film set for the elicitation of emotion in research: A comprehensive catalog derived from four decades of investigation. Behav Res Methods. Behavior Research Methods; 2017; https://doi.org/10.3758/s13428-016-0842-x PMID: 28078572

9. Goeleven E, De Raedt R, Leyman L, Verschuere B. The Karolinska Directed Emotional Faces: A validation study. Cogn Emot. 2008; 22: 1094–1118. https://doi.org/10.1080/02699930701626582

10. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. Presentation and validation of the Radboud Faces Database. Cogn Emot. 2010; 24: 1377–1388. https://doi.org/10.1080/02699930903485076

11. Calvo RA, D'Mello SK. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Trans Affect Comput. 2010; 1: 18–37. https://doi.org/10.1109/T-AFFC.2010.1

12. Bradley MM, Lang PJ. The International Affective Picture System (IAPS) in the study of emotion and attention. In: Coan JA, Allen JJB, editors. Handbook of Emotion Elicitation and Assessement. New York: Oxford University Press; 2007. pp. 29–46.

13. Nosek BA, Spies JR, Motyl M. Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. Perspect Psychol Sci. 2012; 7: 615–631. https://doi.org/10.1177/1745691612459058 PMID: 26168121

14. Thomee B, Shamma DA, Friedland G, Elizalde B, Ni K, Poland D, et al. YFCC100M: The new data in multimedia research. Commun ACM. 2016; 59: 64–73. https://doi.org/10.1145/2812802

15. Harenski CL, Antonenko O, Shane MS, Kiehl KA. A functional imaging investigation of moral deliberation and moral intuition. Neuroimage. 2010; 49: 2707–2716. https://doi.org/10.1016/j.neuroimage.2009.10.062 PMID: 19878727

16. Luo Q, Nakic M, Wheatley T, Richell R, Martin A, Blair RJR. The neural basis of implicit moral attitude—An IAT study using event-related fMRI. Neuroimage. 2006; 30: 1449–1457. https://doi.org/10.1016/j.neuroimage.2005.11.005 PMID: 16418007

17. Moll J, de Oliveira-Souza R, Eslinger PJ, Bramati IE, Andreiuolo PA, Pessoa L. The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. J Neurosci. 2002; 22: 2730–2736. PMID: 11923438

18. Feinberg M, Willer R. The moral roots of environmental attitudes. Psychol Sci. 2013; 24: 56–62. https://doi.org/10.1177/0956797612449177 PMID: 23228937

19. Day M V, Fiske ST, Downing EL, Trail TE. Shifting liberal and conservative attitudes using moral foundations theory. Personal Soc Psychol Bull. 2014; 40: 1559–73. https://doi.org/10.1177/0146167214551152 PMID: 25286912

20. Kidwell B, Farmer A, Hardesty DM. Getting liberals and conservatives to go green: Political ideology and congruent appeals. J Consum Res. 2013; 40: 350–367. https://doi.org/10.1086/670610

21. Sagi E, Dehghani M. Measuring moral rhetoric in text. Soc Sci Comput Rev. 2013; 32: 132–144. https://doi.org/10.1177/0894439313506837

22. Christensen JF, Gomila A. Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. Neurosci Biobehav Rev. 2012; 36: 1249–64. https://doi.org/10.1016/j.neubiorev.2012.02.008 PMID: 22353427

23. McGuire J, Langdon R, Coltheart M, Mackenzie C. A reanalysis of the personal/impersonal distinction in moral psychology research. J Exp Soc Psychol. 2009; 45: 577–580. https://doi.org/10.1016/j.jesp.2009.01.002

24. Gray K, Keeney JE. Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. Soc Psychol Personal Sci. 2015; 6: 859–868. https://doi.org/10.1177/1948550615592241

25. Trémolière B, De Neys W. Methodological concerns in moral judgement research: Severity of harm shapes moral decisions. J Cogn Psychol. 2013; 25: 989–993. https://doi.org/10.1080/20445911.2013.841169

26. Clifford S, Iyengar V, Cabeza R, Sinnott-Armstrong W. Moral Foundations Vignettes: A standardized stimulus database of scenarios based on moral foundations theory. Behav Res Methods. 2015; 47: 1178–1198. https://doi.org/10.3758/s13428-014-0551-2

27. Chadwick RA, Bromgard G, Bromgard I, Trafimow D. An index of specific behaviors in the moral domain. Behav Res Methods. 2006; 38: 692–697. https://doi.org/10.3758/BF03193902 PMID: 17393841

28. Lotto L, Manfrinati A, Sarlo M. A new set of moral dilemmas: Norms for moral acceptability, decision times, and emotional salience. J Behav Decis Mak. 2014; 27: 57–65. https://doi.org/10.1002/bdm.1782

29. Knutson KM, Krueger F, Koenigs MR, Hawley A, Escobedo JR, Vasudeva V, et al. Behavioral norms for condensed moral vignettes. Soc Cogn Affect Neurosci. 2010; 5: 378–84. https://doi.org/10.1093/scan/nsq005 PMID: 20154053

30. May MA, Hartshorne H. Objective methods of measuring character. Pedagog Semin J Genet Psychol. 1925; 32: 45–67. https://doi.org/10.1080/08856559.1925.10532317

31. Pittel SM, Mendelsohn GA. Measurement of moral values: A review and critique. Psychol Bull. 1966; 66: 22–35. https://doi.org/10.1037/h0023425 PMID: 5329602

32. Boccia M, Dacquino C, Piccardi L, Cordellieri P, Guariglia C, Ferlazzo F, et al. Neural foundation of human moral reasoning: an ALE meta-analysis about the role of personal perspective. Brain Imaging Behav. 2016; https://doi.org/10.1007/s11682-016-9505-x PMID: 26809288

33. Chapman HA, Anderson AK. Things rank and gross in nature: A review and synthesis of moral disgust. Psychol Bull. 2013; 139: 300–327. https://doi.org/10.1037/a0030964 PMID: 23458435

34. Judd CM, Westfall J, Kenny DA. Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. J Pers Soc Psychol. 2012; 103: 54–69. https://doi.org/10.1037/a0028347 PMID: 22612667

35. Wells GL, Windschitl PD. Stimulus sampling and social psychological experimentation. Personal Soc Psychol Bull. 1999; 25: 1115–1125. https://doi.org/10.1177/01461672992512005

36. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen. 2014; 143: 2020–2045. https://doi.org/10.1037/xge0000014 PMID: 25111580

37. Whitsett DD, Shoda Y. An approach to test for individual differences in the effects of situations without using moderator variables. J Exp Soc Psychol. 2014; 50: 94–104. https://doi.org/10.1016/j.jesp.2013.08.008 PMID: 24550572

38. Westfall J, Nichols TE, Yarkoni T. Fixing the stimulus-as-fixed-effect fallacy in task fMRI. Wellcome Open Res. 2016; 1: 23. https://doi.org/10.12688/wellcomeopenres.10298.2 PMID: 28503664

39. Chiong W, Wilson SM, D'Esposito M, Kayser AS, Grossman SN, Poorzand P, et al. The salience network causally influences default mode network activity during moral reasoning. Brain. 2013; https://doi.org/10.1093/brain/awt066 PMID: 23576128

40. Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. An fMRI investigation of emotional engagement in moral judgment. Science (80-). 2001; 293: 2105–2108. https://doi.org/10.1126/science.1062872 PMID: 11557895

41. Ahn W-Y, Kishida KT, Gu X, Lohrenz T, Harvey A, Alford JR, et al. Nonpolitical images evoke neural predictors of political ideology. Curr Biol. 2014; 24: 2693–2699. https://doi.org/10.1016/j.cub.2014.09.050 PMID: 25447997

42. Haidt J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychol Rev. 2001; 108: 814–834. https://doi.org/10.1037//0033-295X PMID: 11699120

43. Van Bavel JJ, FeldmanHall O, Mende-Siedlecki P. The neuroscience of moral cognition: From dual processes to dynamic systems. Curr Opin Psychol. 2015; 6: 167–172. https://doi.org/10.1016/j.copsyc.2015.08.009

44. Greene JD. Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. Ethics. 2014; 124: 695–726. https://doi.org/10.1086/675875

45. Buon M, Seara-Cardoso A, Viding E. Why (and how) should we study the interplay between emotional arousal, Theory of Mind, and inhibitory control to understand moral cognition? Psychon Bull Rev. 2016; https://doi.org/10.3758/s13423-016-1042-5 PMID: 27169411

46. Nichols S. Norms with feeling: Towards a psychological account of moral judgment. Cognition. 2002; 84: 221–236. https://doi.org/10.1016/S0010-0277(02)00048-3 PMID: 12175573

47. Mikhail J. Universal moral grammar: Theory, evidence and the future. Trends Cogn Sci. 2007; 11: 143–52. https://doi.org/10.1016/j.tics.2006.12.007 PMID: 17329147

48. Gantman AP, Van Bavel JJ. The moral pop-out effect: Enhanced perceptual awareness of morally relevant stimuli. Cognition. 2014; 132: 22–9. https://doi.org/10.1016/j.cognition.2014.02.007 PMID: 24747444

49. Gray K, Schein C, Ward AF. The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. J Exp Psychol Gen. 2014; 143: 1600–15. https://doi.org/10.1037/a0036149 PMID: 24635184

50. Cameron CD, Payne BK, Sinnott-Armstrong W, Scheffer JA, Inzlicht M. Implicit moral evaluations: A multinomial modeling approach. Cognition. 2017; 158: 224–241. https://doi.org/10.1016/j.cognition.2016.10.013 PMID: 27865113

51. Imura M, Burkley M, Brown RP. Honor to the core: Measuring implicit honor ideology endorsement. Pers Individ Dif. 2014; 59: 27–31. https://doi.org/10.1016/j.paid.2013.10.025

52. Cutler A. Making up materials is a confounded nuisance, or: Will we able to run any psycholinguistic experiments at all in 1990? Cognition. 1981; 10: 65–70. https://doi.org/10.1016/0010-0277(81)90026-3 PMID: 7198562

53. Michel J-B, Shen YK, Aiden AP, Veres A, Gray MK, The Google Books Team, et al. Quantitative analysis of culture using millions of digitized books. Science (80-). 2011; 331: 176–182. https://doi.org/10.1126/science.1199644 PMID: 21163965

54. De Houwer J, Hermans D. Differences in the affective processing of words and pictures. Cogn Emot. 1994; 8: 1–20. https://doi.org/10.1080/02699939408408925

55. Slovic P, Västfjäll D, Erlandsson A, Gregory R. Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. Proc Natl Acad Sci. 2017; 1–5. https://doi.org/10.1073/pnas.1613977114 PMID: 28074038

56. Liberman N, Trope Y. Traversing psychological distance. Trends Cogn Sci. 2014; 18: 364–369. https://doi.org/10.1016/j.tics.2014.03.001 PMID: 24726527

57. Trope Y, Liberman N. Construal-level theory of psychological distance. Psychol Rev. 2010; 117: 440–463. https://doi.org/10.1037/a0018963 PMID: 20438233

58. Carnevale JJ, Fujita K, Han HA, Amit E. Immersion versus transcendence: How pictures and words impact evaluative associations assessed by the implicit association test. Soc Psychol Personal Sci. 2015; 6: 92–100. https://doi.org/10.1177/1948550614546050

59. Rim S, Amit E, Fujita K, Trope Y, Halbeisen G, Algom D. How words transcend and pictures immerse: On the association between medium and level of construal. Soc Psychol Personal Sci. 2015; 6: 123–130. https://doi.org/10.1177/1948550614548728

60. Amit E, Greene JD. You see, the ends don't justify the means: Visual imagery and moral judgment. Psychol Sci. 2012; 23: 861–868. https://doi.org/10.1177/0956797611434965 PMID: 22745347

61. Eyal T, Sagristano MD, Trope Y, Liberman N, Chaiken S. When values matter: Expressing values in behavioral intentions for the near vs. distant future. J Exp Soc Psychol. 2009; 45: 35–43. https://doi.org/10.1016/j.jesp.2008.07.023 PMID: 21822329

62. Luguri JB, Napier JL, Dovidio JF. Reconstruing intolerance: Abstract thinking reduces conservatives' prejudice against nonnormative groups. Psychol Sci. 2012; 23: 756–763. https://doi.org/10.1177/0956797611433877 PMID: 22653799

63. Napier JL, Luguri JB. Moral mind-sets: Abstract thinking increases a preference for "Individualizing" over "Binding" moral foundations. Soc Psychol Personal Sci. 2013; 4: 754–759. https://doi.org/10.1177/1948550612473783

64. Rogers R, Vess M, Routlege C. Construal level shapes associations between political conservatism and reactions to male same-sex intimacy. Soc Psychol (Gott). 2016; 47: 87–97. https://doi.org/10.1027/1864-9335/a000261

65. Woodrow H. A picture-preference character test. J Educ Psychol. 1926; 17: 519–531. https://doi.org/10.1037/h0073723

66. McGrath MC. A story of the moral development of children. Psychol Monogr. 1923; 32: i–190. https://doi.org/10.1037/h0093194

67. Wayne I. American and Soviet themes and values: A content analysis of pictures in popular magazines. Public Opin Q. 1956; 20: 314. https://doi.org/10.1086/266620

68. Eberhart JC. The use of pictures in the estimation of the seriousness of property offenses. Pedagog Semin J Genet Psychol. 1940; 56: 411–437. https://doi.org/10.1080/08856559.1940.10534510

69. Von Baldegg KC-M. The best thing since sliced bread: A brief history of sliced bread. The Atlantic. 2012: 3–7. Available: http://www.theatlantic.com/video/archive/2012/07/the-best-thing-since-sliced-bread-a-brief-history-of-sliced-bread/260202/

70. Steckler CM, Hamlin JK, Miller MB, King D, Kingstone A. Moral judgement by the disconnected left and right cerebral hemispheres: a split-brain investigation. R Soc Open Sci. 2017; 4: 170172. https://doi.org/10.1098/rsos.170172 PMID: 28791143

71. Nosek BA, Hawkins CB, Frazier RS. Implicit social cognition: From measures to mechanisms. Trends Cogn Sci. 2011; 15: 152–159. https://doi.org/10.1016/j.tics.2011.01.005 PMID: 21376657

72. Tusche A, Kahnt T, Wisniewski D, Haynes J-D. Automatic processing of political preferences in the human brain. Neuroimage. 2013; 72: 174–82. https://doi.org/10.1016/j.neuroimage.2013.01.020 PMID: 23353599

73. McLean SP, Garza JP, Wiebe SA, Dodd MD, Smith KB, Hibbing JR, et al. Applying the flanker task to political psychology: A research note. Polit Psychol. 2014; 35: 831–840. https://doi.org/10.1111/pops.12056

74. Chakroff A, Dungan JA, Koster-Hale J, Brown A, Saxe R, Young LL. When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. Soc Cogn Affect Neurosci. 2016; 11: 476–484. https://doi.org/10.1093/scan/nsv131 PMID: 26628642

75. Lee JJ, Sohn Y, Fowler JH. Emotion regulation as the foundation of political attitudes: Does reappraisal decrease support for conservative policies? PLoS One. 2013; 8: e83143. https://doi.org/10.1371/journal.pone.0083143 PMID: 24367583

76. Koenigs MR, Young LL, Adolphs R, Tranel D, Cushman FA, Hauser MD, et al. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature. 2007; 446: 908–911. https://doi.org/10.1038/nature05631 PMID: 17377536

77. Uhlmann EL, Pizarro DA, Diermeier D. A person-centered approach to moral judgment. Perspect Psychol Sci. 2015; 10: 72–81. https://doi.org/10.1177/1745691614556679 PMID: 25910382

78. Piazza J, Goodwin GP, Rozin P, Royzman EB. When a virtue is not a virtue: Conditional virtues in moral evaluation. Soc Cogn. 2014; 32: 528–558. https://doi.org/10.1521/soco.2014.32.6.528

79. Carr AR, Paholpak P, Daianu M, Fong SS, Mather M, Jimenez EE, et al. An investigation of care-based vs. rule-based morality in frontotemporal dementia, Alzheimer's disease, and healthy controls. Neuropsychologia. 2015; 78: 73–79. https://doi.org/10.1016/j.neuropsychologia.2015.09.033 PMID: 26432341

80. Glenn AL, Iyer R, Graham J, Koleva SP, Haidt J. Are all types of morality compromised in psychopathy? J Pers Disord. 2009; 23: 384–398. https://doi.org/10.1521/pedi.2009.23.4.384 PMID: 19663658

81. Blair RJR, White SF, Meffert H, Hwang S. Emotional learning and the development of differential moralities: Implications from research on psychopathy. Ann N Y Acad Sci. 2013; 1299: 36–41. https://doi.org/10.1111/nyas.12169 PMID: 25684831

82. Marshall J, Watts AL, Lilienfeld SO. Do psychopathic individuals possess a misaligned moral compass? A meta-analytic examination of psychopathy's relations with moral judgment. Personal Disord Theory, Res Treat. 2016; https://doi.org/10.1037/per0000226 PMID: 27797544

83. Sinnott-Armstrong W, Wheatley T. Are moral judgments unified? Philos Psychol. 2013; 27: 451–474. https://doi.org/10.1080/09515089.2012.736075

84. Graham J, Haidt J, Motyl M, Meindl P, Iskiwitch C, Mooijman M. Moral Foundations Theory: On the advantages of moral pluralism over moral monism. In: Gray K, Graham J, editors. The Atlas of Moral Psychology: Mapping Good and Evil in the Mind. New York: Guilford Press; 2016.

85. Graham J, Haidt J, Koleva SP, Motyl M, Iyer R, Wojcik SP, et al. Moral Foundations Theory: The pragmatic validity of moral pluralism. Adv Exp Soc Psychol. 2013; 47: 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002–4

86. Haidt J, Joseph CM. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. Daedalus. 2004; 133: 55–66. https://doi.org/10.1162/0011526042365555

87. Haidt J, Joseph CM. The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In: Carruthers P, Laurence S, Stich S, editors. The Innate Mind, Volume 3: Foundations and the Future. New York: Oxford University Press; 2007. pp. 367–391. https://doi.org/10.1093/acprof:oso/9780195332834.003.0019

88. Iyer R, Koleva SP, Graham J, Ditto PH, Haidt J. Understanding libertarian morality: The psychological dispositions of self-identified libertarians. PLoS One. 2012; 7: e42366. https://doi.org/10.1371/journal.pone.0042366 PMID: 22927928

89. Haidt J. The righteous mind: Why good people are divided by politics and religion. London: Penguin Books; 2012.

90. Suhler CL, Churchland PS. Can innate, modular "foundations" explain morality? Challenges for Haidt's Moral Foundations Theory. J Cogn Neurosci. 2011; 23: 2103–2116. https://doi.org/10.1162/jocn.2011.21637 PMID: 21291315

91. Smith KB, Alford JR, Hibbing JR, Martin NG, Hatemi PK. Intuitive ethics and political orientations: Testing moral foundations as a theory of political ideology. Am J Pol Sci. 2016; https://doi.org/10.1111/ajps.12255

92. Cameron CD, Lindquist KA, Gray K. A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. Personal Soc Psychol Rev. 2015; 19: 371–394. https://doi.org/10.1177/1088868314566683 PMID: 25587050

93. Thompson B, Kirby S, Smith K. Culture shapes the evolution of cognition. Proc Natl Acad Sci. 2016; 113: 4530–4535. https://doi.org/10.1073/pnas.1523631113 PMID: 27044094

94. Haidt J, Joseph CM. How moral foundations theory succeeded in building on sand: A response to Suhler and Churchland. J Cogn Neurosci. 2011; 23: 2117–2122. https://doi.org/10.1162/jocn.2011.21638

95. Royzman EB, Kim K, Leeman RF. The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. Judgm Decis Mak. 2015; 10: 296–313.

96. Gutierrez R, Giner-Sorolla R. Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. Emotion. 2007; 7: 853–868. https://doi.org/10.1037/1528-3542.7.4.853 PMID: 18039054

97. Frimer JA, Tell CE, Haidt J. Liberals condemn sacrilege too: The harmless desecration of Cerro Torre. Soc Psychol Personal Sci. 2015; 6: 878–886. https://doi.org/10.1177/1948550615597974

98. Rottman J, Kelemen D, Young LL. Tainting the soul: Purity concerns predict moral judgments of suicide. Cognition. 2014; 130: 217–26. https://doi.org/10.1016/j.cognition.2013.11.007 PMID: 24333538

99. Ekman P. An argument for basic emotions. Cogn Emot. 1992; 6: 169–200. https://doi.org/10.1080/02699939208411068

100. Ekman P, Cordaro D. What is meant by calling emotions basic. Emot Rev. 2011; 3: 364–370. https://doi.org/10.1177/1754073911410740

101. Larsen JT, McGraw AP. The case for mixed emotions. Soc Personal Psychol Compass. 2014; 8: 263–274. https://doi.org/10.1111/spc3.12108

102. Berrios R, Totterdell P, Kellett S. Eliciting mixed emotions: a meta-analysis comparing models, types, and measures. Front Psychol. 2015; 6: 1–15. https://doi.org/10.3389/fpsyg.2015.00001

103. Trampe D, Quoidbach J, Taquet M. Emotions in everyday life. PLoS One. 2015; 10: e0145450. https://doi.org/10.1371/journal.pone.0145450 PMID: 26698124

104. Riegel M, Żurawski Ł, Wierzba M, Moslehi A, Klocek Ł, Horvat M, et al. Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE). Behav Res Methods. 2016; 48: 600–612. https://doi.org/10.3758/s13428-015-0620-1 PMID: 26205422

105. Stevenson RA, Mikels JA, James TW. Characterization of the Affective Norms for English Words by discrete emotional categories. Behav Res Methods. 2007; 39: 1020–1024. https://doi.org/10.3758/BF03192999 PMID: 18183921

106. Libkuman TM, Otani H, Kern R, Viger SG, Novak N. Multidimensional normative ratings for the International Affective Picture System. Behav Res Methods. 2007; 39: 326–334. https://doi.org/10.3758/BF03193164 PMID: 17695361

107. Haberkamp A, Glombiewski JA, Schmidt F, Barke A. The DIsgust-RelaTed-Images (DIRTI) database: Validation of a novel standardized set of disgust pictures. Behav Res Ther. Elsevier Ltd; 2017; 89: 86–94. https://doi.org/10.1016/j.brat.2016.11.010 PMID: 27914317

108. Stevenson RA, James TW. Affective auditory stimuli: Characterization of the International Affective Digitized Sounds (IADS) by discrete emotional categories. Behav Res Methods. 2008; 40: 315–321. https://doi.org/10.3758/BRM.40.1.315 PMID: 18411555

109. Wierzba M, Riegel M, Wypych M, Jednoróg K, Turnau P, Grabowska A, et al. Basic emotions in the Nencki Affective Word List (NAWL BE): New method of classifying emotional stimuli. PLoS One. 2015; 10: e0132305. https://doi.org/10.1371/journal.pone.0132305 PMID: 26148193

110. Nabi RL. The theoretical versus the lay meaning of disgust: Implications for emotion research. Cogn Emot. 2002; 16: 695–703. https://doi.org/10.1080/02699930143000437

111. Russell PS, Giner-Sorolla R. Bodily moral disgust: What it is, how it is different from anger, and why it is an unreasoned emotion. Psychol Bull. 2013; 139: 328–351. https://doi.org/10.1037/a0029319 PMID: 23458436

112. Sousa P, Piazza J. Harmful transgressions qua moral transgressions: A deflationary view. Think Reason. 2014; 20: 99–128. https://doi.org/10.1080/13546783.2013.834845

113. Fiske AP, Rai TS. Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships. Cambridge, UK: Cambridge University Press; 2015.

114. Simpson A, Laham SM, Fiske AP. Wrongness in different relationships: Relational context effects on moral judgment. J Soc Psychol. 2016; 156: 594–609. https://doi.org/10.1080/00224545.2016.1140118 PMID: 26751010

115. Waytz A, Dungan JA, Young LL. The whistleblower's dilemma and the fairness-loyalty tradeoff. J Exp Soc Psychol. 2013; 49: 1027–1033. https://doi.org/10.1016/j.jesp.2013.07.002

116. Fernando JW, Kashima Y, Laham SM. Multiple emotions: A person-centered approach to the relationship between intergroup emotion and action orientation. Emotion. 2014; 14: 722–732. https://doi.org/10.1037/a0036103 PMID: 24749637

117. Dhami MK, Hertwig R, Hoffrage U. The role of representative design in an ecological approach to cognition. Psychol Bull. 2004; 130: 959–988. https://doi.org/10.1037/0033-2909.130.6.959 PMID: 15535744

118. Todd PM, Gigerenzer G. Environments that make us smart: Ecological rationality. Curr Dir Psychol Sci. 2007; 16: 167–171. https://doi.org/10.1111/j.1467-8721.2007.00497.x

119. Royzman EB. Are experiments possible? The limitations of a posteriori control in experimental behavior analysis: The case of clinical process research. Theory Psychol. 2000; 10: 171–196. https://doi.org/10.1177/0959354300102002

120. Dahlsgaard K, Peterson C, Seligman MEP. Shared virtue: The convergence of valued human strengths across culture and history. Rev Gen Psychol. 2005; 9: 203–213. https://doi.org/10.1037/1089-2680.9.3.203

121. Gray K, Young LL, Waytz A. Mind perception is the essence of morality. Psychol Inq. 2012; 23: 101–124. https://doi.org/10.1080/1047840X.2012.651387 PMID: 22754268

122. Janoff-Bulman R, Carnes NC. Surveying the moral landscape: Moral motives and group-based moralities. Personal Soc Psychol Rev. 2013; 17: 219–36. https://doi.org/10.1177/1088868313480274 PMID: 23504824

123. Hofmann W, Wisneski DC, Brandt MJ, Skitka LJ. Morality in everyday life. Science (80-). 2014; 345: 1340–1343. https://doi.org/10.1126/science.1251560 PMID: 25214626

124. Cacioppo JT, Berntson GG. Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. Psychol Bull. 1994; 115: 401–423. https://doi.org/10.1037//0033-2909.115.3.401

125. Alves H, Koch A, Unkelbach C. Why good is more alike than bad: Processing implications. Trends Cogn Sci. 2017; 21: 69–79. https://doi.org/10.1016/j.tics.2016.12.006 PMID: 28063663

126. Rozin P, Royzman EB. Negativity bias, negativity dominance, and contagion. Personal Soc Psychol Rev. 2001; 5: 296–320. https://doi.org/10.1207/S15327957PSPR0504_2

127. Knobe J. Intentional action and side effects in ordinary language. Analysis. 2003; 63: 190–194. https://doi.org/10.1093/analys/63.3.190

128. Wiltermuth SS, Monin B, Chow RM. The orthogonality of praise and condemnation in moral judgment. Soc Psychol Personal Sci. 2010; 1: 302–310. https://doi.org/10.1177/1948550610363162

129. Baveye Y, Dellandréa E, Chamaret C, Liming Chen. LIRIS-ACCEDE: A video database for affective content analysis. IEEE Trans Affect Comput. 2015; 6: 43–55. https://doi.org/10.1109/TAFFC.2015.2396531

130. Lindeman M, Koirikivi I, Lipsanen J. Pictorial Empathy Test (PET): An easy-to-use method for assessing affective empathic reactions. Eur J Psychol Assess. 2016; 1–11. https://doi.org/10.1027/1015-5759/a000353

131. Buhrmester MD, Kwang T, Gosling SD. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? Perspect Psychol Sci. 2011; 6: 3–5. https://doi.org/10.1177/1745691610393980 PMID: 26162106

132. Mason WA, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods. 2012; 44: 1–23. https://doi.org/10.3758/s13428-011-0124-6 PMID: 21717266

133. Crump MJC, McDonnell J V, Gureckis TM. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PLoS One. 2013; 8: e57410. https://doi.org/10.1371/journal.pone.0057410 PMID: 23516406

134. Graham J, Nosek BA, Haidt J, Iyer R, Koleva SP, Ditto PH. Mapping the moral domain. J Pers Soc Psychol. 2011; 101: 366–85. https://doi.org/10.1037/a0021847 PMID: 21244182

135. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychol Bull. 1979; 86: 420–428. https://doi.org/10.1037//0033-2909.86.2.420 PMID: 18839484

136. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996; 1: 30–46. https://doi.org/10.1037/1082-989X.1.1.30

137. LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. Organ Res Methods. 2008; 11: 815–852. https://doi.org/10.1177/1094428106296642

138. Gamer M, Lemon J, Fellows I, Singh P. irr: Various coefficients of interrater reliability and agreement [Internet]. 2012. Available: http://cran.r-project.org/package=irr

139. Cicchetti D V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994; 6: 284–290. https://doi.org/10.1037/1040-3590.6.4.284

140. Shweder RA, Much NC, Mahapatra M, Park L. The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. In: Brandt A, Rozin P, editors. Morality and Health. New York: Routledge; 1997.

141. Schein C, Gray K. The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. Personal Soc Psychol Bull. 2015; 41: 1147–1163. https://doi.org/10.1177/0146167215591501 PMID: 26091912

142. Gray K, Keeney JE. Disconfirming Moral Foundations Theory on its own terms: Reply to Graham (2015). Soc Psychol Personal Sci. 2015; 6: 874–877. https://doi.org/10.1177/1948550615592243

143. Kievit RA, Frankenhuis WE, Waldorp LJ, Borsboom D. Simpson's paradox in psychological science: a practical guide. Front Psychol. 2013; 4: 1–14. https://doi.org/10.3389/fpsyg.2013.00001

144. Robinson WS. Ecological correlations and the behavior of individuals. Am Sociol Rev. 1950; 15: 351. https://doi.org/10.2307/2087176

145. Kuppens P, Tuerlinckx F, Russell JA, Barrett LF. The relation between valence and arousal in subjective experience. Psychol Bull. 2013; 139: 917–940. https://doi.org/10.1037/a0030811 PMID: 23231533

146. Koleva SP, Graham J, Iyer R, Ditto PH, Haidt J. Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. J Res Pers. 2012; 46: 184–194. https://doi.org/10.1016/j.jrp.2012.01.006

147. Cushman FA, Young LL, Greene JD. Our multi-system moral psychology: Towards a consensus view. The Moral Psychology Handbook. Oxford University Press; 2010. pp. 47–71. https://doi.org/10.1093/acprof:oso/9780199582143.003.0003

148. Greene JD, Haidt J. How (and where) does moral judgment work? Trends Cogn Sci. 2002; 6: 517–523. https://doi.org/10.1016/S1364-6613(02)02011-9 PMID: 12475712

149. Miller RM, Cushman FA. Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. Soc Personal Psychol Compass. 2013; 7: 707–718. https://doi.org/10.1111/spc3.12066

150. Corradi-Dell'Acqua C, Tusche A, Vuilleumier P, Singer T. Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. Nat Commun. 2016; 7: 10904. https://doi.org/10.1038/ncomms10904 PMID: 26988654

151. Cheng JS, Ottati VC, Price ED. The arousal model of moral condemnation. J Exp Soc Psychol. 2013; 49: 1012–1018. https://doi.org/10.1016/j.jesp.2013.06.006

152. Hibbing JR, Smith KB, Alford JR. Differences in negativity bias underlie variations in political ideology. Behav Brain Sci. 2014; 37: 297–307. https://doi.org/10.1017/S0140525X13001192 PMID: 24970428

153. Baumert A, Gollwitzer M, Staubach M, Schmitt M. Justice sensitivity and the processing of justice-related information. Eur J Pers. 2011; 25: 386–397. https://doi.org/10.1002/per.800

154. Bartoszek G, Cervone D. Toward an implicit measure of emotions: Ratings of abstract images reveal distinct emotional states. Cogn Emot. 2016; https://doi.org/10.1080/02699931.2016.1225004 PMID: 27603515

155. Murphy ST, Zajonc RB. Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. J Pers Soc Psychol. 1993; 64: 723–739. https://doi.org/10.1037/0022-3514.64.5.723 PMID: 8505704

156. Baron-Cohen S, Wheelwright S, Hill J, Raste Y, Plumb I. The "Reading the Mind in the Eyes" Test Revised Version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. J Child Psychol Psychiatry. 2001; 42: 241–251. https://doi.org/10.1111/1469-7610.00715 PMID: 11280420

157. Gantman AP, Van Bavel JJ. Exposure to justice diminishes moral perception. J Exp Psychol Gen. 2016; 145: 1728–1739. https://doi.org/10.1037/xge0000241 PMID: 27935734

158. Navarick DJ. Moral ambivalence: Modeling and measuring bivariate evaluative processes in moral judgment. Rev Gen Psychol. 2013; 17: 443–452. https://doi.org/10.1037/a0034527

159. Schneider IK, Veenstra L, Harreveld F Van, Schwarz N, Sander LK. Let's not be indifferent about neutrality: Neutral ratings in the International Affective Picture System (IAPS) mask mixed affective responses. Emotion. 2016; 16: 426–430. https://doi.org/10.1037/emo0000164 PMID: 26950363

160. Jonas K, Broemer P, Diehl M. Attitudinal ambivalence. Eur Rev Soc Psychol. 2000; 11: 35–74. https://doi.org/10.1080/14792779943000125

161. Conner M, Sparks P. Ambivalence and attitudes. Eur Rev Soc Psychol. 2002; 12: 37–70. https://doi.org/10.1080/14792772143000012

162. Larsen JT, Norris CJ, McGraw AP, Hawkley LC, Cacioppo JT. The evaluative space grid: A single-item measure of positivity and negativity. Cogn Emot. 2009; 23: 453–480. https://doi.org/10.1080/02699930801994054

163. Dodd MD, Hibbing JR, Smith KB. The politics of attention: Differences in visual cognition between liberals and conservatives. Psychology of Learning and Motivation. Elsevier Ltd; 2016. pp. 277–309. https://doi.org/10.1016/bs.plm.2016.04.003

164. Michałowski JM, Droździel D, Matuszewski J, Koziejowski W, Jednoróg K, Marchewka A. The Set of Fear Inducing Pictures (SFIP): Development and validation in fearful and nonfearful individuals. Behav Res Methods. 2016; 1–13. https://doi.org/10.3758/s13428-015-0567-2

165. Sloan DM, Sege CT, McSweeney LB, Suvak MK, Shea MT, Litz BT. Development of a Borderline Personality Disorder—Relevant Picture Stimulus Set. J Pers Disord. 2010; 24: 664–675. https://doi.org/10.1521/pedi.2010.24.5.664 PMID: 20958174

166. Eddie D, Bates ME. Toward validation of a Borderline Personality Disorder–relevant picture set. Personal Disord Theory, Res Treat. 2016; https://doi.org/10.1037/per0000173 PMID: 27046392

167. Greenwald AG. There is nothing so theoretical as a good method. Perspect Psychol Sci. 2012; 7: 99–108. https://doi.org/10.1177/1745691611434210 PMID: 26168438

168. Jarudi I, Kreps T, Bloom P. Is a refrigerator good or evil? The moral evaluation of everyday objects. Soc Justice Res. 2008; 21: 457–469. https://doi.org/10.1007/s11211-008-0082-z

169. Cavrak SE, Kleider-Offutt HM. Pictures are worth a thousand words and a moral decision or two: Religious symbols prime moral judgments. Int J Psychol Relig. 2015; 25: 173–192. https://doi.org/10.1080/10508619.2014.921111

170. Becker JC, Butz DA, Sibley CG, Barlow FK, Bitacola LM, Christ O, et al. What do national flags stand for? An exploration of associations across 11 countries. J Cross Cult Psychol. 2017; 48: 335–352. https://doi.org/10.1177/0022022116687851

171. Chandler JJ, Mueller P, Paolacci G. Nonnaïveté among Amazon Mechanical Turk workers: consequences and solutions for behavioral researchers. Behav Res Methods. 2014; 46: 112–30. https://doi.org/10.3758/s13428-013-0365-7 PMID: 23835650

172. Stewart N, Ungemach C, Harris AJL, Bartels DM, Paolacci G, Chandler JJ. The average laboratory samples a population of 7,300 Amazon Mechanical Turk workers: The size of the MTurk population. Judgm Decis Mak. 2015; 10: 479–491.

173. Klein RA, Ratliff KA, Vianello M, Adams RB, Bahník Š, Bernstein MJ, et al. Investigating variation in replicability. Soc Psychol (Gott). 2014; 45: 142–152. https://doi.org/10.1027/1864-9335/a000178

174. Open Science Collaboration. Estimating the reproducibility of psychological science. Science (80-). 2015; 349: aac4716-1–aac4716-8. https://doi.org/10.1126/science.aac4716 PMID: 26315443

175. Schweinsberg M, Madan N, Vianello M, Sommer SA, Jordan J, Tierney W, et al. The pipeline project: Pre-publication independent replications of a single laboratory's research pipeline. J Exp Soc Psychol. 2016; 66: 55–67. https://doi.org/10.1016/j.jesp.2015.10.001

176. Westfall J, Judd CM, Kenny DA. Replicating studies in which samples of participants respond to samples of stimuli. Perspect Psychol Sci. 2015; 10: 390–399. https://doi.org/10.1177/1745691614564879 PMID: 25987517

177. Monin B, Oppenheimer DM. The limits of direct replications and the virtues of stimulus sampling. Soc Psychol (Gott). 2014; 45: 299–300. https://doi.org/10.1027/1864-9335/a000202

178. Bahník Š, Vranka MA. If it's difficult to pronounce, it might not be risky. Psychol Sci. 2017; 95679761668577. https://doi.org/10.1177/0956797616685770 PMID: 28406381

179. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14: 365–76. https://doi.org/10.1038/nrn3475 PMID: 23571845

180. Bauman CW, McGraw AP, Bartels DM, Warren C. Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. Soc Personal Psychol Compass. 2014; 8/9: 536–554. https://doi.org/10.1111/spc3.12131

181. Adolphs R, Nummenmaa L, Todorov A, Haxby J V. Data-driven approaches in the investigation of social perception. Philos Trans R Soc Lond B Biol Sci. 2016; 371. https://doi.org/10.1098/rstb.2015.0367 PMID: 27069045

182. Tamir DI, Thornton MA, Contreras JM, Mitchell JP. Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. Proc Natl Acad Sci. 2015; 201511905. https://doi.org/10.1073/pnas.1511905112 PMID: 26621704

183. Skerry AE, Saxe R. Neural representations of emotion are organized around abstract event features. Curr Biol. 2015; 25: 1945–1954. https://doi.org/10.1016/j.cub.2015.06.009 PMID: 26212878

184. Iliev RI, Dehghani M, Sagi E. Automated text analysis in psychology: methods, applications, and future developments. Lang Cogn. 2015; 7: 265–290. https://doi.org/10.1017/langcog.2014.30

185. Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, et al. Automatic personality assessment through social media language. J Pers Soc Psychol. 2015; 108: 934–952. https://doi.org/10.1037/pspp0000020 PMID: 25365036

186. Yarkoni T. Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. J Res Pers. 2010; 44: 363–373. https://doi.org/10.1016/j.jrp.2010.04.001 PMID: 20563301

187. Kern ML, Park G, Eichstaedt JC, Schwartz HA, Sap M, Smith LK, et al. Gaining insights from social media language: Methodologies and challenges. Psychol Methods. 2016; https://doi.org/10.1037/met0000091 PMID: 27505683

188. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol. 2009; 29: 24–54. https://doi.org/10.1177/0261927X09351676

189. Olney AM, Dale R, D'Mello SK. The World Within Wikipedia: An Ecology of Mind. Information. 2012; 3: 229–255. https://doi.org/10.3390/info3020229

190. Mehdi M, Okoli C, Mesgari M, Nielsen FA, Lanamäki A. Excavating the mother lode of human-generated text: A systematic review of research that uses the wikipedia corpus. Inf Process Manag. 2017; 53: 505–529. https://doi.org/10.1016/j.ipm.2016.07.003

191. Chen J,Hsieh G, Mahmud JU, Nichols J. Understanding individuals' personal values from social media word use. Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing. New York, New York, USA: ACM Press; 2014. pp. 405–414. https://doi.org/10.1145/2531602.2531608

192. Haber EM. On the stability of online language features: How much text do you need to know a person? arXiv Prepr. 2015; 1504.06391v1. Available: http://arxiv.org/abs/1504.06391

193. Fulgoni D, Carpenter J, Ungar LH,Preot D. An empirical exploration of moral foundations theory in partisan news sources. Proceedings of the 10th edition of the Language Resources and Evaluation Conference. Portorož, Slovenia; 2016.

194. Kaur R, Sasahara K. Quantifying moral foundations from various topics on Twitter conversations. arXiv Prepr. 2016; 1610.02991v1. Available: http://arxiv.org/abs/1610.02991

195. Bardi A, Calogero RM, Mullen B. A new archival approach to the study of values and value—behavior relations: Validation of the value lexicon. J Appl Psychol. 2008; 93: 483–497. https://doi.org/10.1037/0021-9010.93.3.483 PMID: 18457482

196. Graham J, Haidt J, Nosek BA. Liberals and conservatives rely on different sets of moral foundations. J Pers Soc Psychol. 2009; 96: 1029–1046. https://doi.org/10.1037/a0015141 PMID: 19379034

197. Teernstra L, Putten P van der, Noordegraaf-Eelens L, Verbeek F. The morality machine: Tracking moral values in tweets. In: Boström H, Knobbe A, Soares C, Papapetrou P, editors. Advances in Intelligent Data Analysis XV. Springer; 2016. pp. 26–37. https://doi.org/10.1007/978-3-319-46349-03

198. Recchia G, Louwerse MM. Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. Q J Exp Psychol. 2014; 1–15. https://doi.org/10.1080/17470218.2014.941296 PMID: 24998307

199. Van Rensbergen B, De Deyne S, Storms G. Estimating affective word covariates using word association data. Behav Res Methods. 2016; 48: 1644–1652. https://doi.org/10.3758/s13428-015-0680-2 PMID: 26511372

200. You Q, Luo J,Jin H, Yang J. Building a large scale dataset for image emotion recognition: The fine print and the benchmark. Proceedings of the 30th Conference on Artificial Intelligence (AAAI 2016). 2016. pp. 308–314. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12272

201. Armstrong BC, Watson CE, Plaut DC. SOS! An algorithm and software for the stochastic optimization of stimuli. Behav Res Methods. 2012; 44: 675–705. https://doi.org/10.3758/s13428-011-0182-9 PMID: 22351612

202. Murawski C, Bossaerts P. How humans solve complex problems: The case of the knapsack problem. Sci Rep. Nature Publishing Group; 2016; 6: 34851. https://doi.org/10.1038/srep34851 PMID: 27713516

203. Graham J. Explaining away differences in moral judgment: Comment on Gray and Keeney (2015). Soc Psychol Personal Sci. 2015; https://doi.org/10.1177/1948550615592242

204. Crone DL, Laham SM. Utilitarian preferences or action preferences? De-confounding action and moral code in sacrificial dilemmas. Pers Individ Dif. 2017; 104: 476–481. https://doi.org/10.1016/j.paid.2016.09.022

205. Kahane G, Wiech K, Shackel N, Farias M, Savulescu J, Tracey I. The neural basis of intuitive and counterintuitive moral judgment. Soc Cogn Affect Neurosci. 2012; 7: 393–402. https://doi.org/10.1093/scan/nsr005 PMID: 21421730

206. Paxton JM, Bruni T, Greene JD. Are "counter-intuitive" deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). Soc Cogn Affect Neurosci. 2014; 9: 1368–71. https://doi.org/10.1093/scan/nst102 PMID: 23887818

207. Firestone C, Scholl BJ. Enhanced visual awareness for morality and pajamas? Perception vs. memory in "top-down" effects. Cognition. 2014; https://doi.org/10.1016/j.cognition.2014.10.014 PMID: 25547483

208. van Casteren M, Davis MH. Match: A program to assist in matching the conditions of factorial experiments. Behav Res Methods. 2007; 39: 973–978. https://doi.org/10.3758/BF03192992 PMID: 18183914

209. Huber S, Dietrich JF, Nagengast B, Moeller K. Using propensity score matching to construct experimental stimuli. Behav Res Methods. 2017; 49: 1107–1119. https://doi.org/10.3758/s13428-016-0771-8 PMID: 27421975

210. Constantinescu AC, Wolters M, Moore AB, MacPherson SE. A cluster-based approach to selecting representative stimuli from the International Affective Picture System (IAPS) database. Behav Res Methods. 2017; 49: 896–912. https://doi.org/10.3758/s13428-016-0750-0 PMID: 27287449

211. Peirce JW. PsychoPy—Psychophysics software in Python. J Neurosci Methods. 2007; 162: 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017 PMID: 17254636

212. Colden A, Bruder M, Manstead ASR. Human content in affect-inducing stimuli: A secondary analysis of the international affective picture system. Motiv Emot. 2008; 32: 260–269. https://doi.org/10.1007/s11031-008-9107-z

213. Delplanque S N'diaye K, Scherer KR, Grandjean D. Spatial frequencies or emotional effects? A systematic measure of spatial frequencies for IAPS pictures by a discrete wavelet analysis. J Neurosci Methods. 2007; 165: 144–150. https://doi.org/10.1016/j.jneumeth.2007.05.030 PMID: 17629569

214. Rust NC, Movshon JA. In praise of artifice. Nat Neurosci. 2005; 8: 1647–1650. https://doi.org/10.1038/nn1606 PMID: 16306892

215. Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW. Controlling low-level image properties: The SHINE toolbox. Behav Res Methods. 2010; 42: 671–684. https://doi.org/10.3758/BRM.42.3.671 PMID: 20805589

216. Liu M, Breuel T, Kautz J. Unsupervised image-to-image translation networks. Adv Neural Inf Process Syst. 2017

217. Zhu J, Park T, Isola P, Efros AA. Unpaired image-to-mage translation using cycle-consistent adversarial networks. arXiv Prepr. 2017; https://doi.org/10.1109/ICCV.2017.244

218. Lakens D, Fockenberg DA, Lemmens KPH, Ham J, Midden CJH. Brightness differences influence the evaluation of affective pictures. Cogn Emot. 2013; 27: 1225–1246. https://doi.org/10.1080/02699931.2013.781501 PMID: 23639173

219. Salerno JM. Seeing red: Disgust reactions to gruesome photographs in color (but not in black and white) increase convictions. Psychol Public Policy, Law. 2017; https://doi.org/10.1037/law0000122

220. Laham SM, Alter AL, Goodwin GP. Easy on the mind, easy on the wrongdoer: Discrepantly fluent violations are deemed less morally wrong. Cognition. 2009; 112: 462–466. https://doi.org/10.1016/j.cognition.2009.06.001 PMID: 19573863

221. Sherman GD, Clore GL. The color of sin: White and black are perceptual symbols of moral purity and pollution. Psychol Sci. 2009; 20: 1019–1025. https://doi.org/10.1111/j.1467-9280.2009.02403.x PMID: 19619180

222. Zarkadi T, Schnall S. "Black and White" thinking: Visual contrast polarizes moral judgment. J Exp Soc Psychol. 2013; 49: 355–359. https://doi.org/10.1016/j.jesp.2012.11.012

223. Fincher KM, Tetlock PE. Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. J Exp Psychol Gen. 2016; 145: 131–146. https://doi.org/10.1037/xge0000132 PMID: 27045281

224. Kotabe HP, Kardan O, Berman MG. The order of disorder: Deconstructing visual disorder and its effect on rule-breaking. J Exp Psychol Gen. 2016; https://doi.org/10.1037/xge0000240 PMID: 27736133

225. Gan T, Fang W, Ge L. Colours' impact on morality: Evidence from event-related potentials. Sci Rep. 2016; 6: 38373. https://doi.org/10.1038/srep38373 PMID: 28004749

226. Poldrack RA, Barch DM, Mitchell JP, Wager TD, Wagner AD, Devlin JT, et al. Toward open sharing of task-based fMRI data: the OpenfMRI project. Front Neuroinform. 2013; 7: 12. https://doi.org/10.3389/fninf.2013.00012 PMID: 23847528

227. Vinciarelli A, Pantic M, Heylen D, Pelachaud C, Poggi I, D'Errico F, et al. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. IEEE Trans Affect Comput. 2012; 3: 69–87. https://doi.org/10.1109/T-AFFC.2011.27

228. Zhang Z, Luo P, Loy CC, Tang X. Learning social relation traits from face images. Proc IEEE Int Conf Comput Vis. 2016;11–18–Dece: 3631–3639. https://doi.org/10.1109/ICCV.2015.414

229. Wallach W, Allen C. Moral machines: Teaching robots right from wrong. Oxford: Oxford University Press; 2009.

230. Conitzer V, Sinnott-Armstrong W, Borg JS, Deng Y, Kramer M. Moral decision making frameworks for artificial intelligence. Association for the Advancement of Artificial Intelligence. 2017.

231. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically aligned design: A vision for prioritizing wellbeing with artificial intelligence and autonomous systems [Internet]. 2016. Available: http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf

232. Creative Commons. State of the Commons [Internet]. 2015. Available: https://stateof.creativecommons.org/2015/sotc2015.pdf

233. Levay KE, Freese J, Druckman JN. The demographic and political composition of Mechanical Turk samples. SAGE Open. 2016; 6. https://doi.org/10.1177/2158244016636433

234. Fiske AP. Metarelational models: Configurations of social relationships. Eur J Soc Psychol. 2012; 42: 2–18. https://doi.org/10.1002/ejsp.847

235. Schuster S, Krishna R, Chang A, Fei-Fei L, Manning CD. Generating semantically precise scene graphs from textual descriptions for improved image retrieval. Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal; 2015. pp. 70–80. Available: http://www.emnlp2015.org/proceedings/VL/pdf/VL12.pdf

236. Torralba A, Efros AA. Unbiased look at dataset bias. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE; 2011. pp. 1521–1528. https://doi.org/10.1109/CVPR.2011.5995347

237. Ferraro F, Mostafazadeh N, Huang T-H, Vanderwende L, Devlin J, Galley M, et al. A survey of current datasets for vision and language research. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics; 2015. pp. 207–213. Available: http://aclweb.org/anthology/D15-1021

238. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci. 2010; 33: 1–23. https://doi.org/10.1017/S0140525X09991105

239. Miles A, Vaisey S. Morality and politics: Comparing alternate theories. Soc Sci Res. Elsevier Inc.; 2014; 53: 252–269. https://doi.org/10.1016/j.ssresearch.2015.06.002 PMID: 26188452

240. Landy JF. Representations of moral violations: Category members and associated features. Judgm Decis Mak. 2016; 11: 496–508.

241. Kugler M, Jost JT, Noorbaloochi S. Another look at Moral Foundations Theory: Do Authoritarianism and Social Dominance Orientation explain liberal-conservative differences in "moral" intuitions? Soc Justice Res. 2014; 27: 413–431. https://doi.org/10.1007/s11211-014-0223-5

242. Sinn JS, Hayes MW. Replacing the Moral Foundations: An evolutionary-coalitional theory of liberal-conservative differences. Polit Psychol. 2016; xx. https://doi.org/10.1111/pops.12361

243. Landy JF, Bartels DM. Inductive ethics: A bottom-up taxonomy of the moral domain. Annual Meeting of the Cognitive Science Society. Philadelphia, PA; 2016. pp. 2303–2308.

244. Skitka LJ. The psychology of moral conviction. Soc Personal Psychol Compass. 2010; 4: 267–281. https://doi.org/10.1111/j.1751-9004.2010.00254.x

245. Baron J, Spranca M. Protected values. Organ Behav Hum Decis Process. 1997; 70: 1–16. https://doi.org/10.1006/obhd.1997.2690

246. Tetlock PE. Thinking the unthinkable: Sacred values and taboo cognitions. Trends Cogn Sci. 2003; 7: 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9 PMID: 12860191

247. Gantman AP, Van Bavel JJ. Moral perception. Trends Cogn Sci. 2015; 19: 631–633. https://doi.org/10.1016/j.tics.2015.08.004 PMID: 26440123