A Functional Neuroimaging Investigation of Moral Foundations Theory

Maria Khoudary[1,2], Eleanor Hanna[5], Kevin O'Neill[2], Vijeth Iyengar[6], Scott Clifford[7], Roberto Cabeza[2,3], Felipe De Brigard[1,2,3], Walter Sinnott-Armstrong[1,2,4]

[1]Department of Philosophy, Duke University; [2]Center for Cognitive Neuroscience, Duke University; [3]Duke Institute for Brain Sciences; [4]Kenan Institute for Ethics, Duke University; [5]Valassis Digital; [6]U.S. Department of Health and Human Services; [7]Department of Political Science, University of Houston

BACKGROUND

Moral Foundations Theory (MFT) posits that moral cognition takes place in functionally specialized neurocognitive systems, or modules, that evolved in response to unique pressures from the social environment.[1]

The foundations—*harm (physical and emotional)*, *fairness*, *loyalty*, *purity*, *authority*, *liberty*—can be organized into superordinate categories that emphasize the value of individuals (individualizing: *harm and fairness*) vs. the value of group unity (binding: *loyalty*, *purity*, *authority*).[2,3]
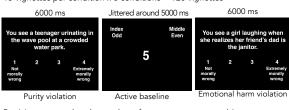
We use spatiotemporal partial least squares (PLS) analyses on functional neuroimaging (fMRI) data to investigate MFT's claim of modularity at the levels of the individual foundations and the superordinate categories.

STUDY DESIGN

27 healthy, right-handed adults (14 male; age=24.65 (4.21) years) judged the Moral Foundations Vignettes[3] while in the scanner (TR=2000ms).
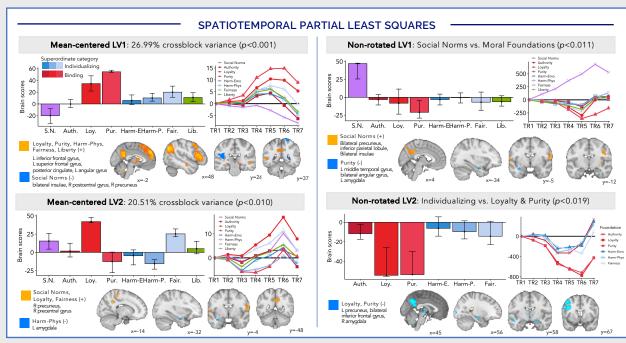
Vignettes were normed and validated to activate only one foundation. Vignettes depicting violations of amoral social norms served as a control.

15 vignettes per condition x 8 conditions = 120 vignettes

6000 ms — Jittered around 5000 ms — 6000 ms

You see a teenager urinating in the wave pool at a crowded water park.
1 Not morally wrong 2 3 4 Extremely morally wrong

Index Odd — Middle Even
5

You see a girl laughing when she realizes her friend's dad is the janitor.
1 Not morally wrong 2 3 4 Extremely morally wrong

Purity violation — Active baseline — Emotional harm violation

Participants completed a number of post-scan surveys probing memory for the vignettes, emotional reactions to them, and various sociopolitical trait-level attributes (e.g., empathy, conservatism, sensitivity to disgust)

1. Haidt, J., & Graham, J. (2007). *Soc. Justice Res.* 2. Graham, J., et al. (2009). *J Pers Soc Psychol.* 3. Iyer, R., et al. (2012). *PloS One.* 4. Clifford, S., et al. (2015). *Behav. Res. Methods.*

maria.khoudary@duke.edu | @maria_khoudary | mariakhoudary.com

SPATIOTEMPORAL PARTIAL LEAST SQUARES

Mean-centered LV1: 26.99% crossblock variance ($p<0.001$)

Superordinate category: Individualizing, Binding

Loyalty, Purity, Harm-Phys, Fairness, Liberty (+)
L inferior frontal gyrus, L superior frontal gyrus, posterior cingulate, L angular gyrus

Social Norms (-)
bilateral insulae, R postcentral gyrus, R precuneus

Mean-centered LV2: 20.51% crossblock variance ($p<0.010$)

Social Norms, Loyalty, Fairness (+)
R precuneus, R precentral gyrus

Harm-Phys (-)
L amygdala

Non-rotated LV1: Social Norms vs. Moral Foundations ($p<0.011$)

Social Norms (+)
Bilateral precuneus, inferior parietal lobule, Bilateral insulae

Purity (-)
L middle temporal gyrus, bilateral angular gyrus, L amygdala

Non-rotated LV2: Individualizing vs. Loyalty & Purity ($p<0.019$)

Loyalty, Purity (-)
L precuneus, bilateral inferior frontal gyrus, R amygdala

SUMMARY

A data-driven analysis (mean-centered PLS) of whole-brain activity returned a unique neurocognitive signature for only some of the individual foundations (Loyalty, Purity, and Fairness). It coarsely differentiated superordinate categories (individualizing and binding).

The non-rotated PLS, which contrasts conditions of interest against each other, corroborated (i) the neurocognitive distinction between judgments of Social Norm and Moral Foundation violations and (ii) that Authority does not bear a neurocognitive resemblance to the other binding foundations.