Values in Words: Using Language to Evaluate and Understand Personal Values

Ryan L. Boyd[1*], **Steven R. Wilson**[2*], **James W. Pennebaker**[1], **Michal Kosinski**[3], **David J. Stillwell**[4], **and Rada Mihalcea**[2]

[1]University of Texas at Austin, [2]University of Michigan, [3]Stanford University, [4]Cambridge University

ryanboyd@utexas.edu, steverw@umich.edu, pennebaker@mail.utexas.edu, michal@michalkosinski.com, ds617@cam.ac.uk, mihalcea@umich.edu

Abstract

People's values provide a decision-making framework that helps guide their everyday actions. Most popular methods of assessing values show tenuous relationships with everyday behaviors. Using a new Amazon Mechanical Turk dataset ($N = 767$) consisting of people's language, values, and behaviors, we explore the degree to which attaining "ground truth" is possible with regards to such complicated mental phenomena. We then apply our findings to a corpus of Facebook user ($N = 130,828$) status updates in order to understand how core values influence the personal thoughts and behaviors that users share through social media. Our findings suggest that self-report questionnaires for abstract and complex phenomena, such as values, are inadequate for painting an accurate picture of individual mental life. Free response language data and language modeling show greater promise for understanding both the structure and content of concepts such as values and, additionally, exhibit a predictive edge over self-report questionnaires.

Introduction

The increasing amount of publicly available web data has provided a new lens through which we can study how people are thinking, behaving, and feeling (Lazer et al. 2009). Recent developments in natural language processing and information retrieval techniques have allowed researchers to better understand and model social and psychological processes such as personality (Yarkoni 2010), emotion (Strapparava and Mihalcea 2008), and online behaviors (Zhang et al. 2011). We can now study psychological traits and their links to behaviors on a larger scale than ever before through the analysis of social media data.

The current research explores the psychological construct of values, their measurement, and their relationship with behaviors. Using natural language processing techniques, we analyze the ways in which people describe their personal values and behaviors, then compare them with closed (i.e., "forced choice") self-reports. We then expand our study of how values and behaviors are revealed in language to a large corpus of Facebook status updates. This project raises a central question: How should we measure values? That is, are values best measured through traditional self-reports or can we better assess them through the analysis of natural language? Finally, how are values – as measured either through questionnaires or language – related to behaviors?

Values and Value Research

Psychologists, historians, and other social scientists have long argued that people's basic values predict their behaviors (Ball-Rokeach, Rokeach, and Grube 1984; Rokeach 1968). Further, human values are thought to generalize across broad swaths of time and culture (Schwartz 1992) and are deeply embedded in the language that people use on a day-to-day basis (Chung and Pennebaker 2014; Lepley 1957).

In psychological research, the term *value* is typically defined as a network of ideas that a person views to be desirable and important (Rokeach 1973). Values are usually thought of as relatively abstract, giving rise to a broad constellation of related attitudes and behaviors. For example, a person who values "honesty" will typically hold a very negative attitude towards dishonest politicians and, accordingly, will be less likely to vote for them in the future (for a discussion of the links between values and attitudes, see Kristiansen and Zanna, 1988). Such core values are pervasive and often internalized at a very young age (Aronson 2004). It is generally believed that the values which people hold tend to be reliable indicators of how they will actually think and act in value-relevant situations (Rohan 2000).

Over the years, many researchers have conceptualized different frameworks that are believed to cover nearly all core human values (Rokeach 1968). One of the most accepted and widely used of these frameworks was developed by Schwartz and colleagues around two decades ago (Schwartz et al. 2012). The most prevalent form of this approach to the study of values suggests that there are ten primary values organized into a circumplex. This circumplex serves as the umbrella under which the majority of human value judgments fall. These 10 value types are as follows:[1]

[*]We note equal contributions to the current work.

[1]Recent work by Schwartz and colleagues has specified refinements to the classic version of the theory that identifies 10 core values (Cieciuch, Schwartz, and Vecchione 2013). However, the 10 values that we use for the current research have remained across theoretical iterations.

- Self-direction (S-D)
- Stimulation (Stim)
- Hedonism (Hed)
- Achievement (Achiev)
- Power (Pow)
- Security (Sec)
- Conformity (Conf)
- Tradition (Trad)
- Benevolence (Benev)
- Universalism (Univ)

Schwartz's 10-value model has seen great success in psychological research as well as other fields. The basic circumplex model has been applied to the understanding of culture (Schwartz 1994; 2004), religion (Schwartz and Huismans 1995), cognitive development (Bubeck and Bilsky 2004), and politically-motivated behaviors (Caprara et al. 2006), to name but a few domains. Generally speaking, the vast majority of this research has been built upon the Schwartz Value Survey (SVS), an internally reliable self-report questionnaire commonly used to assess the theorized ten core human values (Schwartz 1992). The SVS's greatest asset is that it is now the common currency of values researchers around the world.

As impressive as the Schwartz approach to values is, it is constructed on the foundation of people's self-theories. That is, the SVS requires people to evaluate themselves along a predetermined group of 10 values that are assumed to take a specific structure constituted of specific content. Ultimately, this structure and content are imposed upon research participants by the fact that they are inherently built into the questionnaire and its scoring methods – a necessary practice for nearly all self-report questionnaires. Importantly, this is a very different approach than simply asking people for their own thoughts on the question of "What are your personal values that guide your decisions and behaviors?" Indeed, if asked this question, many people might answer "to work hard", "be faithful to my religion", or "be a good mother". Such professed values are not inherently contradictory to the SVS. Rather, the SVS lacks the ability to concretely reflect those *specific* values that people hold in their own personal value constellations.

An even more complex problem arises when studying the relationship between values and behaviors. Unfortunately, most studies attempting to examine value–behavior links have simply compared self-reported SVS values with other self-report attributes such as personality, likes, and dislikes. This creates a problem wherein researchers are often ultimately exploring the relationships between different facets of people's explicit self-concepts rather than studying more organic and real-world instantiations of values and behaviors. In fact, Schwartz has pushed for researchers to explore behaviors in more detail. This undertaking seems promising and has been the focus of recent research that seeks to build a set of self-report behaviors that correspond to the values measured by the SVS (Butenko and Schwartz 2013). Unfortunately, many of the self-reported behaviors thus far have been general abstractions rather than concrete behaviors. For example, the behavioral measure for the value of "stimulation" was "change plans spontaneously", and for the value of "humility", "play down my achievements or talent."

A related issue with which all social scientists struggle is the question of how to measure behaviors efficiently and effectively. Self-reports of behaviors via forced choice questionnaires ultimately suffer from the same problem as other self-report measures: the questionnaires only contain questions that researchers think to ask. By adopting such an approach, researchers run the risk of imposing a potentially skewed, and sometimes inaccurate, structure on behavioral patterns. These are intractable features inherent to virtually all closed-format self-report questionnaires. In most cases, we would like to know what behaviors our respondents are actually doing and thinking about without relying upon questionnaire prompts. Currently, researchers are beginning to acquire greater amounts of objective behavioral measures such as buying behaviors, movement information, and even reading pattern data as the "big data" revolution continues to grow (Kolb and Kolb 2013). In the interim, however, researchers now have access to an endless stream of open-ended reports of mental life in the form of social media. A principal benefit of these reports is that they are ecologically valid and driven entirely by what people say they are doing and thinking in their own words.

The current study examines values and behaviors that emerge from open-ended text. The first of two projects relies on an online survey. This survey involved multiple randomized tasks that included 1) asking people to describe in detail the basic values that guide their lives, 2) asking people to describe the behaviors in which they engaged within the past week, and 3) participant completion of the self-reported SVS.[2] Using a topic modeling technique called the meaning extraction method (Chung and Pennebaker 2008; Kramer and Chung 2011), values and behaviors were inductively extracted from the texts. Value- and behavior-relevant thematic factors were then compared with each other and with the SVS data.

The second project adapted the results of the first project and applied them to status updates from over 130,000 Facebook users; these data are part of the myPersonality project (Kosinski, Stillwell, and Graepel 2013). Although a relatively small number of the cases ($N = 1,260$) included the SVS, the primary analyses revealed intuitive links between the MEM-derived values and MEM-derived behaviors. The work presented here, then, constitutes a proof-of-concept study demonstrating the utility of relying on natural language markers of abstract psychological phenomena, including values, to better predict and understand their connections to behaviors and thought in a broader sense.

Project 1: Values and Behavior in an Online Survey Sample

To begin, we sought to determine how well the SVS captures prevalent values as described by people when discussing the things that are most important to them (i.e., their core personal values) in their own words. Additionally, we sought to explore the links between values (both from the SVS

[2]For this study, we also collected data in the form of closed questionnaires about recent behaviors from all participants. These items corresponded very strongly with the free-response behavioral reports provided by participants. Results are available from the authors.

and people's free responses) and human behaviors as they manifest themselves in the real world. Theoretically, values should exhibit a discernible influence upon behaviors, including language use. As such, we expected to see that the values reflected in a person's descriptions of their guiding principles would show relatively intuitive, predictive links to everyday behaviors. To capture this information, we designed a social survey using the Qualtrics Research Suite;[3] the survey was then distributed using Amazon Mechanical Turk (AMT). [4] Survey takers were presented with a series of randomized tasks that included the following:

- **Values Essay.** In order to assess participants' values in their own words, they were asked to respond to the following prompt:

  For the next 6 minutes (or more), write about your central and most important values that guide your life. Really stand back and explore your deepest thoughts and feelings about your basic values. You might think about the types of guiding principles that you use to make difficult decisions, interact with other people, and determine the things that are important in your life and the lives of those around you. Try to describe each of these values and their relationship to who you are. Once you begin writing, try to write continuously until time runs out."

- **Behavior Essay.** Similarly, a prompt was given with the aim of collecting natural language related to everyday behaviors. This prompt was not intended to acquire a list of all behaviors in which all participants engaged. Rather, our goal was to acquire a natural language behavioral inventory that reflected common, psychologically meaningful behaviors. The writing prompt read as follows:

  For the next 6 minutes (or more), write about everything that you have done in the past 7 days. For example, your activities might be simple, day-to-day types of behaviors (such as eating dinner with your family, making your bed, writing an e-mail, and going to work). Your activities in the past week might also include things that you do regularly, but not necessarily every day (such as going to church, playing a sport, writing a paper, having a romantic evening) or even rare activities (such as skydiving, taking a trip to a new place). Try to recall each activity that you have engaged in, starting a week ago and moving to the present moment. Be specific. Once you begin writing, try to write continuously until time runs out."

- **Schwartz Value Survey** Respondents were asked to assign integers in the range [-1,7] to the 57 different value items of the SVS based on how important they perceived them to be as guiding principles in their own lives. With this scale, higher numbers indicate greater personal importance – responses were made using a Likert-type scale. Scores for the ten values were then calculated by taking the mean of the individual items that characterize each particular value type, with corrections being performed to address respondents' differences in use of the response scale. This step involves computing the average score for each individual across all 57 survey items, then centering each item's score around that average value (Schwartz 2009).

Tasks were presented in a randomized fashion between participants in order to minimize the potential for order effects, placing boundaries on any effects that may have been present. Participants were allowed to take as much time as needed to complete each section of the study and were encouraged to be as comprehensive as possible in their responses to the writing prompts. In order to filter out spam and careless responses, multiple "catch" items were randomly interspersed throughout the survey. These items asked users to select a particular answer that could be easily verified (e.g., "For this question, please select the third option") – participants who failed to respond to catch items were excluded from all analyses. Additionally, each of the essay writing samples was manually checked for coherence and plagiarism. Between the months of May and July, 2014, surveys successfully completed by 767 respondents (64.5% female, 77.1% Caucasian, 70.0% aged 26-54) were retained using the aforementioned criteria.

Analysis

In order to model the natural language data from participants into statistically actionable metrics, we employed the meaning extraction method (MEM). The MEM is an approach to topic modeling for natural language data that possesses demonstrated utility in understanding psycholog-

Table 1: Themes extracted by the MEM for the values essay writing task, Project 1.

Theme	Example Words
Faith (Positive)	God, Christian, Faith, Bible, Church
Empathy	People, Treat, Respect, Kind, Compassion
Family Growth	Family, Good, Child, Parent, Raise
Work	Work, Best, Hard, Job, Goal
Decision Making	Make, Feel, Decision, Situation, Difficult
Honesty	Honest, Trust, Lie, Truth, Loyalty
Faith (Negative)	Belief, Bad, Wrong, Religion, Problem
Social	Life, Love, Friend, Relationship, Enjoy
Growth	Life, Learn, Live, Grow, Easy
Indulgence	Money, Enjoy, Spend, Free, Change
Caring/ Knowledge	Know, Care, Give, Allow, Truth
Openness	Happy, Mind, Open, Positive, See
Knowledge Gain	Better, Learn, Understand, Experience, Realize
Principles	Guide, Principle, Situation, Central, Follow
Freedom	Strive, Action, Nature, Personal, Free
Certainty	Right, Sure, Strong, Stand, Thought

[3]qualtrics.com/research-suite/

[4]requester.mturk.com

ical phenomena, including both cognition (Chung and Pennebaker 2008) and behaviors (Ramirez-Esparza et al. 2008). In essence, the MEM allows researchers to discover words that repeatedly co-occur across a corpus. When considering modest to large numbers of observations together, the co-occurrence of words can converge to identify emergent and psychologically meaningful themes. These themes are then treated as independent dimensions of thought along which all texts can be quantified. Like most topic modeling methods, the MEM omits closed-class (function) words and low-frequency open-class (content) words to ensure reliability and validity. For the current research, we used software designed specifically to automate topic modeling and lemmatization procedures (Boyd 2014a). With the MEM approach, we identified 16 themes from the language generated during the values essay task (Table 1) and 27 themes from the behavior essay task (Table 2).[5]

Table 2: Themes extracted by the MEM for the behaviors essay writing task, Project 1.

Theme	Example Words
Time	Night, Sunday, Friday, Thursday, Today
Daily Routine	Work, TV, Shower, Wake, Sleep
Fiscal Concerns	Need, Spend, Money, Buy, Make
Family Care	Husband, School, Nap, Child, Birthday
Chores	House, Clean, Laundry, Cook, Wash
Errands	Grocery, Store, Doctor, Bank, Dinner
Personal Care	Shower, Dress, Brush, Hair, Party
Time Awareness	Day, Year, Yesterday, Week, Hour
Gaming	Play, Game, Online, TV, Video
Routine (Meta)	Early, Week, Routine, Activity, Schedule
Media Consumption	Online, Listen, Music, Show, Internet
Enjoyment	Friend, Drink, Weekend, Party, Fun
Exhaustion	Drove, Slept, Late, Doctor, Tire
Social Maintenance	Friend, Family, Call, Phone, Visit
Car/Bill	Car, Bill, Paid, Hard, Facebook
Information Consumption	Watch, Read, Book, News, Usual
Yard work	Water, Garden, Yard, Plant, Mow
Relaxing Afternoon	Stay, Enjoy, Rest, Afternoon, Time
Car Body	Car, Minute, Fix, Gas, Gym
Task Preparation	Start, Coffee, Begin, Prepare, Sit
Petcare	Water, Cat, Fed, Feed
Secondary Fiscal	MTurk, Coffee, Fix, Mail, Bank
Relaxation	Watch, Move, Relax, Pizza, Summer
Travel	Walk, Drive, Park, Trip, Swim
Meetings	School, Church, Class, Meeting, Attend
Student	Work, Job, Parent, Relax, Hour
Momentary Respite	Outside, Television, Cooking, Bath, Snack

	Conformity	Tradition	Security	Power	Achievement	Hedonism	Stimulation	Self-Direction	Universalism	Benevolence
Religion	●	●				○	○	○	○	•
Empathy					○				•	•
FamilyGrowth	•	•	•				○	◯	○	
Work										
DecisionMaking										
Honesty										•
NegativeReligion										
Social		•						○	○	
Growth										
Indulgence							•			
CaringKnowledge										
Openness										
KnowledgeGain	○	○	○				•	•	•	
Principles										
Freedom	○		○			•	•			
Certainty										

Table 3: Relationships between SVS values and MEM-derived value themes, Project 1.
Positive relationship: • = $R^2 \geq .01$, ● = $R^2 \geq .04$.
Negative relationship: ○ = $R^2 \geq .01$, ◯ = $R^2 \geq .04$.

The MEM-derived value themes capture the various semantic topics that people generate and, more broadly, tend to focus on when asked to reflect upon and discuss their values. Such themes lack the constraints of a forced choice questionnaire and, like other assessment methods, allow for nuance and variability between individuals. After performing the standard MEM procedures for theme extraction, we sought to determine how these topics correspond to the 10 values as defined in the SVS. To quantify each MEM-derived theme for individual respondents, we used word counting software (Boyd 2014b) to measure the rate of words from each theme as they appeared in each essay response. For example, an individual who used 4 "empathy" words out of 100 total words would attain a score of 4% for this theme. Following these calculations, we then correlated scores for the MEM-derived values with the values quantified by the SVS. This comparison is summarized in Table 3.

The established relationships among the SVS values seem to exhibit themselves here. For each of the SVS value dimensions, the correlations tend to exhibit an expected sinusoidal trend against the MEM-derived themes. Additionally, we see relatively intuitive correlations between MEM-derived values and the SVS in a way that might be expected. Peoples' use of words from the "religion" theme align well with the SVS Tradition value and fall in opposi-

[5]Like other topic modeling methods, researchers have some degree of leeway in determining the number of themes extracted. For the MEM, theme interpretability is typically a key determining factor in deciding how many themes to retain. While other potential solutions were available, the adoption of an alternate number of themes does not impact the conclusions that we draw from the current research.

Value	Score	Value	Score
Achiev	.03	Sec	$-.32$
Benev	.08	S-D	.88
Conf	$-.22$	Stim	$-.05$
Hed	.61	Trad	.28
Pow	-1.72	Univ	$-.22$

Table 4: SVS Scores for "Participant Z".

tion to the SVS value of Self-Direction. We see small positive correlations between theme-score pairs such as Honesty/Benevolence, KnowledgeGain/Universalism, and Indulgence/Stimulation. However, we note that the correlations between the MEM-derived values and the SVS value scores are considerably weaker than would be expected were they reflecting identical constructs. Given their hypothetical measurement of the same broad construct (i.e., "values"), convergence would be expected to a rather high degree, reflected by moderately strong effect sizes; this was not the case. In other words, the ideas that people described when asked about their core personal values appear to show divergence from the top-down, theory driven set of values offered by the SVS. To illustrate the discrepancy, consider an example of one respondent's description of their core personal values. The following text is the entire description provided by a single participant, heretofore referred to as *Participant Z*, in response to the previously described "Values Essay" writing prompt:

> *Mainly in my life I try to maintain a moral standing with everyone I meet. I like to branch out and speak with others when they appear to be happy and in the mood to socialize. I try to work hard and make money in an honest fashion so that I may live a healthy and normal life. I try my best to maintain a positive attitude and outlook every day. I live life hoping for the best and looking forward instead of back.*

Consider Participant Z's scores along the SVS dimensions (Table 4). While this person's scores along the 10 theorized value dimensions of the SVS provide no indication of any particularly strong or cohesive values, a casual reading suggests that this respondent does possess a coherent network of ideas that they believe guides their daily behaviors. In this example, the SVS offers little insight into Participant Z's values, yet the quantification of their values from language appear to show some rather strong indications of their guiding principles, particularly when considered in relation to the sample's means (Table 5). Additionally, the MEM-derived value themes afford relatively transparent interpretation of the relative importance of each theme, even without consideration of the broader sample. These results should not be taken to suggest an inherent inferiority of the SVS. Rather, we emphasize that all self-report questionnaires designed to assess personal values would likely show similar discrepancies.

Viewing values as constructs that inherently influence people's behavior, we also expect to see meaningful relationships between people's values and measurements of common, everyday behaviors in which they engage. To examine these links, we performed simple Pearson's correlations between the 27 behavioral themes extracted from participant behavior essays (quantified in a fashion parallel to the values themes) and values as assessed by both the SVS and MEM-derived themes (results are presented in Table 6). The results of this analysis show that the SVS values exhibit low predictive coverage of themes related to everyday behaviors, yet the themes extracted from value descriptions show connections (i.e., effect sizes of $R^2 \geq .01$) to more than twice as many common behavior topics. In other words, of the 27 behavioral themes extracted, only 6 are predicted by participant SVS scores. On the other hand, the MEM-derived value themes exhibit correlations with 14 behavioral themes. The behavior themes "Relaxation" and "Meetings" were the only themes that exhibited relationships exclusively with SVS values and none of the MEM-derived value themes. Beyond these small relationships, SVS coverage of behavioral themes was in no place stronger than that afforded by the MEM-derived value themes.

In summation, the SVS dimensions are theorized to be those values that are universal and, importantly, such values are consciously accessible and able to be explicitly reported by the individual (Schwartz et al. 2012). However, in using an open-ended method for assessing a person's values where we can rely upon their own words, we see a constellation of values not captured by the top-down, theory driven approach of the SVS, which necessarily captures a limited semantic breadth. Furthermore, our language-based assessment of values exhibits better predictive coverage of an established criterion: everyday behaviors. As such, Project 1 provides further support for previous work suggesting that a person's values are predictive of behaviors. Importantly, however, we find that the network of values that are able to be captured from a person's own words appear to show predictive validity above and beyond that of a traditional self-report.

MEM-derived Value	Respondent Score	Sample Mean
Faith (Positive)	0.00	0.53
Empathy	3.57	2.93
Family Growth	0.00	1.51
Work	4.76	1.10
Decision Making	1.19	1.20
Honesty	1.19	0.86
Faith (Negative)	0.00	0.89
Social	3.57	3.43
Growth	5.95	2.48
Indulgence	1.19	0.83
Caring/ Knowledge	0.00	0.65
Openness	2.38	1.12
Knowledge Gain	0.00	0.08
Principles	-1.19	0.71
Freedom	0.00	0.43
Certainty	1.19	0.34

Table 5: MEM-derived value scores for "Participant Z".

	Time	Daily Routine	Fiscal Concerns	Family Cares	Chores	Errands	Personal Care	Time Awareness	Gaming	Routine (Meta)	Media Consumption	Enjoyment	Exhaustion	Social maintenance	Car Bill	Information Consumption	Yardwork	Relaxing Afternoon	Car Body	Task Preparation	Petcare	Secondary Fiscal	Relaxation	Travel	Meetings	Student	Momentary Respite
Schwartz Values																											
Achievement																											
Benevolence				⬤																					●		
Conformity				●																							
Hedonism				○																			●		○		
Power																										●	
Security																											
Self-Direction				○																							
Stimulation				○								●															
Tradition				●								○															
Universalism		○		○																						○	
MEM Values																											
Religion				●			●																				
Empathy						●					●			●		●								●			
FamilyGrowth				⬤		●		○						●													
Work				●																							
DecisionMaking																											
Honesty												●							●	●							
NegativeReligion		●																								●	
Social				●										●													
Growth																											
Indulgence				●					●																		
CaringKnowledge																											
Openness																											
KnowledgeGain				○																							
Principles																											
Freedom																											
Certainty		●																								⬤	

Table 6: Coverage of MEM-derived behavioral themes by SVS values and MEM-derived value themes in Project 1. Positive relationship: • = $R^2 \geq .01$, ● = $R^2 \geq .04$. Negative relationship: ○ = $R^2 \geq .01$, ◯ = $R^2 \geq .04$.

Project 2: Values in Social Media

The primary goal of Project 2 was to conceptually replicate the results from Project 1 in a real-world social media sample. To do so, we began by examining the relationship between social media users' SVS scores and the 16 MEM-derived value topics from our original AMT sample. For this project, we used an extensive sample of social media user data is available from the myPersonality project (Kosinski, Stillwell, and Graepel 2013). This dataset consists of approximately 150,000 Facebook user's status updates. Additionally, various subsamples of these users have completed some portion of a battery of dozens of questionnaires pertaining to personality assessment, demographics, and values.

While our AMT sample in Project 1 revealed value themes using language explicitly related to people's core values, value-laden language is also prevalent in everyday life (Chung and Pennebaker 2014). In Project 2, language pertaining to values and behaviors are not inherently differentiated, as all language was acquired exclusively from user status updates. As such, we used the MEM-derived value lexicon created within Project 1 as our "ground truth" for value-relevant words in Project 2. MEM-derived values for Facebook users were measured using word counting software (Boyd 2014b) to scan user status updates for the predetermined value-relevant words; this procedure was parallel to the language-based value quantification method described for Project 1.

To ensure reliability, all participants were required to have a minimum of 200 words used across all status updates (participants meeting criteria: $N = 130,828$). Those users included in the myPersonality dataset who had completed demographic surveys reported an average age of 25.3 years ($SD = 11.1$), and 56% identified themselves as female. Additionally, a subsample of the myPersonality dataset included Facebook users who had also completed the SVS online ($N = 1,260$).[6]

[6]Average SVS scores were generally analogous to those from Project 1's AMT sample.

Analysis

As a first step, SVS scores for Facebook users were correlated with the MEM-derived value themes as they were present in the users' status updates (Table 7). Again, we see only partial coverage of value-relevant language in terms of value dimensions captured by the SVS. However, in this sample, we see a decrease in the predictive coverage of the SVS with regard to value-laden words in participant status updates. The weakened correspondence between these two measures is to be expected – unlike Project 1, participants are not likely to be explicitly enumerating their core values. However, these results also suggest that those constructs measured by the SVS may not permeate into everyday life to the extent that researchers have typically assumed, whereas value-laden language does.

As with Project 1, we also sought to examine the links between Facebook users' core values and other aspects of mental life, primarily behavior. As was described for the first project, we first used the MEM to extract topical themes from the entire myPersonality corpus that met our minimum word count inclusion criteria. This procedure resulted in 30 broad themes found within Facebook user status updates (Table 8).[7] A few of the behavioral themes derived from the Facebook users' language have analogs to those themes found in the AMT behavior essay responses (e.g., "Day to Day" and "Daily Routine", "Children" and "Family Care") but, in general, many of the themes derived from Facebook status updates pertain to qualitatively novel topics. Unlike the behavioral themes from the first project, the topics in the status updates give us insight not only into what people are doing in behavioral terms (e.g., eating, studying, expressing gratitude, playing games), but also the things about which they are thinking (e.g., privacy, national issues, illness).

Importantly, many of the behavioral themes that were extracted from the corpus included words that were also found within the MEM-derived value themes found in Project 1. Many behaviors in which people engage will necessarily be value-laden to some degree, however, we sought to minimize effect size inflation due to shared word use between Project 1's MEM-derived value themes and Project 2's MEM-derived behavioral themes. As such, words that appeared in both sets of themes were systematically omitted from the behavioral themes prior to quantification. As with value-relevant words, each Facebook user's entire set of posts was then quantified along each MEM-derived behavioral dimension using the same word counting approach described above.

[7] Additional themes could be extracted, however, themes not intuitively reflecting cognition or behavior were excluded. Extraneous themes largely reflected culture (e.g., specific word spelling such as "neighbour" and "arse" from the U.K.) or verbal fries (e.g., "gurl", "cuz"). Retention of these themes did not alter the results or conclusions.

	Conformity	Tradition	Security	Power	Achievement	Hedonism	Stimulation	Self-Direction	Universalism	Benevolence
Religion	○			○				•	•	
Empathy										
FamilyGrowth									•	
Work										
DecisionMaking										
Honesty										
NegativeReligion										
Social								•	•	
Growth								•	•	
Indulgence			○						•	
CaringKnowledge								•	•	
Openness									•	
KnowledgeGain										
Principles										
Freedom										
Certainty										

Table 7: Relationships between SVS values and MEM-derived value themes, Project 2.
Positive relationship: • = $R^2 \geq .01$, ● = $R^2 \geq .04$.
Negative relationship: ○ = $R^2 \geq .01$, ◯ = $R^2 \geq .04$.

Theme	**Example Words**
Achievement	Success, Courage, Achieve, Ability
Daily Routine	Dinner, Sleep, Shower, Nap, Laundry
Going to Events	Ticket, Event, Contact, Free, Tonight
Wonderful	Sky, Dream, Heart, Soul, Star
Student Responsibility	Class, Study, Paper, Homework, Exam
Recreation Planning	Weekend, Flight, Beach, Summer
Religiosity	Lord, Jesus, Bless, Worship, Pray
Eating & Cooking	Soup, Sandwich, Pizza, Delicious, Cooking
Fun Personality	Cute, Loveable, Funny, Goofy
Anticipation	Amaze, Excite, Birthday, Tomorrow
Sports	Team, Game, Win, Baseball, Football
Celebration	Birthday, Christmas, Anniversary
Swearing	Ass, Bitch, Dick, Fucker
Internet Movies	Watch, Movie, YouTube, Episode
Privacy Declaration	Settings, Information, Account, Privacy
Nationalism	Liberty, America, Nation, Flag, Unite
Parental Protection	Childhood, Violence, Campaign, Abuse
Cancer Support	Cancer, Patient, Cure, Illness
Musicianship	Band, Guitar, Rehearsal, Perform
Friendship Gratitude	Cherish, Friendship, Post
Farmville	Farmville, Stable, Barn, Gift
Group Success	Succeed, Hug, Cheer
Web Links	HTTP, ORG, PHP
Concern for Underprivileged	Elderly, Homeless, Veteran
Proselytizing	Deny, Believer, Christ, Heaven
Celebrity Concerns	Marriage, Britney, Spears, Jesse
Severe Weather	Severe, Thunderstorm, Tornado, Warning

Table 8: Themes extracted using the MEM on Facebook status updates.

	Achievement	Daily Routine	Going to Events	Wonderful	Student Resp.	Recreation Planning	Religiosity	Eating/Cooking	Fun Personality	Anticipation	Sports	Celebrations	Swearing	Internet/Movies	Privacy Concerns	Nationalism	Parental Protection	Cancer Support	Musicianship	Friendship	Farmville	Group Success	Web Links	Underprivileged	Proselytizing	Celebrity Concerns	Severe Weather
Schwartz Values																											
Achievement							∘						•												∘		
Benevolence																											
Conformity							•					•										•			•		
Hedonism	∘						∘						•												∘		
Power																											
Security																			∘						•		
Self-Direction																											
Stimulation							∘					∘															
Tradition							●					•													•		
Universalism																											
MEM Values																											
Religion	•						●					●													●		
Empathy	•			•					•								•	•		•							
FamilyGrowth		●				•	●	•		●		●						•		•	•			•			
Work		●			•	●				●		●															
DecisionMaking	•	•		•																							
Honesty	●	∘		●		∘		∘		∘																	
NegativeReligion	•			•																							
Social	●			●	∘		●		•	•		●						•		•		●			•		
Growth	●			●		∘	•	∘																			
Indulgence	•	●				•	•			•		●			•		•	•		•	•			•			
CaringKnowledge	●	∘		●	∘	∘	•	∘										•		•					•		
Openness	•			•			•		•	•		●										•					
KnowledgeGain	•			•			∘					∘								∘							
Principles		○			○	○		∘		○		○															
Freedom	•		•				•					•															
Certainty																									•		

Table 9: Coverage of behavior MEM themes by SVS values and value MEM themes, Project 2.
Positive relationship: • = $R^2 \geq .01$, ● = $R^2 \geq .04$. Negative relationship: ∘ = $R^2 \geq .01$, ○ = $R^2 \geq .04$.

Finally, we performed an analysis parallel to that described for Project 1 in order to explore the degree to which the language-derived value themes and SVS value scores corresponded to the self-described behaviors and ideas present in Facebook users' status updates. We emphasize two primary aspects of the results, presented in Table 9. First, we again see a conceptual replication of Project 1 in terms of value-behavior relationships. Scores from the SVS appear to show little correspondence with the actual behaviors and ideas that our sample of Facebook users share with others, whereas language-derived values show considerable and consistent relationships with behavioral topics. Second, whereas the SVS appears to correspond to rather narrow bands of behavioral themes, the language-derived values show extensive coverage of behaviors in predictive terms. In other words, the results from Project 2 not only conceptually replicate the results from Project 1, but demonstrate the applicability of the language-derived value themes to a completely new set of themes pertaining to the common thoughts and behaviors of social media users in the real world.

Conclusions

We have collected and analyzed one new, crowd-sourced dataset and one archival social media user dataset in order to better understand the relationships between people's values and their behaviors using a natural language processing approach. We found that the widely-adopted set of values that are measured by the SVS provide substantially less predictive coverage of real-world behaviors than a set of values extracted from people's own descriptions. Simply asking people what is important to them turns out to be a more informative method for answering the question of what values are, and the simple word counting approach appears to be a viable method for value quantification. Using this approach, we examined a large-scale social media data set to explore whether the language of values would continue to exhibit relationships with the ideas and behaviors that people share in their Facebook status updates. Results offer consistently strong support for language-based value–behavior links.

It is our hope that this study will open more doors to future work in values research. A new set of values has been identified, along with a method that allows for the simple, intuitive lexical representation of values. These methods can be used to study the values of various groups of people across various platforms, languages, time, and space. We note that this approach requires that a large enough body of text be collected for successful research. However, this is easily achieved by using more social media data, blog data, and other forms of prevalent data available in the current big data atmosphere. This approach may also facilitate further exploration of the relationships that exist between values and behavior by encouraging more fine-grained computational models.

Beyond Values

We have shown here a single case in which natural language data provided a more clear picture of people's cognitive and behavioral processes than data collected from a traditional and widely used self-report survey. Additionally, we have demonstrated that the information extracted from natural language exhibited more links (both in terms of quantity and diversity) with behaviors and thoughts than a standardized self-report measure. However, we advocate that the general approach that we have used for the current studies can also be applied much more generally. Indeed, many of the social and psychological phenomena studied using social media are conceptually abstract and difficult to distill into valid metrics. While the standard approach to studying such phenomena is to rely on gathering self-report data in the form of forced-choice questionnaires, this process often requires the collection of data beyond what is already available via social media and may often serve as insufficient "ground truth" when attempting to capture psychology as it exists in the real world.

As described in the current work, we emphasize that already-existing, organically generated social media data can exhibit greater predictive strength for human behaviors and a more dynamic structure than that imposed by closed, forced-choice questionnaires. Additionally, data at the "big data" level are often only available in the form of natural language. In such cases, we have demonstrated that psychological "ground truth" can still be attained, allowing researchers to explore human psychology under conditions where diverse forms of data are unavailable. Finally, the methods described here allow for the inference of many different psychological phenomena from the same data, including the core three components of human psychology (i.e., affect, cognition, and behavior). It is our aim to demonstrate with the work presented here that language is an incredibly flexible form of data that can be used to many great purposes.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation award #1344257, by grant #48503 from the John Templeton Foundation, and by award #W5J9CQ12C0043 from the Army Research Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, or the Army Research Institute.

References

Aronson, E. 2004. *The Social Animal*. New York, New York, USA: Worth Publishers, 9th edition.

Ball-Rokeach, S.; Rokeach, M.; and Grube, J. W. 1984. *The Great American Values Test: Influencing Behavior and Belief Through Television*. New York, New York, USA: Free Press.

Boyd, R. L. 2014a. MEH: Meaning Extraction Helper.

Boyd, R. L. 2014b. RIOT Scan: Recursive Inspection of Text Scanner.

Bubeck, M., and Bilsky, W. 2004. Value structure at an early age. *Swiss Journal of Psychology* 63:31–41.

Butenko, T., and Schwartz, S. 2013. Relations of the new circle of 19 values to behaviors. Technical report, National Research University Higher School of Economics.

Caprara, G. V.; Schwartz, S.; Capanna, C.; Vecchione, M.; and Barbaranelli, C. 2006. Personality and politics: Values, traits, and political choice. *Political Psychology* 27:1–28.

Chung, C. K., and Pennebaker, J. W. 2008. Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of Research in Personality* 42:96–132.

Chung, C. K., and Pennebaker, J. W. 2014. Finding values in words: Using natural language to detect regional variations in personal concerns. In *Geographical psychology: Exploring the interaction of environment and behavior*. 195–216.

Cieciuch, J.; Schwartz, S. H.; and Vecchione, M. 2013. Applying the refined values theory to past data: What can researchers gain? *Journal of Cross-Cultural Psychology* 44(8):1215–1234.

Kolb, J., and Kolb, J. 2013. *The Big Data Revolution*. CreateSpace Independent Publishing Platform.

Kosinski, M.; Stillwell, D.; and Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110(15):5802–5805.

Kramer, A. D., and Chung, C. K. 2011. Dimensions of self-expression in facebook status updates. *ICWSM*.

Kristiansen, C. M., and Zanna, M. P. 1988. Justifying attitudes by appealing to values: A functional perspective. *British Journal of Social Psychology* 27(3):247–256.

Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabasi, A.-L.; Brewer, D.; and Christakis, N. 2009. Computational social science. *Science* 323(5915):721–723.

Lepley, R. 1957. *The Language of Value*. New York: Columbia University Press.

Ramirez-Esparza, N.; Chung, C. K.; Kacewicz, E.; and Pennebaker, J. W. 2008. The psychology of word use in depression forums in english and in spanish: Texting two text analytic approaches. In *ICWSM*.

Rohan, M. J. 2000. A Rose by Any Name? The Values Construct. *Personality and Social Psychology Review* 4(3):255–277.

Rokeach, M. 1968. *Beliefs, Attitudes, and Values.*, volume 34. San Francisco: Jossey-Bass.

Rokeach, M. 1973. *The Nature of Human Values*, volume 70.

Schwartz, S. H., and Huismans, S. 1995. Value Priorities and Religiosity in Four Western Religions. *Social Psychology Quarterly* 58:88.

Schwartz, S. H.; Cieciuch, J.; Vecchione, M.; Davidov, E.; Fischer, R.; Beierlein, C.; Ramos, A.; Verkasalo, M.; Lönnqvist, J.-E.; Demirutku, K.; Dirilen-Gumus, O.; and Konty, M. 2012. Refining the theory of basic individual values. *Journal of Personality and Social Psychology* 103(4):663–688.

Schwartz, S. H. 1992. Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. *Advances in Experimental Social Psychology* 25:1–65.

Schwartz, S. H. 1994. Beyond individualism/collectivism: New cultural dimensions of values. In *Individualism and Collectivis*. 85–119.

Schwartz, S. H. 2004. Mapping and interpreting cultural differences around the world. In Vinken, H.; Soeters, J.; and Ester, P., eds., *Comparing Cultures, Dimensions of Culture in a Comparative Perspective*. Leiden, the Netherlands: Brill. 43–73.

Schwartz, S. H. 2009. Draft users manual: Proper use of the schwartz value survey, version 14.

Strapparava, C., and Mihalcea, R. 2008. Learning to identify emotions in text. *Proceedings of the 2008 ACM symposium on Applied computing - SAC '08* 1556.

Yarkoni, T. 2010. Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality* 44:363–373.

Zhang, Y.; Chen, W.; Wang, D.; and Yang, Q. 2011. User-click modeling for understanding and predicting search-behavior. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* 1388–1396.