Moral Foundations Dictionaries for Linguistic Analyses, 2.0

Summary drafted by: Jeremy Frimer
Collaborators: Jonathan Haidt, Jesse Graham, Morteza Dehgani, & Reihane Boghrati

Abstract

The original Moral Foundations Dictionaries (MFD) have unknown psychometric properties, and fewer words in them (~32 per dictionary) than many conventional linguistic analytic dictionaries. Our objective was to test their validity and develop a larger and potentially more valid set of MFDs. After generating large sets of candidate words, we used word2vec software to estimate the prototypicality of each word within their respective foundations. We then tested the validity of the new dictionaries, and whether the validity diminished by trimming the dictionaries to include only highly prototypic words. Results suggest that the new dictionaries are more valid than the original set, and that using only highly prototypic words did not diminish the dictionaries' validity. We conclude by recommending that future research rely on the full-length MFD 2.0.

Background.

The original MFDs have relatively few words (32 on average; see Table 1). A number of words that are prototypic to the foundations are absent in the dictionaries. For instance, the care vice foundation (harm) does not include the words *murder*, *torture*, or *agony*. Whether or not these omissions limit the validity of the original dictionaries is unknown. Our first goal was to test the validity of the original MFDs.

Our second goal was to develop a new set of MFDs with more words, and test their validity too. Standard linguistic dictionaries in LIWC typically have hundreds of words in each. This led us to wonder whether increasing the number of words might improve the validity of the dictionaries. On the other hand, some of Morteza's recent work suggests that after about 30 words, increasing the number of words in a dictionary does not improve validity. We begin by developing new dictionaries. We then test the validity of the new and original dictionaries.

Step 1. Word Lists Generation

We generated enormous lists of words for each foundation. Jon and Jesse then selected those that they thought were conceptually relevant to each foundation. The result was 210 words per dictionary on average (see Table 1).

Table 1. The number of words/word stems in the original MFD, and the new MFD in its entirety, and with low, medium, and high prototypicality inclusion criteria.

	Care		Fairness		Loyalty		Authority		Sanctity		
	Virtue	Vice	Virtue	Vice	Virtue	Vice	Virtue	Vice	Virtue	Vice	Average
Original MFD	16	35	26	18	29	23	45	37	35	54	**32**
New MFD – full dictionaries	182	288	115	236	142	49	301	130	272	388	**210**
New MFD z > 0.4	72	110	54	115	56	34	118	75	130	160	**92**
New MFD z > 0.8	46	62	34	66	28	28	71	46	72	96	**55**
New MFD z > 1.2	32	40	25	30	17	15	39	20	36	51	**31**

Step 2. Prototypicality Estimation

We wanted to know which words were most and least prototypic of each foundation. To do so, we used word2vec software. It relies on the idea that related words appear close to one another in text passages. For example, the words *protect* and *safety* are likely to appear close to one another in texts because they are conceptually related. Word2vec uses a very large text corpus to estimate the prototypicality ("cosine") of target words to a set of seed words. Jon and Jesse selected seed words for each foundation (see Table 2).

Table 2. Seed words for each foundation that were used to generate prototypicality estimates.

Valence	Foundation				
	Care	Fairness	Loyalty	Authority	Sanctity
Virtue	kindness	fairness	loyal	authority	purity
	compassion	equality	team player	obey	sanctity
	nurture	justice	patriot	respect	sacred
	empathy	rights	fidelity	tradition	wholesome
Vice	suffer	cheat	betray	subversion	impurity
	cruel	fraud	treason	disobey	depravity
	hurt	unfair	disloyal	disrespect	degradation
	harm	injustice	traitor	chaos	unnatural

Morteza and Reihane then computed the prototypicality of each word in each foundation. To establish baseline prototypicality numbers, we also computed the prototypicality of 82 non-moral but common words (e.g., *drink*, *dawn*, *age*). As expected, the prototypicality estimates of foundations words were much higher than non-moral words, across the board (sparing you the details).

To create a common language across the foundations, we included both foundation and non-moral words and computed z-scores of prototypicality ratings within each foundation. For example, within the care virtue foundation, the word *caring* was highly prototypic of the foundations ($z = 2.74$), with words like *nurtures* ($z = 1.25$), *generously* (0.50), *healthiness* (0.00), *relieve* (-0.47), and *consoles* (-1.16) being successively less prototypic. We later use z-scores to set cut-offs for creating smaller dictionaries with only highly prototypic words.

Step 3. Validity Test

The final objective was to test whether word density analyses with the MFDs could successfully distinguish texts of known content. We asked people from around the world to write essays about the foundations, then tested how well the dictionaries picked out the content.

Sample. We recruited 1144 participants on the crowdsourcing website, http://crowdflower.com. Crowdflower is similar to Amazon's Mechanical Turk except that Crowdflower has participants from many more countries. Each participant received $0.50 to write an essay about one of the 10 moral foundations. A research assistant read each essay and identified ones that were not in English (104), respondents that declined the task (34), and incoherent texts (25). Following standard Pennebakerian protocols, we excluded an additional 256 responses that were less than 50 words long because short texts give unreliable word density estimates.

The final sample, $N = 656$, was 37% female, 33 years old on average ($SD = 11$) and from 58 different countries. The most common ones were Venezuela ($n = 79$), Egypt (62), the US (58), Serbia (42), Ukraine (39), India (30), Russia (29), Greece (21), Mexico (21), Italy (17), Spain (16), Canada (16), Germany (14), Philippines (14), Turkey (12), Argentina (12), Croatia (12), the UK (10), and Moldova (10). We asked participants to indicate their political ideology on social issues on a scale ranging from -100 (*extremely liberal*) to 100 (*extremely conservative*). The average participant was slightly liberal (-7) but the sample was quite diverse (*SD* = 50)

Procedure. We randomly assigned participants to write an essay about one of the 10 foundations. For instance, the instructions for the care virtue read:

> Please take a moment to recall a specific event in which a person (protagonist) acted with kindness, compassion, or empathy, or nurtured another person. The person who did this could have been you, or someone you know of. The person could also be a fictional individual from a book, movie, or TV show. When you have an event in mind, please proceed to the next page to answer some questions.

For the other foundations, the words in the first sentence after "person (protagonist)" were replaced with foundation-specific prompts (see Table 3.) Participants wrote in response to three questions: (a) "What led up to the event?" (b) "What did the person (protagonist) do?" and (c) "What were the outcomes and consequences?" After each question, the instructions were to "please explain thoroughly" before writing in three successive text boxes. We combined all the text from each participant to form short essays. Essays were 113 words long on average (*SD* = 100).

Table 3. Instructions for participants for the various foundations.

Foundation	Valence	Instructions
Care	Virtue	…acted with kindness, compassion, or empathy, or nurtured another person.
	Vice	…acted with cruelty, or hurt or harmed another person/animal and caused suffering.
Fairness	Virtue	…acted in a fair manner, promoting equality, justice, or rights.
	Vice	…was unfair or cheated, or caused an injustice or engaged in fraud.
Loyalty	Virtue	…acted with fidelity, or as a team player, or was loyal or patriotic.
	Vice	…acted disloyal, betrayed someone, was disloyal, or was a traitor.
Authority	Virtue	…obeyed, or acted with respect for authority or tradition.
	Vice	…disobeyed or showed disrespect, or engaged in subversion or caused chaos.
Sanctity	Virtue	…acted in a way that was wholesome or sacred, or displayed purity or sanctity.
	Vice	…was depraved, degrading, impure, or unnatural.

For example, a 59-year-old female from the Ukraine was assigned to the *care virtue* condition. She wrote:

> During the Great Patriotic War, an acquaintance of my grandmother lost a family: his wife, son, parents. He came from the war as a hero, but he was alone and did not see the point in life. He adopted a child, an orphan who lost his parents in the same war. The child was 6 years old, he was homeless, hungry, unhappy. This man gave all his love, care for the baby. Two lonely hearts melted, warming each other. A veteran of the war had a meaning in life. He cared for the child, sprouted it, fed it, dressed it, taught it. The child acquired a caring father and very soon ceased to cry in a dream, experiencing all the horrors of the war. The kindness of the former warrior gave the child the opportunity to survive in a difficult post-war period. And the man has a new meaning in life.

We then used LIWC to estimate the density of words from each of the 10 foundations in each essay. Our analyses examined whether the density of words in a foundation (e.g., care virtue) was higher than the density of the same dictionary words in the other 9 foundations. To keep things simple, we present the results as Cohen *d* effect sizes in Table 4. For example, a *d* = +0.56 in the loyalty virtue foundation with the new, full dictionary means that the density of loyalty virtue words was about half a standard deviation higher in the loyalty virtue essays than all other essays.

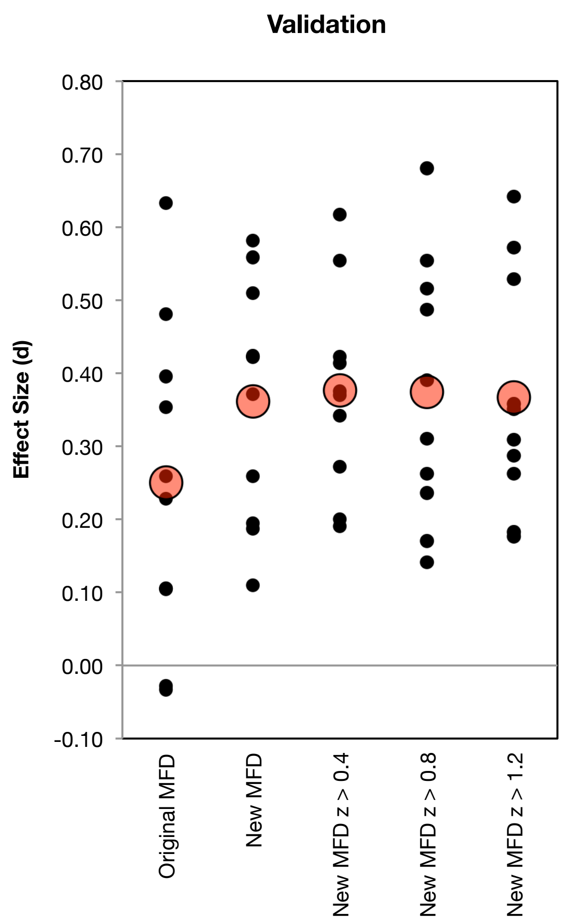
The first two rows of Table 4 show that the average validity of the new dictionaries was higher (*d* = 0.36) than that of the original dictionaries (*d* = 0.25). To test whether we benefit from having so many words in the new dictionaries, we created shorter, tighter MFDs with only highly prototypic words (using cut-offs of $z > 0.40$, 0.80, and 1.20, respectively; see Table 1 for dictionary word counts). Table 4 shows that the average validity neither increased nor decreased much as a result. In sum, the new dictionaries are more valid than the old. And trimming the new dictionaries to include only highly prototypic words seems to have few costs.

Table 4. Validity of the various MFDs. Numbers represent Cohen's d effect sizes distinguishing the density of words corresponding to a specific foundation (e.g., care virtue) to the density of the same dictionary of words in all the other 9 foundations.

	Care		Fairness		Loyalty		Authority		Sanctity		
	Virtue	**Vice**	**Virtue**	**Vice**	**Virtue**	**Vice**	**Virtue**	**Vice**	**Virtue**	**Vice**	**Average**
Original MFD	0.11	0.63	0.35	-0.03	-0.03	0.10	0.48	0.26	0.23	0.40	**0.25**
New MFD – full dictionaries	0.51	0.37	0.19	0.58	0.42	0.56	0.20	0.11	0.42	0.26	**0.36**
New MFD z > 0.4	0.34	0.38	0.20	0.62	0.42	0.55	0.27	0.19	0.41	0.37	**0.38**
New MFD z > 0.8	0.14	0.52	0.31	0.68	0.17	0.55	0.39	0.26	0.49	0.24	**0.37**
New MFD z > 1.2	0.18	0.57	0.29	0.64	0.18	0.53	0.35	0.26	0.36	0.31	**0.37**

Here are the same numbers represented graphically. The new dictionaries are measurably better but by no means knocking it out of the park.

Figure 1. Validity of the various MFDs (Table 4 represented graphically). Black dots represent the validity of a particular foundation. Large red dots represent the average across all foundations.



Conclusion. We recommend that researchers use of the MFD 2.0. Although the full-length version has no better or worse construct validity than the shorter variants, we recommend the full-length version because it more fully captures each foundation.