NeuroImage 159 (2017) 371–387

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Illuminating the conceptual structure of the space of moral violations with searchlight representational similarity analysis

E.A. Wasserman [a,*], A. Chakroff [a], R. Saxe [b], L. Young [a]

[a] Dept. of Psychology, Boston College, Chestnut Hill, MA, United States
[b] Dept. of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States

ARTICLE INFO

Keywords:
Moral psychology
Representational similarity analysis
fMRI
Social neuroscience

ABSTRACT

Characterizing how representations of moral violations are organized, cognitively and neurally, is central to understanding how people conceive and judge them. Past work has identified brain regions that represent morally relevant features and distinguish moral domains, but has not yet advanced a broader account of where and on what basis neural representations of moral violations are organized. With searchlight representational similarity analysis, we investigate where category membership drives similarity in neural patterns during moral judgment of violations from two key moral domains: Harm and Purity. Representations converge across domains in a network of regions resembling the mentalizing network. However, Harm and Purity violation representations respectively converge in different regions: precuneus (PC) and left inferior frontal gyrus (LIFG). Examining substructure within moral domains, Harm violations converge in PC regardless of subdomain (physical harms, psychological harms), while Purity subdomains (pathogen-related violations, sex-related violations) converge in distinct sets of regions – mirroring a dissociation observed in principal-component analysis of behavioral data. Further, we find initial evidence for representation of morally relevant features within these two domain-encoding regions. The present analyses offer a case study for understanding how organization within the complex conceptual space of moral violations is reflected in the organization of neural patterns across the cortex.

1. Introduction

Judging an act as "morally wrong" may subjectively feel easy and instinctive; yet, underlying each judgment may be a complex, feature-rich representation of the act committed. A wrong act may take many physical forms, from pushing a button to pushing a man off a bridge (Greene et al., 2009), from a mere spoken word (Helwig et al., 2001) to a violent stabbing (Cushman et al., 2012). The victim may be another person or the violator themselves (Chakroff et al., 2013). Moral judgments may demand mental state representations: was the actor internally or externally motivated (Chakroff and Young, 2015)? Did she do it on purpose (Young et al., 2007)? At a higher level, the act may be represented as an instance of a more abstract conceptual category, such as 'harm-based' or 'purity-based' violations, and judged accordingly (Graham et al., 2012; Dungan and Young, 2012; Chakroff et al., 2016b).

Understanding the organization of these representations is critical to understanding how humans conceive of and reason about morally charged acts. Indeed, a long tradition of moral psychological work has sought to answer questions of organization: on what basis can moral acts be grouped? Turiel's classic Domain Theory sought to draw a boundary separating *morals* from *conventions*, on the grounds that morals are generalizable: a moral violation is wrong everywhere and always, even if it is socially condoned (Turiel, 1983). Moreover, moral violations are intrinsically *harmful*, unlike norm violations, which may be merely awkward or improper. With a similar goal, Nichols (2002) separates moral from conventional by arguing that morals are "norms with feeling", defining moral violations as conventional violations accompanied by an affective response. Beyond circumscribing the moral sphere, the problem of organizing morals *within* the sphere has been addressed by Moral Foundations Theory (Haidt et al., 1993; Graham et al., 2012), which argues that morals fall into five principal domains, each characterized by a specific value and its antithesis (loyalty/disloyalty, fairness/cheating, authority/rebellion, purity/impurity, or care/harm).

To translate this question of structure among moral representations into the neural realm, we reframe it in terms of hypotheses about two basic organizing principles: similarity and hierarchy. *Similarity* among representations can reveal basic clustering structure within the space of violations, while assessing *hierarchy* can illuminate how the mind nests

* Corresponding author.
E-mail address: emily.wasserman@bc.edu (E.A. Wasserman).

http://dx.doi.org/10.1016/j.neuroimage.2017.07.043
Received 30 March 2017; Received in revised form 26 June 2017; Accepted 19 July 2017
Available online 22 July 2017

similarity-based clusters to achieve balance between structural parsimony and complexity. We use searchlight representational similarity analysis (RSA) to test a particular model of organization, based on a two-domain model derived from past work (Dungan and Young, 2012; Chakroff, 2015; Chakroff et al., 2016a, 2016b), as a case study to investigate how experimentally determined similarity and hierarchy manifest in converging neural representations across the cortex. Further, in exploratory analyses, we examine representational similarity based on a limited set of psychologically plausible features, as a first effort to determine whether morally relevant features are also being represented in the cortical areas most responsible for representing moral-violation concepts.

As in much RSA work, we employ stimuli that have been structured *a priori*, into two moral domains (Harm and Purity) and four moral subdomains. This method may be seen as analogous to the use of supervised learning models (versus unsupervised models) in data analysis. While we cannot directly assess how the brain *naturally* organizes its representations when encountering unstructured sets of violations, we can assess whether and where it is able to replicate a predefined organizational structure.

1.1. Neural representations of violations

Previous neuroscientific work on morality has largely addressed questions of *content* – where morally relevant features are processed – rather than *structure*. For example, this work has found that the ventromedial prefrontal cortex represents social-emotional value for moral judgment (Koenigs et al., 2007; Shenhav and Greene, 2014) and that the right temporoparietal junction (RTPJ) represents and integrates mental state information for moral judgment (Young and Saxe, 2008; Young et al., 2007). Different affective responses to violations – e.g., moral disgust elicited by impure acts versus indignation elicited by harmful acts – are reflected in BOLD activation differences in various brain regions, including bilateral inferior frontal gyri (Moll et al., 2002). To the extent that this work examines structure, it has taken a univariate functional-mapping approach, identifying regions that respond preferentially to violations of a certain type to argue for the functional coherence of certain groups of moral violations. The impure versus harmful distinction mentioned above, when framed as a distinction between the conceptual domains of Purity and Harm themselves rather than between their associated affective states, is reflected in BOLD differences in whole brain and region of interest (ROI) analyses (Parkinson et al., 2011; Borg et al., 2008; Chakroff et al., 2016a).

This approach answers a useful question – which regions are engaged more during the processing of a given violation type – but does not address the question of which regions, if any, show convergence of multivoxel patterns for violations of that type. Theoretically, pattern representations of a certain type of violation could all resemble one another in a given region without that region showing any preferential BOLD response to those violations, and conversely, a higher BOLD signal does not guarantee similarity of the underlying patterns. More recent work has taken a first step toward representational similarity hypotheses by investigating how morally relevant distinctions are reflected in multivariate pattern differences within neural regions. For example, multivoxel pattern classifiers (MVPA) have identified a binary intentional-accidental distinction in RTPJ's voxel patterns (Koster-Hale et al., 2013; Chakroff et al., 2016a), implying some degree of representational similarity within each violation type. Yet a comprehensive account of how moral-violation pattern representations converge differentially across the whole brain – a cortical map of moral-conceptual organization – remains to be discovered.

In other domains, the representational similarity approach has been highly successful in revealing cognitive organization across broad areas of cortex by characterizing the relationships between multivariate neural representations (Kriegeskorte et al., 2008; Davis and Poldrack, 2014). RSA and related methods have been fruitful in characterizing the structure of the space of physical object representations (Kriegeskorte et al., 2008), semantic representations (Handjaras et al., 2016; Huth et al., 2012), and lexical representations (Su et al., 2012) – as well as the key features driving structural organization. Yet their application to conceptual spaces involving social content is so far limited. For example, RSA has been employed to uncover dimensions of social-information representation within the mentalizing network (Tamir et al., 2015; Chavez and Heatherton, 2015) and belief attributions across the cortex (Leshinskaya et al., 2017). The moral representations tested here, as a subclass of social representations, thus present a novel challenge and opportunity for representational similarity analysis. If representational similarity can shed light on the neural and cognitive organization of objects, words, and concepts, can it do the same for moral violations?

2. Method

2.1. Participants (fMRI)

Forty-five adults participated in the study for payment. Six were excluded for missing or improperly recorded data, for a total sample size of 39 ($N = 10$ female), mean age 30.33 years. Of these, 14 ($N = 2$ female) were diagnosed with Autism Spectrum Disorder by a licensed clinician, based on Autism Quotient (AQ) scores. No group differences in RSA maps were found (see Supplementary Materials). All participants were right-handed native English speakers with normal or corrected-to-normal vision, and gave informed consent in line with institutional review procedure at MIT. Subsets of the data collected for this study have been previously reported in two published articles (Koster-Hale et al., 2013; Chakroff et al., 2016a); the sample reported here constitutes the full set of complete data available at the time of analysis.

2.2. Experimental design (fMRI)

Stimuli for the moral judgment task consisted of 60 written scenarios, of which 48 were moral-violation scenarios and 12 neutral social scenarios (for the full text of all scenarios, see Appendix A of the Supplementary Material). Within the moral scenarios, 24 depicted harm-domain violations, of which 12 were physical (e.g., poisoning) and 12 psychological (e.g., insults) violations. The other 24 depicted purity-domain violations, of which 12 were pathogen-based (e.g., drinking human blood) and 12 incest-based (e.g., consensual sex with an adult sibling) violations. Our choice of these two particular domains, as opposed to the five- or seven-domain Moral Foundations framework (Haidt et al., 1993; Graham et al., 2012), was motivated by the large body of existing literature that focuses on the harm-purity distinction in both psychological and neural responses (e.g., Chakroff and Young, 2015; Parkinson et al., 2011), and by our own past work suggesting that a two-type model captures most variation across moral judgments of actions (Dungan and Young, 2012). Each participant viewed all 60 scenarios in pseudorandom order across 6 runs, with condition order counterbalanced across runs and participants; no condition was shown twice in a row.

Each scenario was split into four serially presented segments - Background (6 s), Action (4 s), Outcome (4 s), and Intent (4 s; Fig. 1). In a subsequent 4-s window, participants judged the moral wrongness of the scenario on a scale from 1 (“not at all morally wrong”) to 4 (“very morally wrong”) using a button box. In the Intent segment, information was presented which either specified that the act was committed intentionally, with full knowledge (e.g., you knew that your sexual partner was your sibling and decided to commit incest anyway), or that the act was committed accidentally, in ignorance (e.g., your sexual partner was a long-lost sibling you didn't recognize). Intent was described with three categories of mental-state verbs: knowledge (knew/thought), realization (realized/discovered), and perception (saw/noticed). Half of the scenarios were randomly presented as intentional and half as accidental. No participant saw both versions of the same scenario.

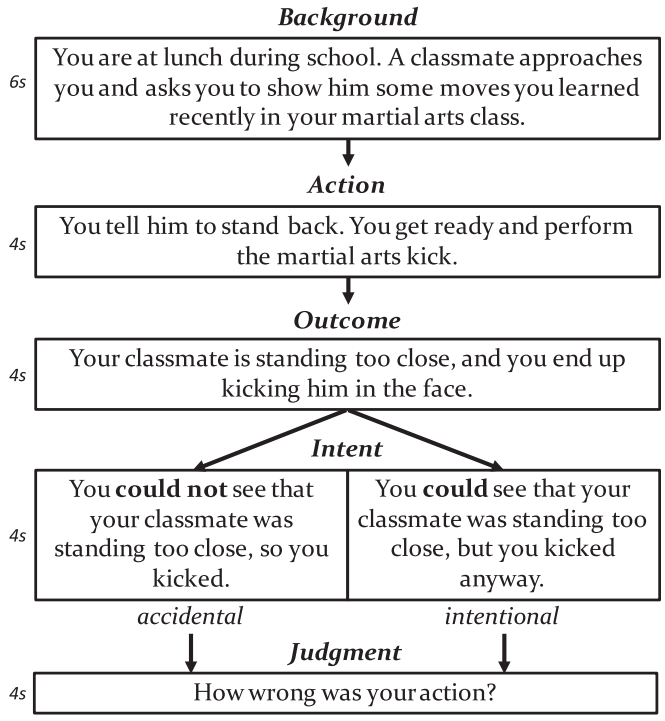

Fig. 1. Illustration of MRI task procedure using a sample scenario from the Physical Harm condition. Each segment was presented separately on a black screen with white text. In the Judgment segment, ratings from 1 to 4 were given using a button box with four buttons.

Prior to the moral judgment task, participants performed a Theory of Mind (false belief) task (Dodell-Feder et al., 2010). Results of this task are not of interest here and have been reported elsewhere (Koster-Hale et al., 2013; Chakroff et al., 2016a).

2.3. fMRI data acquisition and processing

Participants were scanned in a Siemens Trio 3T scanner at MIT's Brain & Cognitive Sciences building in Cambridge, MA, with near-axial slices at 4 mm and in-plane slices at 3x3-mm resolution, at TR = 2 s, TE = 40 ms, with a flip angle of 90°. MATLAB R2015a, SPM12b and custom MATLAB scripts (www.github.com/lypsychlab/RSA) were used to process and analyze all MRI data post-scanner. High-resolution T1-weighted structural images were coregistered and normalized to MNI space, and the parameters used to normalize the functional images. Functional images were slice-time corrected, realigned, and smoothed with a Gaussian kernel of 8-mm FWHM (for univariate analyses) or left unsmoothed (for multivariate analyses). Motion and spike artifact correction was performed with the ART toolbox (www.nitrc.org/projects/artifact_detect). All MRI data are publicly available on OpenFMRI (https://openfmri.org/dataset/ds000212).

After preprocessing, data were modeled either condition-wise (for univariate analysis) or item-wise with an event-related design, with each scenario modeled as a 22-s event beginning with the onset of text presentation. The standard condition-wise modeling procedure included 10 condition regressors (2 intent levels x (4 moral + 1 neutral conditions)), whereas the item-wise procedure modeled each scenario with its own regressor. This latter procedure yielded a single beta image per scenario for a total of 60 beta images per participant, which were used for all subsequent multivariate analyses. ROI masks were constructed with MarsBaR 0.43 (http://marsbar.sourceforge.net).

2.4. Representational similarity analysis

For each representational similarity analysis (RSA), a linear model was constructed with matrix regressors in the lower triangle of 60 × 60 item space (for analyses including neutral items) or 48 × 48 item space (for those including moral items only), in which the (*i*, *j*) entry represents the cross-correlation of voxel vectors for the *i*th and *j*th items (Fig. 2). These regressors were based on either categorical experimental factors, such as domains, or continuous behavioral factors, such as the rated disgustingness of a scenario. Categorical regressors represented hypothesized similarity between pairs of items with 1, i.e., assumed maximal similarity, or 0, i.e., assumed no relationship, to yield similarity-based categorical groupings. For continuous regressors constructed from feature variables, the similarity between mean item ratings, scaled and weighted by the higher-rated of the two items, was computed for each pair of items. We chose to weight similarity ratings to better model the hypothesis that a region which truly represents a certain factor, such as disgust, will encode two items that are similar *and* high on that factor (e.g., two highly disgusting items) more similarly than two items that are similar but low on that factor (e.g., two equally non-disgusting items).

In a searchlight procedure, the center of a 3 × 3 × 3-voxel sphere was moved throughout a canonical grey-matter cortical mask. Within each sphere, voxel patterns for individual scenarios were correlated, and correlations Z-transformed, to produce an empirical similarity matrix of pairwise neural similarities, which was then modeled with our RSA matrix regressors. Separately for each regressor in the model, the modeled parameter value was assigned to the sphere's center voxel and written to an empty template of the same dimensions as the functional images, resulting in a "similarity map" for that regressor. All similarity maps were thresholded using maximum cluster extent values obtained with AFNI 3dClustSim v16.2.02 (voxelwise $p < 0.001$, clusterwise threshold $p < 0.05$, 10,000 iterations). Note that this is a bug-fixed version of 3dClustSim (c.f., Eklund et al., 2016)

Though correlation distance has been recommended over other distance measures (e.g., Euclidean distance) when assessing similarity of representational patterns, it may also be sensitive to mean condition differences in BOLD activation (Walther et al., 2015). As a sanity check, we also remodeled the neural similarities, removing mean signal across within-sphere voxels for each beta image prior to correlation. Demeaning the signal did not significantly change observed clusters.

2.5. Feature-variable collection methods

Behavioral data were collected using Qualtrics and Amazon Mechanical Turk. All participants recruited via Mechanical Turk were located in the United States and had an approval rating of 95% or higher. Each feature variable was collected in a separate sample of 50 participants, except for moral wrongness, which was collected in-scanner. Features collected include ratings of each action's disgustingness, degree of person and situation attribution, weirdness, rationality, badness for self or others, and the extent to which the action made participants think about the physical environment, actions and behaviors, or thoughts and desires (i.e., minds). Features were chosen based on measures previously studied in the context of moral judgments of harm- and purity-based violations – specifically, those studied as potential explanatory factors behind differences in judgment of violations across domains (Chakroff et al., 2013; Chakroff and Young, 2015; Young et al., 2007; Dungan, in prep). See Appendix B for scales and wording of behavioral measures.

Each participant viewed a randomly ordered 12-item random subset of the moral scenarios detailed above, and rated each on a discrete scale. Scenarios were stripped of intent information prior to presentation. Participants gave consent in accordance with Boston College's institutionally approved procedures, and were compensated at an approximate rate of $5/hour, in line with standard online-participation compensation rates.

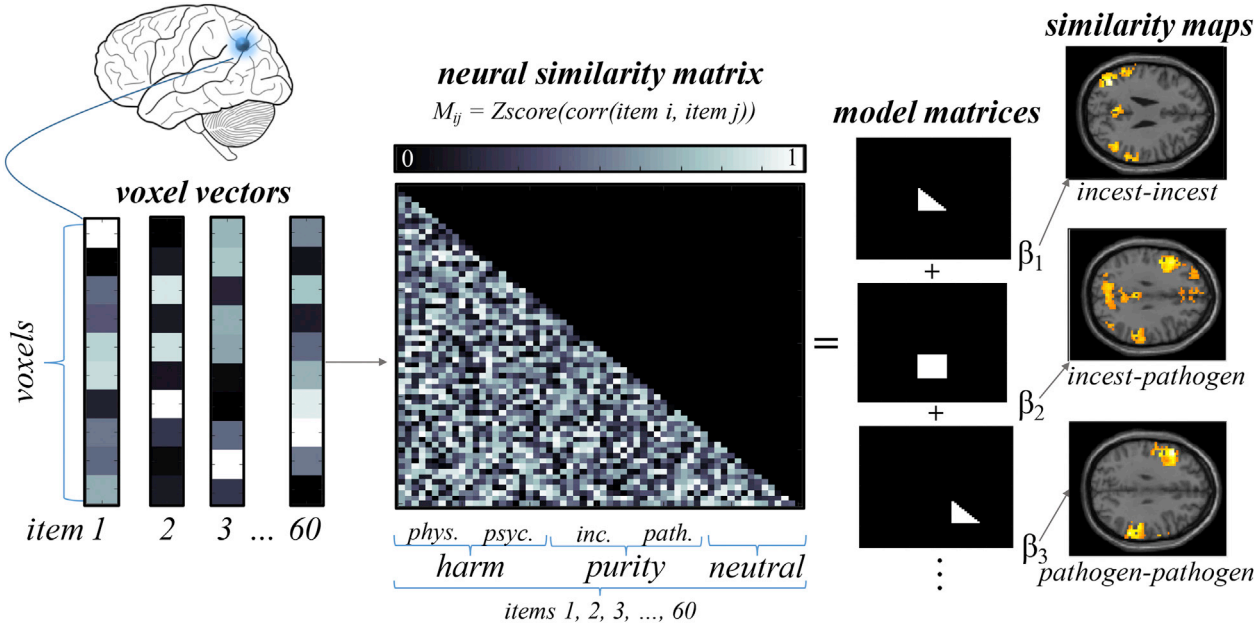

Fig. 2. Illustration of searchlight RSA method. A vector v of per-voxel beta values is extracted from voxels within a sphere, for each item. Item vectors v_1, v_2, …, v_{60} are cross-correlated to yield a neural correlation matrix M, which is modeled linearly with similarity regressors, each representing a hypothesized similarity relationship. Each regressor's corresponding beta weight β is mapped onto the sphere's central voxel. Moving the searchlight sphere throughout the cortex forms a whole-brain statistical image which represents the contribution of that similarity regressor to the model at each cortical location. Lower-triangle similarity matrices are displayed. The full model includes 5 regressors not pictured here (see Results 4.2.1).

3. Results

3.1. Behavioral results

Means and standard deviations by condition for each variable are shown in Table 1. Summary bar plots, including Tukey post-hoc test comparisons for pairwise condition differences, are shown in S.I. Fig. 1. Variables were scaled and centered, and entered into a principal components analysis with prcomp () in R. The first two components accounted for 52.77% of the variance, with marginal (≅ 1%) increases in explained variance after the first 5 components (S.I Fig. 2). A biplot of the items along the first two components is shown in Fig. 3.

In an attempt to recover the experimental categories using principal component scores alone, we performed K-means clustering with 2, 3, and 4 centroids on scores from the first 5 principal components, using R's default Hartigan-Wong algorithm with 100 iterations and 10 random starting positions (Table 2). Misclassification is defined as placement of a category's item within a cluster primarily composed of items from a different category. At $k = 2$, the algorithm recovered the harm-purity distinction almost perfectly, misclassifying 1/24 (4.2%) of harm items. At $k = 3$, the clusters closely reflected groupings for pathogen, incest, and harm items. Misclassification rates were low among purity items: 1/12 (8.4%) and 0/12 for pathogen and incest respectively. However, 8/24 (33.3%) of harm items were misclassified. This classification difference across domains was significant: though the clustering algorithm consistently identified purity items as similar to one another based on their principal component scores, it did not do the same for harm items (X^2 (1) = 6.70, $p = 0.03$; Yates continuity-corrected).

The clustering solution at $k = 4$ also consistently distinguished incest and pathogen subdomains, placing 11/12 items from each in clusters 2 and 4 respectively. Notably, while incest and pathogen items (i.e., purity) could be reliably divided from each other, the same was not true of harm items: both clusters 1 and 3 contained an even mix of physical and psychological harm items, at 4:4 and 8:7 ratios respectively (Fig. 3; Table 2).

3.2. Neural results

Our categorical RSA model was fit in 60 × 60 item space, comprising both moral scenarios (48/60) and non-moral social scenarios (12/60). The model included 8 regressors, each capturing a distinct subspace within the full space of item similarities: Incest-Incest, Incest-Pathogen, Pathogen-Pathogen, Physical-Physical harm, Physical-Psychological harm, Psychological-Psychological harm, Harm-Purity, and Neutral-Neutral[1] similarity. For a given regressor in the model, a higher beta weight means higher neural similarities for those particular item pairs; i.e., a high beta weight assigned to the Incest-Incest regressor in a sphere means that the neural representations of incest items were highly similar to one another in that sphere. Searchlight maps for each regressor are shown in Fig. 4. A statistical map and average map of R^2 values for the model across the cortex is shown in S.I. Fig. 4. (per-subject peak values in Supplementary Table 1).

3.2.1. Representational similarity across moral domains

The Harm-Purity similarity matrix captures regions where moral item patterns resemble one another across domain boundaries – i.e., where features common to moral items regardless of their domain are likely to be represented. These regions comprised the superior and inferior temporal lobes, precuneus, right supramarginal gyrus and caudate, and left angular gyrus, as well as a small cluster in medial prefrontal cortex (Table 3; Fig. 4). We also assessed overall similarity within the set of moral items, including similarity between pairs of items from the same domain, by creating a conjunction of the Harm-Purity map with the Harm and Purity similarity maps (see 3.2.2). This yielded a similar pattern of results to the Harm-Purity analysis alone (S.I. Fig. 3).

3.2.2. Representational similarity within moral domains

At the next level of the categorical hierarchy, we investigated the similarity of neural representations for violations within each moral domain (Harm and Purity), irrespective of subdomains. For each domain, all relevant within-subdomain similarity maps - incest-incest, incest-pathogen, and pathogen-pathogen for Purity, and psychological-psychological harm, psychological-physical harm, and physical-physical harm for Harm - were linearly combined into a single similarity map.

[1] No significant clusters appeared on the neutral-neutral similarity map; thus, we do not discuss it further.

Table 1
Means and standard deviations for feature variables.

Variable	Harm				Purity			
	Physical		Psychological		Incest		Pathogen	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Harm to others	6.29	0.51	4.89	0.67	4.51	0.64	2.83	0.82
Harm to self	4.22	1.13	3.99	1.03	5.47	0.63	5.71	0.99
Situation attribution	4.24	0.85	4.70	0.86	4.06	0.61	3.59	0.91
Person attribution	4.24	0.74	4.59	0.59	4.45	0.58	4.02	0.57
Disgust	2.42	0.56	1.85	0.43	3.34	0.35	3.69	0.22
Irrationality	2.93	0.58	3.28	0.60	2.32	0.49	2.14	0.74
Weirdness	4.35	1.05	4.25	1.27	5.78	0.48	5.20	0.83
Wrongness (acc.)	1.70	0.40	1.71	0.41	1.93	0.36	1.47	0.28
Wrongness (int.)	3.67	0.26	3.17	0.46	3.50	0.28	2.39	0.26
Attention to environment	4.85	0.67	3.77	0.60	3.31	0.33	4.66	0.62
Attention to behaviors	5.36	0.44	5.36	0.62	5.51	0.46	5.11	0.76
Attention to minds	3.43	0.39	4.11	0.76	4.88	0.54	3.53	0.88

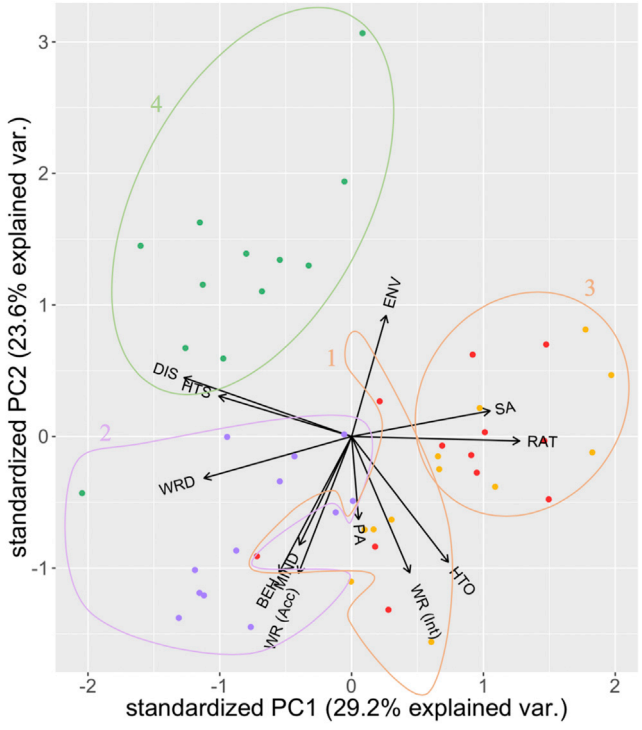

Fig. 3. Biplot showing loadings for each of the 48 moral items on principal components 1 & 2, with vectors corresponding to feature variables. Enclosed areas indicate clusters derived from K-means clustering at K = 4 (see Table 2), and are colored according to their item composition. Abbreviations for feature variables are: ENV (attention to environment); SA (situation attribution); RAT (rationality); HTO (harmfulness to others); WR (Int) (wrongness when intentional); PA (person attribution); WR (Acc) (wrongness when accidental); MIND (attention to minds); BEH (attention to behaviors); WRD (weirdness); HTS (harmfulness to self); DIS (disgust).

Comparing the two domain maps, a dissociation emerged between representational similarity for Harm, in precuneus (PC) and both temporoparietal junctions (Table 3), and for Purity, in an extended network of cortical regions, including left inferior frontal gyrus (LIFG), left-lateralized regions of temporal and parietal cortices, and right angular gyrus, as well as subcortical structures, including midline thalamus and bilateral insula (Table 3). As the respective regions with the largest voxel extent[2] and highest peak intensity (t = 7.21, 7.80), PC and LIFG were chosen for ROI-based analyses (see *d*). Note that while the peak voxel of this ROI fell within LIFG, contiguous voxels extended into other left-hemisphere regions, including parts of superior and inferior temporal gyri.

[2] The Purity map also shows a large region (Table 2) which extends across many functionally and anatomically distinct regions, including portions of the left angular gyrus, superior, middle, and inferior temporal lobe, fusiform gyrus, and cerebellum. For our purposes, we consider this region to be a set of distinct regions, rather than a single functional ROI.

Table 2
Clusters resulting from K-means analysis (K = 2; 3; 4) of principal component loadings.

Cluster	Harm		Purity	
	Physical	Psychological	Incest	Pathogen
K = 2				
1	11	12	0	0
2	1	0	12	12
K = 3				
1	9	7	0	0
2	0	0	0	11
3	3	5	12	1
K = 4				
1	4	4	1	0
2	0	1	11	1
3	8	7	0	0
4	0	0	0	11

3.2.3. *Representational similarity within and across moral subdomains*

We investigated the finest level of experimentally defined categorical structure with four within-subdomain similarity maps and two cross-subdomain similarity maps. Within-subdomain regressors captured similarity between patterns of items in the same subdomain - Physical and Psychological harms, within Harm, and Incest and Pathogen violations, within Purity. Cross-subdomain maps – Incest-Pathogen and Physical-Psychological - captured similarity across items from different subdomains within the same domain.

Physical-harm and Psychological-harm subdomains, as well as Physical-Psychological similarity, were represented in overlapping regions within PC (Table 4; Fig. 4). Bilateral TPJ also appeared on the Physical-Psychological similarity map, and left TPJ only on the Psychological-Psychological map. In contrast, each of the two Purity subdomains showed a distinct representational pattern across the cortex: Incest primarily in bilateral temporal lobes, TPJ, and precuneus, and Pathogen in a broader set of regions including the cuneus, cingulate cortex, bilateral superior parietal cortices, medial prefrontal cortex, and left inferior/middle frontal gyrus (LIFG; Table 5; Fig. 4). LIFG also encoded Incest-Pathogen similarity. Like the Purity similarity map, the Incest-Pathogen similarity map included large regions of high similarity comprised of different yet contiguous functional regions (Table 5).

3.2.4. *Representations of morally relevant features*

Neural representations of moral violations clustered in strikingly different regions depending on moral domain (4.2.2), opening the possibility that these regions represent different types of morally relevant

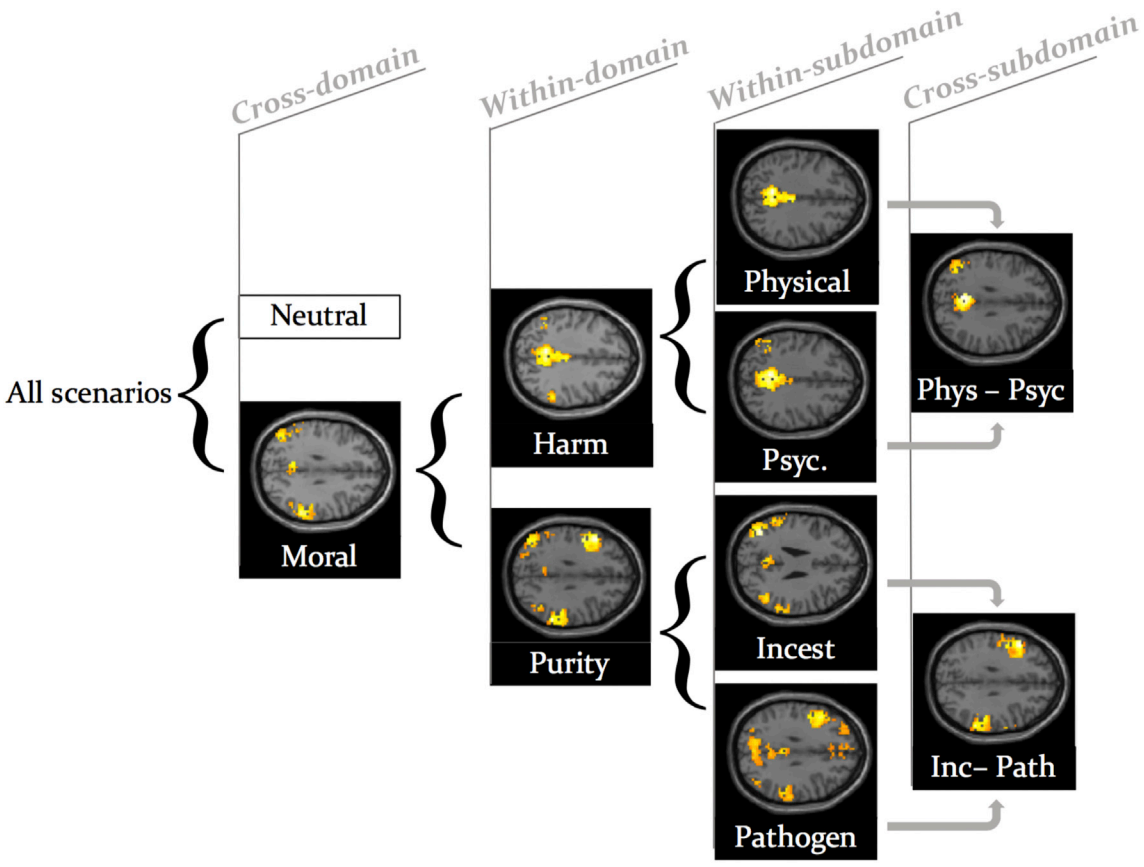

Fig. 4. Similarity maps displaying areas of peak representational similarity at each level of categorical hierarchy within the space of moral item similarities. Maps thresholded using cluster extent thresholds generated by 3dClustSim v16.2.02 (voxelwise $p < 0.001$, clusterwise $p < 0.05$). MNI coordinates: z = 31 (Moral), 37 (Harm), 31 (Purity), 37 (Physical), 31 (Psychological), 37 (Phys – Psyc), 28 (Incest), 31 (Pathogen), 31 (Inc – Path).

Table 3
MNI peak coordinates from maps for within-domain and cross-domain similarity.

Region	*Hemisphere*	*x*	*y*	*z*	*k*
Harm-purity similarity					
Precuneus	R/L	−9	−55	31	416
Superior temporal gyrus	L	−51	−16	−8	396
Supramarginal gyrus	R	57	−43	37	294
Caudate cortex	R	45	8	4	267
Angular gyrus	L	−45	−70	31	245
Inferior temporal lobe	L	−39	−43	−14	163
Middle temporal gyrus	R	57	−13	−8	101
Medial prefrontal cortex	R/L	0	41	−8	63
Harm-harm similarity					
Precuneus	R/L	−6	−52	37	541
Temporoparietal junction	L	−45	−55	37	131
Temporoparietal junction	R	54	−43	37	65
Purity-purity similarity					
Inf. temporal/Fusiform/Angular gyrus	L	−42	−43	−14	1727
Inferior frontal gyrus	L	−48	17	31	648
Inferior parietal lobule/TPJ	R	60	−43	31	331
Angular gyrus	R	45	−64	28	90
Insula	R	45	5	7	79
Thalamus	R/L	3	−19	2	89
Insula	L	−36	−13	1	65
Cerebellum	R/L	3	−70	−20	62
Supramarginal gyrus	L	−48	−40	25	61

content. To better understand the representation of specific moral-violation features within these regions, an RSA model with similarity regressors for each of the 12 feature variables was fit to neural similarity matrices constructed from cross-correlated vectors of voxels within masked ROIs of contiguous voxels around the peak voxel in PC ($x = -6$, $y = -52$, $z = 37$, $k = 541$) or LIFG ($x = -48$, $y = 17$, $z = 31$, $k = 648$) from the Harm and Purity similarity maps respectively. Mean beta weights with associated 95% confidence bounds for the feature-variable similarity regressors are shown in Fig. 5. All participants were included in this analysis, although the categorical-only RSA model was not significant at $\alpha = 0.05$ for 6 participants (in PC) and 5 participants (in LIFG).

Within PC, beta weights for harm to others, person attribution, attention to environment, and attention to minds were significantly above zero across participants ($M_{HTO} = 0.08$, t (38) = 2.29, $p = 0.03$; $M_{PA} = 0.33$, t (38) = 4.58, $p < 0.001$; $M_{ENV} = 0.37$, t (38) = 7.25, $p < 0.001$; $M_{MIND} = 0.12$, t (38) = 2.26, $p = 0.03$), as was disgust within LIFG ($M_{DISG} = 0.08$, t (38) = 2.43, $p = 0.02$). Additionally, a significant negative weight was associated with disgust within PC ($M_{DISG} = -0.10$, t (38) = 4.75, $p < 0.001$), and with rationality within LIFG ($M_{RAT} = -0.20$, t (38) = 2.44, $p < 0.02$).

As suggested in the principal component analysis (4.1), some significant covariation existed between features (see Supplementary Table 2). To ensure that this covariation did not cause instability in model estimates, the above analysis was iterated 100 times for each participant. Parameter estimates were stable across all iterations within each participant. However, given the presence of covariation, the features isolated above should be interpreted as useful indicator variables, which capture some part of the meaningful variation in neural similarities yet may nevertheless share some of their explanatory power with other conceptually related features.

3.2.5. *Permutation tests within core ROIs*

The results above indicate that the two core domain-encoding regions PC and LIFG may also encode continuous feature information: specifically, harm to others, person attribution, and attention to environment/

minds in PC, and disgust in LIFG. However, as these behavioral regressors also show high within-domain uniform similarity, it is unclear whether this result indicates representation of specific feature information in these ROIs, or only representation of uniform within-domain similarity.

We conducted model comparisons between categorical-only models and models including behavioral variables associated with significant positive effects in the 12-variable regression (see Fig. 5). For each participant, within each ROI, we fit both the categorical RSA model and a model augmented with a single behavioral similarity-matrix regressor. We then compared R^2 between the two models across participants with a paired-sample *T*-test. In both PC and LIFG, augmenting the categorical-only model with one of the selected behavioral regressors resulted in a small but significant R^2 boost (PC: all t (38) $\geq$ 3.97, $p < 0.001$; LIFG: t (38) = 5.14, $p < 0.001$). As these models contained categorical regressors for the category originally used to define each ROI (see 4.2.4), this method allowed us to isolate the contribution of the feature variable beyond the contribution of category information and thereby minimize potential effects of nonindependent sampling (Kriegeskorte et al., 2009) – a risk we consider generally low, given that our stimulus categories were constructed based on past work on moral domains, and thus not designed to group stimuli based on any specific feature or features.

To ensure that these results were due to the behavioral regressor's informational content, rather than the mere presence of another regressor, we conducted 100-iteration permutation tests for each ROI. On every iteration, the behavioral similarity matrix was randomly permuted and the resultant matrix added as a regressor to the categorical-only model. Adding a scrambled regressor significantly improved R^2 compared to the categorical-only model, in 100/100 iterations in both ROIs, for all feature variables tested (PC: all t (38) $\geq$ 2.86, all $p \leq 0.007$; LIFG: all t (38) = 3.32, all $p \leq 0.002$). However, the mean R^2 boost from adding an information-containing behavioral regressor was significantly greater than that from adding a scrambled regressor in 100/100 iterations (PC: all t (38) $\geq$ 3.43, all p $\leq$ 0.001; LIFG: all t (38) $\geq$ 4.24, all $p \leq 0.001$). In sum, the specific structure of the feature-variable similarity matrices was critical to improving model fit in every case, for both ROIs.

4. Discussion

In seeking to map the space of moral violations, we focused on two organizational principles: *similarity*, i.e., which violations are close together, forming a category, and which are distant; and *hierarchy*, i.e., how similarity-based clusters nest within each other. Across the cortex, our searchlight RSA approach found evidence for moral categorization in distinct brain regions at each level of our experimentally defined hierarchy, from the moral space as a whole to moral domains to moral subdomains. Neural representations of violations across both domains (Harm and Purity) converged in a set of regions resembling a social-cognition or default-mode network, comprising right and left TPJ and temporal lobes, precuneus, and vmPFC (Saxe and Kanwisher, 2003; Schurz et al., 2014). At the next level of the hierarchy, representations of items from the two moral domains converged in distinct regions: PC, for harm-based violations, and a network of regions including LIFG, for purity-based violations. These moral domains also proved distinct in their substructures. For subdomains, the finest level of categorization tested in the present work, incest- and pathogen-based purity violations converged in divergent sets of regions, while all harm violations converged in PC, regardless of whether they were physical or psychological.

Our data also hint that within-category similarity may be related to representations of *continuous* features of moral violations. Each of the two domain-encoding regions also represented morally relevant features: harm to others, person attribution, and attention to minds/environment in PC, and disgust in LIFG. While correlational in nature, these data provide initial candidates for features that may drive the organization of moral-violation representations into conceptual categories.

Table 4
MNI peak coordinates from maps for within- and cross-subdomain similarity, within Harm.

Region	*Hemisphere*	*x*	*y*	*z*	*k*
Physical-physical similarity					
Precuneus	R/L	6	−52	37	430
Psychological-psychological similarity					
Precuneus	R/L	−6	−55	31	526
Temporoparietal junction	L	−48	−58	34	144
Physical-psychological similarity					
Precuneus	R/L	−6	−52	37	595
Temporoparietal junction	L	−48	−55	34	212
Temporoparietal junction	R	57	−43	34	56

Table 5
MNI peak coordinates from maps for within- and cross-subdomain similarity, within Purity.

Region	*Hemisphere*	*x*	*y*	*z*	*k*
Incest-incest similarity					
Temporoparietal junction	L	−42	−67	28	512
Middle temporal gyrus	L	−51	−10	−20	437
Precuneus	R/L	−9	−46	43	230
Superior temporal gyrus	R	51	−58	25	172
Supramarginal gyrus	R	60	−43	34	156
Middle temporal gyrus	R	57	−13	−17	147
Pathogen-pathogen similarity					
Cingulate cortex/Cuneus	R/L	6	−37	37	2002
Inferior frontal gyrus	L	−45	14	31	655
Orbitofrontal cortex	R/L	6	44	10	305
Inferior temporal gyrus/Fusiform	L	−42	−49	−14	260
Superior parietal lobule	R	42	−64	52	255
Insula	R	39	2	7	202
Inferior parietal lobule	R	51	−31	28	167
Dorsomedial prefrontal cortex	R/L	6	32	43	119
Incest-pathogen similarity					
Temporoparietal junction/insula	R	60	−43	31	957
Superior temporal gyrus	L	−48	17	31	899
Cerebellum	R	−39	−43	−14	656
Precuneus	R	12	−52	67	276
Caudate cortex	L	−15	8	−5	269
Precuneus	L	−12	−58	58	143
Cuneus	R	15	−70	1	140
Middle occipital gyrus	R	21	−103	7	106
Cingulate cortex	L	−12	−1	43	77
Thalamus	R/L	3	−22	2	74
Temporoparietal junction	L	−51	−40	22	57
Superior frontal gyrus	R	24	50	31	55

4.1. Delineating the moral space

The striking similarity between the network of regions in which cross-domain moral violations are encoded similarly and the ToM network leads us to ask: could the key features distinguishing moral acts as a category from nonmoral acts be features about social agents? Greene and Haidt (2002) identified a similar set of regions across multiple imaging studies of moral judgment, and implicate ToM as a 'likely function' in all of them. Since then, many studies have implicated mental state attribution mediated by the ToM network in moral judgments (Young et al., 2007; Young and Saxe, 2008; Young et al., 2010; for review see Moll and Schulkin, 2009), making it plausible that morally relevant scenarios, as a category, are similar insofar as the social-cognitive processes they elicit are similar. The mere presence of social information is most likely not a critical factor in separating morally relevant from neutral scenarios, as our neutral stimuli were deliberately chosen to contain information about agents. However, if judgments of moral wrongness rely on representing agents' beliefs and intentions, we might expect that judging scenarios perceived as morally wrong recruits these social-cognitive processes *more* than judging neutral scenarios, reflected in higher BOLD signal in ToM regions for moral > neutral scenarios – as is the case in this dataset (Chakroff et al., 2016a; Supplementary Analyses). The convergence of multivariate representations across moral items in these regions lends further support for this account: in these regions, moral

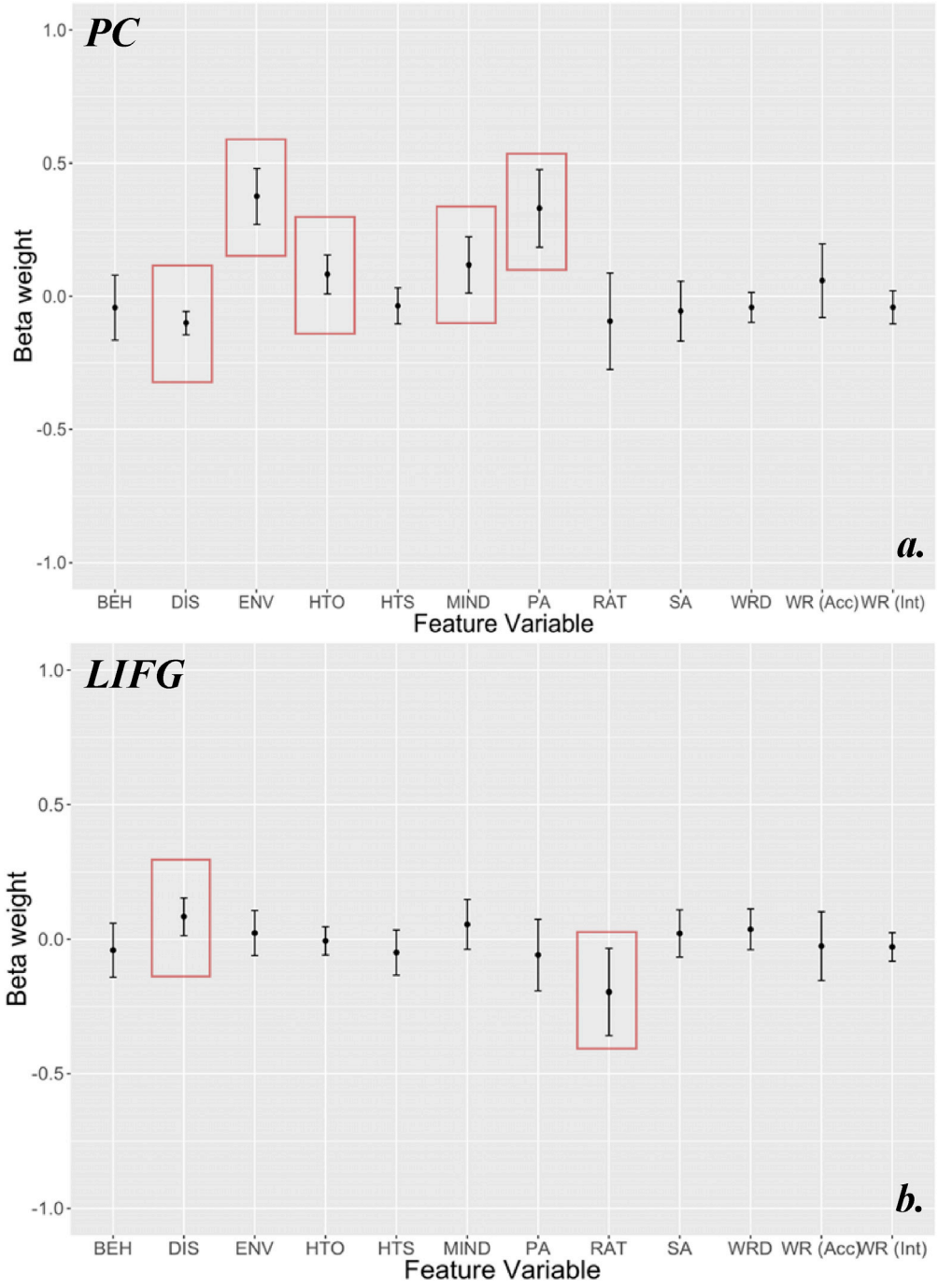

Fig. 5. Mean RSA beta weights and 95% confidence intervals across participants for feature-variable similarity regressors within a) PC and b) LIFG. Highlighted variables include disgust (DIS), attention to environment (ENV), harm to others (HTO), attention to minds (MIND), person attribution (PA), and rationality (RAT). See Fig. 3 & Appendix B for all feature variable abbreviations.

scenarios share not only heightened BOLD signal relative to neutral scenarios, indicating enhanced mental state processing, but also common neural patterns, potentially reflecting common social content. Determining precisely which social features moral scenarios share will require "deconstructing" the construct of ToM into more basic processes – e.g., face processing, action understanding, or representation of others' beliefs – to examine which do, and do not, drive similarity across neural representations of moral scenarios (Schaafsma et al., 2015).

4.2. Two encoding regions for moral domains

Previous work using univariate contrasts has found an association between harm-based violations and PC (Parkinson et al., 2011; Greene et al., 2001; Moll et al., 2002; Heekeren et al., 2005; Chakroff et al., 2016a, [3]) and between purity-based violations and LIFG (Parkinson et al., 2011; Moll et al., 2002, 2005; Borg et al., 2008.; Chakroff et al., 2016a). The present findings strengthen these associations from a different angle, showing that not only do these regions, PC and LIFG, show increased activation for harm and purity respectively, but also they represent information about the categories themselves, in the form of convergent neural representations across items within each domain. Across domains,

[3] Note that this is a subsample of the same data analyzed here.

our results further bolster a harm-purity dissociation observed previously in behavioral (Dungan and Young, 2012; Chakroff et al., 2013, 2016b; Chakroff and Young, 2015; Graham et al., 2011) and neural work (Chakroff et al., 2016a; Parkinson et al., 2011; Borg et al., 2008). Crucially, each domain was primarily encoded within at least one distinct region, suggesting that specific functional regions may handle representation of moral violation categories.

These regions also appear to encode some information about particular features of the moral violations. In LIFG, one part of the connection between features and domain is straightforward: a wealth of previous work links disgust to moral judgment of purity-based violations (Rozin and Fallon, 1987; for reviews see Chapman and Anderson, 2013; Pizarro and Helion, 2011). Our results do not answer whether purity violations are considered immoral *because* they are disgusting, as posited by theories of purity-based moral values as a disease-avoidance mechanism (Curtis et al., 2011). They do, however, support the claim that disgust is a key organizational axis within the moral space, by which gross *and* immoral acts (e.g., sex with a sibling) are separated from immoral but not-gross acts (Dungan and Young, 2015) – a claim further supported by our behavioral results showing that disgust drives organization of violations along the first principal component (Fig. 3).

Similarly, the representation of harm to others in PC is consistent with the definition of harm-based violations. However, the representation of other feature variables in this region is unexpected: why would person attribution, attention to environment, and attention to minds be represented here? A clue comes from functional connectivity work identifying PC as a "hub" region, common to multiple functional networks and thus ideally positioned to integrate disparate sources of information (Buckner et al., 2008, 2009). If representing a scenario of dyadic harm (Gray et al., 2014) depends on combining representations of both agents and the scene itself, PC is one of the most plausible candidate regions to perform that integratory role.

It is also possible that the multifunctional nature of PC (see Cavanna and Trimble, 2006; for a review) makes it difficult to trace a clean connection between the representation of any particular feature and representation of the harm domain with our method. Perhaps none of the features tested is primarily responsible for organizing harm representations in PC, and PC's association with person attribution and attention to environment/minds is unrelated to its role in representing harm violations. Further, the PC ROI defined in our study is relatively large (541 voxels), raising the possibility that what appears to be representation of different features across the entire region is, in reality, representation of each individual feature by distinct functional subregions (Zhang and Li, 2012). Future research will therefore need to use more targeted methods, for example, directly manipulating particular features of harm violations to assess the impact on representation in specific subregions of PC.

Finally, we note that regions in PC appeared across many similarity maps in our study, including the Harm-Purity, Pathogen-Pathogen, and Incest-Pathogen maps. This is congruent with PC's proposed role as a hub region; while the specific content of a simulated moral violation differs drastically across scenarios, the integrational cognition supported by such a hub region – for example, combining the spatial location of a violation with the action itself – could plausibly be similar, regardless of scenario type. Under this interpretation, it is not that PC is preferentially engaged in representing harm-based violations and the features specific to them, but rather that all neural representations of *all* violations tend to converge in PC, and purity-based violations are represented in a broader set of auxiliary regions, including LIFG.

4.3. Unified harm, divergent purity: cross-domain differences in substructure

Physical and psychological harms are tied together in many linguistic metaphors (Semino, 2010; Lakoff and Johnson, 1980), and people may experience the resultant pain via related mechanisms (MacDonald and Leary, 2005; Eisenberger et al., 2003; Kross et al., 2011). Our results here show that a specific area of the brain – PC – may also represent them similarly in the case of harms suffered by others. Representations of items within each harm subdomain, and across the harm domain, converged in PC (Fig. 4). But despite similar metaphorical commonalities for sociomoral and physical purity violations (e.g., Lee and Schwarz, 2010), multivariate representations of incest- and pathogen-based violations converged in distinct sets of neural regions. Principal-components and k-means clustering analysis of the behavioral data alone revealed a similar split: psychological and physical violations were not easily separable based on principal component scores, for any number of clusters, but pathogen and incest violations were. Together, the data argue that despite their common perceived immorality and disgustingness, and despite cross-subdomain similarity in LIFG, incest and pathogen violations are conceptually distinct categories in a way that physical and psychological harms are not.

There were also fundamental differences in the number and relationship of agents present within our purity, but not harm, scenarios. While both psychological and physical harm scenarios involved a standard dyadic template of "agent harms patient" (Gray et al., 2012), each dyadic partner's role is blurred in the case of incest (if both partners are active participants, which partner is "the agent", and which "the patient"?). And in pathogen scenarios, one person is both agent and patient. Intriguingly, representations of incest violations (but not pathogen violations) converged in regions linked to social cognition, including right and left TPJ and left superior temporal sulcus, and several of the behavioral factors contributing to separation of incest and pathogen violations in principal component space – weirdness, attention to behaviors, and attention to minds – are associated with social cognition. It is therefore plausible that, due to the heightened social cognitive demands of processing an interaction between multiple agents, representations of incest violations depend on representations of social cognitive features to a greater extent than do representations of single-actor pathogen violations.

4.4. Limitations

While the predetermined organization of our stimuli into categories is a strength, giving us a concrete organizational scheme to test, it is also a weakness. Our data cannot speak to the question of how the brain *naturally* organizes representations of moral violations: the inherent structure of the moral space. Nor can our data definitively answer whether the key features driving categorization here would play the same role for moral violations in other datasets. Addressing this requires both ecologically valid stimuli, with no prior structure, and data-driven methods. Hofmann et al., 2014identified seven moral categories in a dataset of everyday moral events volunteered by users of an experience-sampling app; however, they used a predetermined coding scheme rather than a data-driven approach. Chakroff (2015) used PCA to show evidence for binary organization of moral violations, with clusters corresponding roughly to harm and purity, within a dataset of unstructured violations volunteered by participants on Amazon Mechanical Turk. Though in this case the fMRI data (for representations in dmPFC specifically) did not match the two-factor behavioral solution, this study provides a model for future representational analyses of the moral space.

As the data presented here constitutes, to our knowledge, the first analysis of its type on moral stimuli, we cannot speak directly to the question of generalizability across different sets of stimuli. However, as an extension to the work presented here, we plan to collect an unstructured dataset of ecologically valid moral violations rated along a wider set of features, which will both address the question of inherent categorical structure and provide a more solid indication of which features are critical to conceptual organization, in a way that is not linked to the particular design of our experimenter-generated dataset.

A further design drawback is the split between intentional and accidental versions of each item. While this allows us to make more general inferences about moral representations here, it also means that

intentionality could be considered another feature of these violations, potentially altering their associated neural patterns and thereby shifting their placement within the moral space. We did not find any evidence for representational similarity based on the dimension of intent alone (Supplementary Information). But future work should explore whether the conceptual space of moral violations is stable or flexibly reorganized when features are added or removed.

The neural organization of moral acts may also differ across individuals. However, we did not find strong evidence for differences between NT and ASD participants here, despite prior work identifying autism-related differences in moral judgment (Gleichgerrcht et al., 2013; Moran et al., 2011; Zalla et al., 2011) and even differences in neural responses to moral scenarios in an alternate analysis of this dataset (Koster-Hale et al., 2013). Evidence for ASD-NT similarity in neural responses to Theory of Mind tasks (Dufour et al., 2013) hints that at least in some contexts the social-cognitive mechanisms underlying moral judgments may be similar across these groups, leading to similarity in neural representations of moral acts as well.

5. Conclusions

The present results add to a growing line of work that attempts to understand the neural organization of high-level, complex stimulus representations using more specific informational hypotheses (Chavez and Heatherton, 2015; Handjaras et al., 2016; Kriegeskorte et al., 2006). Not only can these approaches illuminate the neurocognitive correlates of these stimuli, as our results do here, but they can also provide specific organizational models to be tested and refined. In that vein, the current data suggest concrete priors for future studies: we expect to find broad similarity across moral violations in the mentalizing network, distinct regions of representation for moral violations in different domains, and greater diversity of neural representations within Purity than within Harm. The results also suggest concrete questions to be tested: which social cognitive features does the brain use to determine a scenario's moral relevance? What computational roles do PC and LIFG play in representing harm and purity violations? What kinds of fine-grained distinctions does the brain make between violations *within* domains? Ultimately, approaches like this one will contribute to a more specific, as well as more integrated, model of the vast and varied space of actions our minds deem 'morally wrong'.

Acknowledgements

This work was supported by funding from the Alfred P. Sloan Foundation, the Simons Foundation, the John Templeton Foundation (grant 5001695), and the National Institutes of Health (grant 1R01 MH096914-01A1). Data collection was supported by the Athinoula A. Martinos Center for Biomedical Imaging. The authors would like to thank Lily Tsoi, James Dungan, Jordan Theriault, Brendan Gaesser, and Larisa Heiphetz for helpful suggestions during data analysis and comments on earlier drafts of the paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.neuroimage.2017.07.043.

Appendix A. Scenarios

Stories were presented in cumulative segments:

(1) Background (6 s)
(2) Action (4 s)
(3) Outcome (4 s)
(4) Intent (4 s): accidental versus intentional

A.1. Physical harm

You are about to leave your friend's house when the carbon monoxide detector sounds. You check the back of the detector to see what might be going on.
You take a look, turn off the detector, and leave the house.
The detector detects dangerous levels of carbon monoxide in your friend's house.
From the back of the detector, you thought the detector just needed new batteries.
From the back of the detector, you knew that the detector was working properly.
You are house-sitting for a neighbor when pest control arrives. All people and pets must evacuate before fumigation begins, or else they will get very sick from the gases.
You leave and tell pest control they can begin fumigating.
The neighbor's dog is still in the house when the gases start getting pumped in.
From the sound of barking, you thought your neighbor's dog was safe in the backyard.
From the sound of barking, you knew your neighbor's dog was still inside the house.
You are teaching a kids' fire safety class. In the class, kids learn how to deal with real smoke in case of a fire. You decide who should participate.
You decide everyone must participate and begin to teach the class.
A boy in the class has asthma and cannot participate without risk of an asthma attack.
From the health forms, you did not realize this boy has asthma.
From the health forms, you fully realized that this boy has asthma.
You are a new employee at a sky diving company. A customer is looking to buy one of the store's parachutes. You consult with your manager first.
You sell the customer one of the parachutes. He plans to try it out tomorrow.
The chute failed in-house quality control and should have been tossed.
Based on what the manager said, you did not realize at all the chute was faulty.

Based on what the manager said, you definitely realized the chute was faulty.

You are traveling with your cousin. Your cousin is hot and wants to go swimming in the pond ahead. You look up information on the pond in your travel guide.

You tell your cousin it's safe to go swimming. He eagerly jumps in.

The pond actually contains chemical pollution and is not safe for swimming.

Because of what the guide book said, you did not realize the pond was unsafe.

Because of what the guide book said, you realized the pond was unsafe.

You are grocery shopping for your grandmother. Bagged spinach had recently been recalled for *E. coli* contamination, but some markets have begun carrying it again.

You buy spinach for your grandmother. You use it to make her a large salad.

The spinach is contaminated with *E. coli* and will make your grandmother very sick.

You had checked online and did not realize the spinach at your market was contaminated.

You had checked online, so you realized the spinach at your market was contaminated.

Your classmate wants to borrow your bike to go mountain biking. Your bike's brakes had not been working properly. Your bike has just come back from the repair shop.

You lend your classmate your bike, and he leaves the next day for a bike trip.

The brakes are still not working, and the bike is unsafe to ride.

After talking to the folks at the repair shop, you thought the brakes were fully fixed.

After talking to the folks at the repair shop, you knew the brakes were still broken.

You are at a Mexican restaurant. It is a slow day. There are no waiters nearby and only one other customer. This customer is sitting at the next table, and he starts coughing loudly.

You ignore the man's coughing and continue eating your meal.

He is actually choking on a piece of food and needs help.

Judging from the man's expression, you can't tell at all that he's choking on food.

Judging from the man's expression, you can absolutely tell he's choking on his food.

You and a friend are in a two person kayak in the ocean. The sun is beating down on you, and it would be cool and refreshing to take a swim in the surrounding water.

You tell your friend to jump in for a swim while you man the boat.

There are lots of jellyfish in the water that deliver painful stings to swimmers.

You looked and did not see any jellyfish in the water at any point along your ride.

You looked and actually saw jellyfish in the water at many points along your ride.

You and your friend are in the park roller-skating. You skate ahead and sit down behind a tree. You try to get comfortable, but there is a large stick in your way.

You toss the stick aside to make yourself comfortable, and it lands on the park path.

Your friend skates over the stick, and breaks his ankle.

You could not see that your friend was about to skate by, so you tossed the stick.

You saw that your friend was about to skate by, and you still tossed the stick.

You are at lunch during school. A classmate approaches you and asks you to show him some moves you learned recently in your martial arts class.

You tell him to stand back. You get ready and perform the martial arts kick.

Your classmate is standing too close, and you end up kicking him in the face.

You could not see that your classmate was standing too close, so you kicked.

You could see that your classmate was standing too close, but you kicked anyway.

Your family is over for dinner. You're taking cooking classes and wish to show off your culinary skills. For one of the dishes, adding peanuts will really bring out the flavor.

You grind up some peanuts, add them to that dish, and serve everyone.

Your cousin, one of your dinner guests, is severaly allergic to peanuts.

You had absolutely no idea about your cousin's allergy when you added the peanuts.

You knew about your cousins's peanut allergy when you added the peanuts to the dish.

A.2. Psychological harm

You are at a dinner party with friends when there is a lull in the conversation. You just heard a joke about a boy and his puppy that could liven things up.

You have everyone's attention, so you proceed to tell the joke.

Your friend's puppy was just hit by a car, and she is very upset by your joke.

You did not hear your friend's puppy died until after you told the joke.

You had just heard your friend's puppy died before you told the joke.

You and another classmate are about to give a formal presentation to many important faculty members. Your classmate is especially anxious about making a good impression.

The two of you take a deep breath and begin the presentation.

Someone in the audience points to your classmate's open fly, mortifying him.

You did not notice his open fly before that and could not have warned him.

You did notice his open fly before the presentation and could have warned him.

Your friend is a marine and was recently shipped to Iraq. You are watching the news when his name comes up as someone who has died in a recent bombing.

You rush over to your friend's parents to say you're sorry about their son's death.

The young man who died is not actually their son, just someone with the same name.

On the news, there had been no picture with the name, so you didn't realize it wasn't him.

On the news, there had been a picture with the name, so you realized it was not him.

Your sister is babysitting, and you go visit to go over to see what's up. When you get to the house, you find the baby but not your sister, who is in the basement fixing the fuse.

You take the baby home with you without telling your sister.

When your sister cannot find the baby, she panics and calls.

You thought your sister left, and you were doing her a favor by taking the baby.

You knew your sister had not left, and that she would panic if you took the baby.

You are sitting in math class on the first day of high school. As an introduction, the teacher asks everyone in class for an entertaining anecdote or a joke.

You decide to tell a joke. It is a joke about an elf and a midget.

A new student in your class happens to be a midget. He is hurt by your joke.

This student's sitting far behind you. You didn't realize he was in your class.

This student's sitting right in front of you. You realized he was in your class.

You are sitting in class without much to do. You notice a note of some kind sitting on the floor next to your desk.

You show the note to the girl behind you.

It is a poem written by your close friend about this girl, his secret crush.

The poem was not signed, so you really did not think your friend wrote it.

The poem is in your friend's writing, so you knew your friend wrote it.

You are taking a class on Shakespeare. On the very first day, the teacher asks for volunteers to deliver a monologue in front of the class.

You volunteer a classmate who is in the bathroom at the time.

When she returns, she has to perform for everyone. She is terrified of public speaking.

You just met this classmate and did not realize she has this fear.

You know this classmate from before and realized she has this fear.

You are in English class, prepping for the AP test at the end of the year. Your teacher passes out a sample essay that you are to discuss openly and honestly.

You suggest the essay must have been written by a third-grader.

The student who wrote the essay is in your class listening to your critique.

The essay was typed, so you completely didn't realize who had written it.

The essay was handwritten, so you realized right away who had written it.

You are at an ice skating rink with a large group of friends. One of your friends shows up in an ugly, frumpy looking sweater that his grandmother gave to him.

You make a joke about how his grandmother has terrible taste in sweaters.

His grandmother died very recently, and he is still very sad.

You hadn't heard anything about his grandmother passing away.

You had heard from someone that his grandmother passed away.

It's your little brother's birthday, and you are in charge of the surprise birthday party. You make your brother's party circus themed.

You dress up like a clown and jump out yelling 'Surprise' when he arrives.

Your little brother has an extreme phobia of clowns. He starts shaking.

You had never been to the circus with him and didn't know about his phobia.

You had been to the circus with him before, and so you knew about this phobia.

You are giving a PowerPoint presentation for your Psychology class. Your laptop is not compatible with the classroom equipment needed to display your presentation.

You announce that you're borrowing the laptop of a classmate who is in the bathroom.

When you hook it up, pornography pops up on his computer. The whole class sees.

You didn't see that a porn site was open because the browser was hidden.

You saw that a porn site was open because the browser was in front.

You receive an email from a distant acquaintance in your dorm. The email says she is committing suicide tomorrow. She says it is a secret.

You cannot bear this burden alone.

You forward her entire email out to everyone in your dorm.

You think you're forwarding her email to the school psychiatrist only.

You know that everyone in your dorm will read her private email.

A.3. Incest

You are walking your dog in the local park, and you see someone walking a dog of the same breed. The two of you strike up a conversation about dog walking.

After the third date, you decide to go home and sleep together.

The two of you are related by blood. You are half siblings.

You do not discover this until the next date after you'd already slept together.

You discovered this on the second date before you'd actually slept together.

You are on vacation by yourself in a national park, hiking and camping. After a day or so, you run into someone who happens to be from the same city as you.

A day later, you decide to have sex in your tent, using two forms of birth control to be safe.

The person you have sex with in your tent is your first cousin.

You didn't realize you're first cousins, as you're from estranged parts of the family.

You came to realize you're first cousins, as soon as soon as you met and started talking.

You are on a singles' cruise. On the first day you decide you've met your soulmate. Luckily, this person agrees, and you decide to move in together after the cruise.

After the trip, the two of you finally consummate your relationship on a waterbed.

You and this person you say is your soulmate are long lost twins.

You didn't know this until after you had sex, in talking about family drama.

You knew about this well before you had sex, in talking about family drama.

You are at the library doing some research for work. You end up chatting with a younger, attractive person who happens to be reading the novel that you are reading.

A coffee date and two dinner dates later, you end up in bed with this person.

It turns out that this person is the child you gave up for adoption decades ago.

In conversation, after that night, you find out this person is your child.

In conversation, before that night, you found out this person is your child.

You recently started chatting with someone in an online chat community. You live on opposite coasts, but you have been chatting nightly for weeks now.

You engage in cyber sex. For this, both of you pleasure yourselves on the computer camera.

Your cyber sex partner is your older sibling.

The camera shots were of the body only, so you didn't know it was your sibling.

The camera shots were of the body and face, so you knew it was your sibling.

You and your co-workers are at a strip club. Some of the dancers are wearing masks. Your co-workers buy you a dance, where you and the dancer go to a private booth.

The dancer is about to get fully undressed. You are very aroused by the dancer.

The dancer is your own child from your former marriage.

The dancer was wearing a mask, so you could not see that it's your own child.

The dancer had taken off the mask, so you could see that it's your own child.

You are at a dorm party, and you have a good time with someone there. This person feels the same way about you, like you've known each other forever.

At the end of the night, you decide to have sex, using a condom and a dental dam.

You two are actually long lost siblings.

The next day, you discover that you're siblings when you talk about family.

Earlier that night, you discovered you're siblings in talking about family.

You were separated from your fraternal twin at birth. You two have never met each other. Years later, you are on a blind date. Your friends at work set you two up.

After a stimulating date, you have sex in the cab ride back.

This person is actually your fraternal twin.

You didn't talk about your shared past until after, so you didn't know.

You talked a lot about your shared past at dinner, so you definitely knew.

You were adopted at birth and have never met either of your parents. At your college reunion, you go to your school's football game and meet someone a bit older.

That night, you two end up sneaking back into the stadium and having sex on the field.

The person you have sex with is actually your biological parent.

You did not know this was your parent, because you had never met.

You knew this was your parent, because you kept a photo with you.

You are interested in getting some minor plastic surgery and go to a clinic your aunt suggested. The receptionist who greets you is extremely attractive.

You end up having sex with the receptionist in one of the medical exam rooms.

The person you have sex with is your aunt's child, your own cousin.

Your cousin has had multiple plastic surgeries, so you did not recognize your cousin.

Your cousin has had multiple plastic surgeries, but you still recognized your cousin.

You at a family reunion. You find many members of your family to be very boring, but then you meet someone you have never seen before. The two of you start talking.

That night, you two sleep together, making sure to use birth control.

The two of you live on different continents, but you are first cousins.

You couldn't tell this person's related to you, and not just a family friend.

You could tell this person's directly related to you, not just a family friend.

You have been out of touch with your brother for nearly twenty years and finally decide to reunite. You go to his house, but he's out. Someone else answers the door.

The two of you instantly hit it off and have sex in the kitchen before your brother returns.

This person happens to be your brother's adult child.

You thought that this person was just a house sitter, not related to you.

You knew this person was your brother's child, not just a house sitter.

A.4. Pathogen

You are at a friend's party. People are chatting outside on the patio where you left your drink. Later in the night, you find that you are

extremely thirsty.

You go out to the patio, find your drink, and drink the whole thing.

Your drink contains a fair amount of someone's urine.

You did not realize that someone peed in your drink, so you finished it.

You realized that someone peed in your drink, but you finished it anyway.

You are in your garden planting when you decide to get a drink inside. Meanwhile, a dog wanders by where you were sitting. You go back outside to continue planting.

You scoop up a pile of dirt with your bare hands, getting some on your face.

The dirt pile is actually a pile of wet steaming poop left behind by the dog.

You didn't see the dog poop, so you continued without gloves.

You saw the dog poop, but you continued anyway without gloves.

You are at your neighbors' house with a glass of tomato juice when their vampire-obsessed -year-old daughter enters the kitchen. She starts talking about vampires.

You take your tomato juice to drink and finish it in a few gulps.

What you drank was half juice half human blood that this girl poured in.

You weren't looking when she poured the blood, so you drank your juice.

You were looking when she poured in the blood, but you drank it anyway.

You're preparing dinner for yourself, and decide to make rice pilaf. The last time you put the rice in the cupboard nearly a month ago, you left the container open.

You scoop up a large cup of rice, cook it, and eat all of it for dinner.

Your rice is infested with maggots, which you end up cooking and eating.

You didn't see the maggots when you scooped up the rice, so you had it for dinner.

You saw the maggots when you scooped up the rice but decided not to waste food.

You are having a picnic at the beach with friends. One of your friends is already very sunburnt and has been compulsively peeling flakes of skin off his body all day.

You and your friends decide to go swim for a while. When you come back, you eat your food.

While you were swimming, a breeze blew all the skin flakes onto your potato salad.

You didn't realize your potato salad was covered in skin flakes, so you ate it.

You realized your potato salad was covered in skin flakes, but you ate it anyway.

You are working at a free health care clinic. While you are sitting, somebody walking by spills iced coffee on your head. You look for something to wipe off your face and neck.

You find a cloth on the table and wipe off your entire face and neck.

The cloth is actually a used bandage covered in dried body fluids.

You did not know that the cloth was a used bandage, so you used it.

You knew that the cloth was a used bandage, but you used it anyway.

You are eating lunch at a new fast food restaurant. You decide to try the new 'super burger' on the menu. You're starving by the time the burger is in your hands.

You scarf down the whole burger along with your soda and fries.

The burger actually contains the tail of a tiny dead mouse that got cooked into your burger.

You did not see the tiny mouse tail at any point, so you finished your meal.

You saw a tiny mouse tail halfway through your meal but continued eating.

A car just killed your beloved dog. You had your dog for many years. Later, to cheer yourself up, you decide to cook dinner for yourself and your quirky housemate.

You take meat out of the freezer, chop it into smaller cubes and make stew for dinner.

The meat was from your dead dog. Your housemate had prepared and frozen it before it spoiled.

The meat was labeled 'beef' so you did not realize you were eating your dog.

The meat was labeled 'dog', so you did realize that you were eating your dog.

You are waiting to brush your teeth while your friend is in the bathroom. When she leaves, you go in. There's an opened pregnancy test on the counter by the sink.

You finish brushing your teeth and use a cup on the counter to rinse out your mouth.

Your friend just peed in that cup for her pregnancy test.

Your friend forgot to tell you she peed in that cup, so you thought it was just mouthwash.

Your friend told you she peed in that cup, so you knew it was urine, and not mouthwash.

Your roommate recently had liposuction around her stomach. She is now one week post operation. She is resting in the living room, when you return from the gym.

You go to take a shower. You've just run out of your favorite body wash, so you decide to use soap.

The soap you used to clean your body was made from your roommate's stomach fat.

The soap was labeled 'Dove', so you had no idea it was actually stomach fat soap.

The soap was labeled 'Fat', so you knew it was your roommate's stomach fat soap.

Your grandpa is in bed with a terrible cough. You decide to bring over a large container of soup for the two of you. You go to the kitchen to get two bowls and silverware.

You divide the soup up into two bowls and finish all of your serving.

When you were in the kitchen, your grandpa tipped a container of his phlegm into the soup.

You had no idea your grandpa spilled his phlegm because he didn't say anything.

You knew your grandpa spilled his phlegm because the container had tipped over.

You are at your uncle's house. Your uncle is somewhat mentally unstable. He collects many strange small objects. You decide to make yourself some

coffee and go to the kitchen.
You grind some coffee beans from a container, and brew some coffee.
The container contained your uncle's toenail clippings that you ground up with the beans.
The clippings were at the bottom, so you did not see anything wrong with the coffee.
The clippings were at the top, so you saw that something was wrong with the coffee.

NEUTRAL

You are taking a walk by the woods near your house when you run into a neighbor of yours, walking one big dog and one small dog. Your neighbor stops to say hi.
You say hi and bend down to pet your neighbor's dogs.
The big dog starts wagging its tail, but the small dog bears its teeth.
You didn't see that the big dog was friendlier than the small dog.
You saw that the big dog was friendlier than the small dog.
You are walking to your local supermarket. It has been raining, and the street is covered in large puddles. You see one in the path in front of you.
You go to jump over the puddle but don't make it, dunking your foot in the water.
Despite this, no water soaks through your shoe, and your foot remains dry.
You knew that your shoes were water resistant.
You didn't know that your shoes were water resistant.
You are a new employee at a popular clothes store in the mall. You go in for your first day of work and meet your first customer. They are looking for a new shirt.
The customer tries on a few shirts but decides not to buy any of them.
The customer leaves their shirts in the dressing room.
You realized this and refold the shirts to be put back.
You didn't realize this and refold the shirts at the end of the day.
You're out to dinner with some friends of yours. As the appetizers arrive, you ask whether anyone's seen any exciting movies or read any interesting books lately.
You tell your friends about a movie you rented last weekend.
Two of your friends saw this movie when it came out last year.
You did not realize this until after you started talking about it.
You realized they'd seen it because they told you a while back.
You are bored, and have started checking your friends' facebook pages to see if there is any news.
You go to the page of a friend you haven't spoken to since high school.
His status has changed from “single” to “in a relationship.
You didn't hear that he was in a relationship.
You already heard that he was in a relationship.
You and a friend are about to give a big presentation in class. You both have been working on it for many weeks are completely prepared.
You two show up for class and give the presentation.
The teacher enjoys it, especially the graphs you included.
You knew the teacher would enjoy the graphs.
You did not know the teacher would enjoy the graphs.
You and your partner are on a week-long vacation together. For the first time in a while, you're totally relaxed and not tied to your computer or Blackberry.
You spend much of the week in the hotel room, sleeping or having sex.
This is your first vacation since your honeymoon two years ago.
You didn't realize how much you needed a vacation.
You realized just how much you needed a vacation.
You are at work when you on break for lunch. You have finished eating, and are making a quick run to the bank to deposit this week's paycheck.
You go into the bank, and get in line for one of the tellers.
You chose this line because you are attracted to the bank teller.
You could tell that the bank teller also finds you attractive.
You couldn't tell that the bank teller also finds you attractive.
You have just graduated college and have started working at your new job. You get along especially well with someone from the office next to you.
You decide to date, and after several months, decide to sleep together.
Several years later this person becomes your spouse.
When you first met, you knew that they could be your spouse one day.
When you first met, you never knew that they could be your spouse one day.
You are in charge of teaching third-graders reading and writing skills. You usually read to the children first and then have them write and read aloud their own stories.
For this class, you have everyone to write about food they ate recently.
Two people in the class write about how they ate potato salad.
This hadn't come up before, so you didn't know they'd eaten potato salad.
This came up before, so you knew these students had eaten potato salad.
You are taking a stroll in the park by your house along the bank of a stream. It is finally nice weather outside, and you are enjoying the fresh cool air.
Every once in a while, you pick up a stone and skip it across the water.
As you reach for a stone, a big toad suddenly hops from next to it into the water.

You did not see the toad before it hopped away.
You could see the toad before it hopped away.
You are very hungry and decide to go to your favorite fast-food restaurant for lunch. You look through the many choices on the menu and decide to get the chicken sandwich.
You get your food and hungrily eat the whole thing.
While you eat, you spill a little bit of ketchup on your pants.
You realize this when it happens and clean it off.
You don't realize until you finish eating, then you clean it off.

Appendix B. List of behavioral measures and associated scales

HARM TO OTHERS.
"Is this action *bad for others*, but not necessarily bad for you (the person performing the action)?" [1: not at all bad for others; 7: extremely bad for others.
HARM TO SELF.
"Is this action *bad for you* (the person performing the action) but not necessarily bad for others?" [1: not at all bad for you; 7: extremely bad for you.
SITUATION ATTRIBUTION.
"There are situations that could lead a person to do this." [1: not at all; 7: absolutely.
PERSON ATTRIBUTION.
"A person is either the type to do this, or the type to never do this." [1: not at all; 7: absolutely.
DISGUST.
"How gross is this situation?" [1, not at all gross; 4: very gross.
RATIONALITY.
"How likely would a reasonable person be to do this?" [1: not at all likely; 7: absolutely likely.
WEIRDNESS.
"How abnormal and/or weird is your behavior?" [1: not at all abnormal/weird; 7: extremely abnormal/weird.
WRONGNESS.
"How wrong was this action?" [1: not at all morally wrong; 4: very morally wrong.
ATTENTION TO ENVIRONMENT/BEHAVIORS/MINDS:
"How much does this story make you think about:

- the physical environment?
- actions and behaviors?
- thoughts and desires?" [1: not at all; 7: very much]

References

Borg, J.S., Lieberman, D., Kiehl, K.A., 2008. Infection, incest, and Iniquity: investigating the neural correlates of disgust and morality. J. Cognit. Neurosci. 1529–1546.
Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 38, 1–38. http://dx.doi.org/10.1196/annals.1440.011.
Buckner, R.L., Sepulcre, J., Talukdar, T., Krienen, F.M., Liu, H., Hedden, T., Johnson, K.A., 2009. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J. Neurosci. 29 (6), 1860–1873. http://dx.doi.org/10.1523/JNEUROSCI.5062-08.2009.
Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. Brain 129, 564–583. http://dx.doi.org/10.1093/brain/awl004.
Chakroff, A., 2015. Mapping the Moral Domain from the Ground Up. Harvard University, Cambridge (Unpublished Doctoral Dissertation).
Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., Young, L., 2016a. When minds matter for moral judgment: intent information is neurally encoded for harmful but not impure acts. Soc. Cognit. Affect. Neurosci. 1–9. http://dx.doi.org/10.1093/biostatistics/manuscript-acf-v5 (August).
Chakroff, A., Dungan, J., Young, L., 2013. Harming ourselves and defiling others: what determines a moral domain? PLoS One 8 (9), e74434. http://dx.doi.org/10.1371/journal.pone.0074434.
Chakroff, A., Russell, P.S., Piazza, J., Young, L., 2016b. From impure to harmful: Asymmetric expectations about immoral agents. J. Exp. Soc. Psychol. 1–9. http://dx.doi.org/10.1016/j.jesp.2016.08.001.
Chakroff, A., Young, L., 2015. Harmful situations, impure people: an attribution asymmetry across moral domains. Cognition 136, 30–37. http://dx.doi.org/10.1016/j.cognition.2014.11.034.
Chapman, H.A., Anderson, A.K., 2013. Things rank and gross in Nature: a review and synthesis of moral disgust. Psychol. Bull. 139 (2), 300–327. http://dx.doi.org/10.1037/a0030964.
Chavez, R.S., Heatherton, T.F., 2015. Representational similarity of social and valence information in the medial prefrontal cortex. J. Cognit. Neurosci. 27 (1), 73–82. http://dx.doi.org/10.1162/jocn.
Curtis, V., de Barra, M., Aunger, R., 2011. Disgust as an adaptive system for disease avoidance behaviour. Philos. Trans. R. Soc. B Biol. Sci. 366 (1563), 389–401. http://dx.doi.org/10.1098/rstb.2010.0117.
Cushman, F., Gray, K., Gaffey, A., Mendes, W.B., 2012. Simulating Murder: the aversion to harmful action. Emotion 12 (1), 2–7. http://dx.doi.org/10.1037/a0025071.
Davis, T., Poldrack, R.A., 2014. Quantifying the internal structure of categories using a neural typicality measure. Cereb. Cortex 24 (7), 1720–1737. http://dx.doi.org/10.1093/cercor/bht014.
Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2010. fMRI item analysis in a theory of mind task. Neuroimage 55 (2). http://dx.doi.org/10.1016/j.neuroimage.2010.12.040.
Dufour, N., Redcay, E., Young, L., Mavros, P.L., Moran, J.M., Triantafyllou, C., Gabrieli, J.D.E., Saxe, R., 2013. Similar brain activation during false belief tasks in a large sample of adults with and without autism. PLoS One 8 (9). http://dx.doi.org/10.1371/journal.pone.0075468.
Dungan, J., Young, L., 2012. The two-type model of morality. A Companion Moral Anthropol. 1–29.
Dungan, J., Young, L., 2015. Understanding the adaptive functions of morality from a cognitive psychological perspective. In: Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource. http://dx.doi.org/10.1002/9781118900772.etrds0376.
Eisenberger, N.I., Lieberman, M.D., Williams, K.D., 2003. Does rejection hurt? An FMRI study of social exclusion. Science 302 (5643), 290–292. http://dx.doi.org/10.1126/science.1089134.
Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. U. S. A. 113 (33), 7900–7905. http://dx.doi.org/10.1073/pnas.1612033113.
Gleichgerrcht, E., Torralva, T., Rattazzi, A., Marenco, V., Roca, M., Manes, F., 2013. Selective impairment of cognitive empathy for moral judgment in adults with high functioning autism. Soc. Cognit. Affect. Neurosci. 8 (7), 780–788. http://dx.doi.org/10.1093/scan/nss067.
Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S.P., Ditto, P.H., 2012. Moral foundations theory: the pragmatic validity of moral pluralism. Adv. Exp. Soc. Psychol. 47, 55–130. http://dx.doi.org/10.1016/B978-0-12-407236-7.00002-4.
Graham, J., Nosek, B.A., Haidt, J., Iyer, R., Koleva, S., Ditto, P.H., 2011. Mapping the moral domain. J. Pers. Soc. Psychol. 101 (2), 366–385. http://dx.doi.org/10.1037/a0021847.

Gray, K., Schein, C., Ward, A.F., 2014. The myth of harmless wrongs in moral cognition: automatic dyadic completion from sin to suffering. J. Exp. Psychol. General 143 (4), 1600–1615. http://dx.doi.org/10.1037/a0036149.
Gray, K., Young, L., Waytz, A., 2012. Mind perception is the essence of morality. Psychol. Inq. 23 (2), 101–124. http://dx.doi.org/10.1080/1047840X.2012.651387.
Greene, J.D., Cushman, F.A., Stewart, L.E., Lowenberg, K., Nystrom, L.E., Cohen, J.D., 2009. Pushing moral buttons: the interaction between personal force and intention in moral judgment. Cognition 111 (3), 364–371. http://dx.doi.org/10.1016/j.cognition.2009.02.001.
Greene, J., Haidt, J., 2002. How (and where) does moral judgment work? Trends Cognit. Sci. 6 (12), 517–523. http://dx.doi.org/10.1016/S1364-6613(02)02011-9.
Greene, J.D., Sommerville, R.B., Nystrom, L.E., Darley, J.M., Cohen, J.D., 2001. An fMRI investigation of emotional engagement in moral judgment. Science 293 (5537), 2105–2108. http://dx.doi.org/10.1126/science.1062872.
Haidt, J., Koller, S.H., Dias, M.G., 1993. Affect, culture, and morality, or is it wrong to eat your dog? J. Pers. Soc. Psychol. 65 (4), 613–628. http://dx.doi.org/10.1037/0022-3514.65.4.613.
Handjaras, G., Ricciardi, E., Leo, A., Lenci, A., Cecchetti, L., Cosottini, M., Pietrini, P., 2016. How concepts are encoded in the human brain: a modality independent, category-based cortical organization of semantic knowledge. NeuroImage 135, 232–242. http://dx.doi.org/10.1016/j.neuroimage.2016.04.063. http://doi.org/.
Heekeren, H.R., Wartenburger, I., Schmidt, H., Prehn, K., Schwintowski, H.P., Villringer, A., 2005. Influence of bodily harm on neural correlates of semantic and moral decision-making. NeuroImage 24 (3), 887–897. http://dx.doi.org/10.1016/j.neuroimage.2004.09.026.
Helwig, C.C., Zelazo, P.D., Wilson, M., 2001. Children's judgements of psychological harm in normal and noncanonical situations. Child. Dev. 72 (1), 66–81. http://dx.doi.org/10.1111/1467-8624.00266.
Hofmann, W., Wisneski, D.C., Brandt, M.J., Skitka, L.J., 2014. Morality in everyday life. Science 345 (6202), 1340–1343. http://dx.doi.org/10.1126/science.1251560.
Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76 (6), 1210–1224. http://dx.doi.org/10.1016/j.neuron.2012.10.014.
Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A., 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. Nature 446 (7138), 908–911. http://dx.doi.org/10.1038/nature05631.
Koster-Hale, J., Saxe, R., Dungan, J., Young, L.L., 2013. Decoding moral judgments from neural representations of intentions. Proc. Natl. Acad. Sci. Unit. States Am. 110 (14), 5648–5653. http://dx.doi.org/10.1073/pnas.1207992110.
Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. Unit. States Am. 103 (10), 3863–3868. http://dx.doi.org/10.1073/pnas.0600244103.
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Bandettini, P.A., 2008. Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron 60 (6), 1126–1141. http://dx.doi.org/10.1016/j.neuron.2008.10.043.
Kriegeskorte, N., Simmons, W.K., Bellgowan, P., Baker, C., 2009. Circular analysis in systems neuroscience - the dangers of double dipping. Nat. Neurosci. 12 (5), 535–540. http://dx.doi.org/10.1038/nn.2303.
Kross, E., Berman, M.G., Mischel, W., Smith, E.E., Wager, T.D., 2011. Social rejection shares somatosensory representations with physical pain. Proc. Natl. Acad. Sci. Unit. States Am. 108 (15), 6270–6275. http://dx.doi.org/10.1073/pnas.1102693108.
Lakoff, G., Johnson, M., 1980. Metaphors We Live By. University of Chicago Press, Chicago, IL.
Lee, S.W.S., Schwarz, N., 2010. Dirty hands and dirty mouths: embodiment of the moral-purity metaphor is specific to the motor modality involved in moral transgression. Psychol. Sci. 21 (10), 1423–1425. http://dx.doi.org/10.1177/0956797610382788.
Leshinskaya, A., Contreras, J.M., Caramazza, A., Mitchell, J., 2017. Neural representations of belief concepts: a representational similarity approach to social semantics. Cereb. Cortex 1–14. http://dx.doi.org/10.1093/cercor/bhw401.
MacDonald, G., Leary, M.R., 2005. Why does social exclusion Hurt? The relationship between social and physical pain. Psychol. Bull. 131 (2), 202–223. http://dx.doi.org/10.1037/0033-2909.131.2.202.
Moll, J., de Oliveira-Souza, R., Eslinger, P.J., Bramati, I.E., Mourão-Miranda, J., Andreiuolo, P.A., Pessoa, L., 2002. The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. J. Neurosci. 22 (7), 2730–2736. http://doi.org/20026214.
Moll, J., de Oliveira-Souza, R., Moll, F.T., 2005. The moral affiliations of disgust a functional MRI study. Cogn. Behav. Neurol. 18 (1), 68–78.
Moll, J., Schulkin, J., 2009. Social attachment and aversion in human moral cognition. Neurosci. Biobehav. Rev. 33, 456–465. http://dx.doi.org/10.1016/j.neubiorev.2008.12.001.
Moran, J.M., Young, L.L., Saxe, R., Lee, S.M., O'Young, D., Mavros, P.L., Gabrieli, J.D., 2011. Impaired theory of mind for moral judgment in high-functioning autism. Proc. Natl. Acad. Sci. Unit. States Am. 108 (7), 2688–2692. http://dx.doi.org/10.1073/pnas.1011734108.
Nichols, S., 2002. Norms with feeling: towards a psychological account of moral judgment. Cognition 84, 221–236. http://dx.doi.org/10.1016/S0010-0277(02)00048-3.
Parkinson, C., Sinnott-Armstrong, W., Koralus, P.E., Mendelovici, A., Mcgeer, V., Wheatley, T., 2011. Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. J. Cognit. Neurosci. 3162–3180.
Pizarro, D., Helion, C., 2011. On disgust and moral judgment. Emot. Rev. 3 (3), 267–268. http://dx.doi.org/10.1177/1754073911402394.
Rozin, P., Fallon, A.E., 1987. A perspective on disgust. Psychol. Rev. 94 (1), 23–41. http://dx.doi.org/10.1037//0033-295X.94.1.23.
Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". NeuroImage 19 (4), 1835–1842. http://dx.doi.org/10.1016/S1053-8119(03)00230-1.
Schaafsma, S.M., Pfaff, D.W., Spunt, R.P., Adolphs, R., 2015. Deconstructing and reconstructing theory of mind. Trends Cognit. Sci. 19 (2), 65–72. http://dx.doi.org/10.1016/j.tics.2014.11.007.
Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34. http://dx.doi.org/10.1016/j.neubiorev.2014.01.009.
Semino, E., 2010. Descriptions of pain, metaphor, and embodied simulation. Metaphor Symbol 25 (4), 205–226. http://dx.doi.org/10.1080/10926488.2010.510926.
Shenhav, A., Greene, J.D., 2014. Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. J. Neurosci. 34 (13), 4741–4749. http://dx.doi.org/10.1523/JNEUROSCI.3390-13.2014.
Su, L., Fonteneau, E., Marslen-Wilson, E., Kriegeskorte, N., 2012. In: Spatiotemporal searchlight representational similarity analysis in EMEG source space. 2012 Second International Workshop on Pattern Recognition in Neuroimaging. http://dx.doi.org/10.1109/PRNI.2012.26.
Tamir, D.I., Thornton, M.A., Contreras, J.M., Mitchell, J.P., 2015. Neural evidence that three dimensions organize mental state representation: Rationality, social impact, and valence. In: Proceedings of the National Academy of Sciences, 201511905. http://dx.doi.org/10.1073/pnas.1511905112.
Turiel, E., 1983. The Development of Social Knowledge. Cambridge University Press, Cambridge.
Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2015. Reliability of dissimilarity measures for multi-voxel pattern analysis. NeuroImage 137, 188–200. http://dx.doi.org/10.1016/j.neuroimage.2015.12.012.
Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., Kanwisher, N.G., 2010. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. Proc. Natl. Acad. of Sci. U. S. A. 107 (15), 6753–6758. http://dx.doi.org/10.1073/pnas.0914826107.
Young, L., Cushman, F., Hauser, M., Saxe, R., 2007. The neural basis of the interaction between theory of mind and moral judgment. Proc. Natl. Acad. of Sci. Unit. States Am. 104 (20), 8235–8240. http://dx.doi.org/10.1073/pnas.0701408104.
Young, L., Saxe, R., 2008. The neural basis of belief encoding and integration in moral judgment. NeuroImage 40 (4), 1912–1920. http://dx.doi.org/10.1016/j.neuroimage.2008.01.057.
Zalla, T., Barlassina, L., Buon, M., Leboyer, M., 2011. Moral judgment in adults with autism spectrum disorders. Cognition 121 (1), 115–126. http://dx.doi.org/10.1016/j.cognition.2011.06.004.
Zhang, S., Li, C.S., 2012. Functional connectivity mapping of the human precuneus by resting state fMRI. NeuroImage 59 (4), 3548–3562. http://dx.doi.org/10.1016/j.neuroimage.2011.11.023.