PREDICTING INDIVIDUAL WELL-BEING THROUGH THE LANGUAGE OF SOCIAL MEDIA

H. ANDREW SCHWARTZ
Stony Brook University, Computer Science
Stony Brook, NY 11794

MAARTEN SAP, MARGARET L. KERN*, JOHANNES C. EICHSTAEDT, ADAM KAPELNER†, MEGHA AGRAWAL, EDUARDO BLANCO‡, LUKASZ DZIURZYNSKI, GREGORY PARK, DAVID STILLWELL§, MICHAL KOSINSKI¶, MARTIN E.P. SELIGMAN, AND LYLE H. UNGAR
University of Pennsylvania,
Positive Psychology Center & Computer and Information Sciences
Philadelphia, PA 19104

We present the task of predicting *individual* well-being, as measured by a life satisfaction scale, through the language people use on social media. Well-being, which encompasses much more than emotion and mood, is linked with good mental and physical health. The ability to quickly and accurately assess it can supplement multi-million dollar national surveys as well as promote whole body health. Through crowd-sourced ratings of tweets and Facebook status updates, we create message-level predictive models for multiple components of well-being. However, well-being is ultimately attributed to people, so we perform an additional evaluation at the user-level, finding that a multi-level cascaded model, using both message-level predictions and user-level features, performs best and outperforms popular lexicon-based happiness models. Finally, we suggest that analyses of language go beyond prediction by identifying the language that characterizes well-being.

1. Introduction

As human beings, we desire "the good life". When the British Broadcasting Cooperation (BBC) asked 1,001 Britons what the prime objective of their government should be – "greatest happiness" or "greatest wealth" – 81% answered with happiness.[1] In other studies, an average of 69% of people globally rate well-being as more important than any other life outcome.[2]

Beyond its popular appeal, another reason to consider well-being is that it is linked with positive life outcomes, including health and longevity.[3–6] Although it is not clear if well-being *causes* good health, it provides an indication of healthier or riskier individual trajectories, long before health problems develop.[7] Thus, the focus on well-being offers a preventative approach to public and personal health, with important economic consequences.

Well-being is more than simply positive emotion or mood. Psychologists, organizations, and governments measuring well-being are now using multi-dimensional measures that include a range of factors including meaning in life, engagement in activities, and the state of one's relationships, in addition to positive emotion.[8] While some language analyses have explored "happiness" based on emotion or mood,[9–14] modeling the broader construct of well-being is a relatively unexplored task.

In this paper, we present the task of predicting well-being based on natural language use. We develop a system that predicts the *satisfaction with life* (an overall evaluation of well-being) of Facebook *users* based on simple lexical and topical features. However, since individual *messages* themselves are the units of expression, we investigate the use of message-level models to improve

*University of Melbourne, Graduate School of Education, Australia
†Queens College, Department of Mathematics
‡University of North Texas, Computer Science
§Psychometrics Centre, University of Cambridge
¶Stanford Business School

user-level predictions. The message data, which was easy to come by through Amazon's Mechanical Turk, allowed us to supplement our *satisfaction with life* data, as well as explore other aspects of well-being, *PERMA* (discussed below). Lastly, for a human-level attribute like well-being, insights toward greater understanding is potentially just as important as prediction. Toward this end, we identify topics that most strongly correlate with well-being as clues to achieving "the good life".

Our unique contributions include: (a) the introduction of the task of predicting *individual* well-being, (b) the finding of a two-level, message-to-user model to perform better than models based on either independently, (c) the development of annotated well-being data across various constructs. Further, (e) we provide an analysis of the linguistic features that we find most significantly associated to individual *satisfaction with life*, and (e) we also release a well-being language model available for researchers (available to download at `wwbp.org/data.html`).

2. Background: Well-Being

Despite a desire for "the good life",[‖] well-being has traditionally been measured *indirectly* as a lack of problems (e.g., lack of depression and psychological disorder, low crime/disease/poverty rates) or by economic prosperity (i.e., gross domestic product). Reference 15 aptly notes: "if our interest is in the good life, we must look explicitly at indices of human thriving" (p. 144). Government agencies around the world are beginning to shift their attention toward *directly* measuring well-being.[16] In 2011, the U.K. Office of National Statistics piloted four well-being questions in their annual national survey, and similar efforts are now underway in Australia, Canada, France, Mexico, South Africa, and other places around the world. These initiatives follow a long history of academics who have attempted to rethink the notion of progress, deemphasizing the sole reliance on economic indicators and arguing instead that the welfare of a nation must be understood more holistically, with consideration of aspects such as social belonging, meaning, and optimism in the population.[8,17,18]

Satisfaction with Life (*SWL*) is a well-established representation of well-being, representing a person's cognitive evaluation of their own life. Measures of life satisfaction have been used reliably for several decades, and are increasingly being utilized by governments and organizations around the world as informative social indicators for policy decisions.[19] It is assessed by asking people to indicate the extent to which they agree with statements such as "In most ways my life is close to my ideal".[20] *SWL* draws on the subjects' evaluative judgments and seems to be highly comparative both within nations and between nations.[21] Overall life satisfaction strongly correlates with other well-being domains, such as meaning in life, relationships, and emotions.

PERMA. Beyond an overall evaluation of well-being, other psychologists break well-being into separate domains.[8,22] In such "dashboard approaches" — just like the "state" of an airplane is not given by a single indicator but instead by a variety of different indicators (altitude, speed, heading, fuel consumption) — well-being is best measured as separate, correlated dimensions.[8,23–25] In his well-being theory, Ref. 8 suggests five major pillars that together contribute to a person's sense of well-being: *Positive Emotions, Engagement, Relationships, Meaning*, and *Accomplish-*

[‖] Although debatable by well-being theorists and philosophers, for the purposes of this work, we consider the "good life", "well-being", and "satisfaction with life" synonymous. "Happiness" at times is equated with well-being, but often is used to denote positive emotion/ mood alone, so we only use this term when referencing other works that use this term.

ment (*PERMA*). Other "dashboard approaches" seek to capture subtle psychological notions such as "autonomy" or "self-acceptance".[22] Although greater specificity could be delineated, we chose the *PERMA* constructs since they capture fairly explicit and often foregrounded ends people pursue (i.e., things discussed in social media).

Positive emotion includes positively valenced emotions such as joy, contentment, and excitement. *Engagement* is a multi-dimensional construct that includes behavioral, cognitive, and affective components. It can refer to involvement and participation in groups or activities, enthusiasm and interest in activities, commitment and dedication to work, and focused attention to tasks at hand.[26,27] For our purposes here, we define it in terms of passion and involvement in life, as opposed to apathy and boredom. *Relationships* (or positive relationships) includes trusting others, perceiving others as being there if needed, receiving social support, and giving to others.[28] Considerable evidence identifies the importance of positive relationships for supporting health, longevity, and other important life outcomes.[29] *Meaning* in life captures having a sense of purpose, significance, and understanding in life.[30,31] It can also include transcending the self, feeling a sense of connection to a higher power or purpose, and provides goals or a course of direction to follow.[32] *Accomplishment* is often defined in terms of awards, honors, and other objective markers of achievement.[25,33] For our purposes here, we focus on the subjective side, in terms of a personal sense of accomplishment. It includes a sense of mastery, perceived competence, and goal attainment.[34]

Both PERMA and SWL are typically measured through Likert scales,[35] where people are presented a statement (e.g. "I am satisfied with my life") and asked the extent to which they agree or disagree with it.

3. Related Tasks

Among others, predicting emotion, mood, personality, and classic sentiment analyses are related to our task. In this section, we provide a cursory review of analyses with social media and related approaches. To our knowledge, well-being prediction as more than positive emotion is a novel task, and considering *message*-level predictions to improve a *user*-level model has not been explored.

Although sentiment analysis usually takes place at the single document or sentence level,[36–39] some have explored multi-level approaches. For example, Ref. 40 utilized a cascaded model where sentence level subjectivity classifiers are used to determine whether a sentence should be included for document-level analysis. Reference 41 expanded the idea by incorporating a joint-model of sentence and document level sentiment annotations. We find this sentence-to-document prediction analogous to our multi-level approach. However, such efforts typically aim at predicting the sentiment of text; they do not model any feature at the *user* level (human attributes). Our average user writes 123 messages, far more than the number of sentences in the typical sentiment-analyzed texts.

Some sentiment tasks do address human-level attributes. For example, signals used in distant supervision,[42,43] where heuristics replace manual annotations (e.g., ":(" = negative polarity), somewhat captures people's emotional states. Reference 44 attempts to differentiate between the emotions of the writer and the reader of content on a microblogging platform with social network characteristics. More directly, some have looked at predicting emotion of text[9] or more specifically learned the language of happy and sad blogposts based on self-annotated moods in an online live journal.[10] Reference 45 identified thesaurus-based topics related to the emotional expressions from an English blog corpus. Reference 46 constructed models from a large corpora of world knowledge to identify the affective tone across six basic emotions in texts.

While mood and sentiment analysis aim at text annotations, there have been a few tasks looking at modeling human-level conditions by the language one uses. Work on personality follows the widely accepted Big-Five personality traits.[47] Reference 48 analyzes correlations with basic Facebook features (number of friends, photos, tags, likes, etc.) and Ref. 49 with LIWC[50] word categories (pronouns, cognitive process, etc.). Personality prediction efforts use LIWC categories coupled with other shallow features[51,52] and n-grams.[53] Other tasks include finding indicators of psychological health,[54,55] or predicting political orientation.[56,57] In 2015, a shared task was organized to detect if a Twitter user suffered from depression or PTSD.[58] While these works predict attributes at the user level, they do not incorporate message-level features.

Recently, Ref. 11 used a premade lexicon of positive and negative emotion words to measure "gross national happiness" and both Refs. 12 and 39 used MTurk ratings of individual words for positive or negative valence. Dodds' MTurk application served as a model for our MTurk data collection, but rather than asking workers to annotate individual words, we had whole messages annotated, thus putting words into context. References 11,12 and 39 demonstrated face validity (i.e., people are happier on the weekends or sad after celebrities die), though they lacked an empirical evaluation of the degree to which they accurately measure happiness. By using these approaches as baselines for *SWL* prediction, we provide an empirical evaluation here. Furthermore, in this paper we model well-being as more than emotion and mood.

4. Method

We build predictive models of well-being, as measured through the *satisfaction with life* (*SWL*)[20] and *PERMA*[25] scales. We describe message-level models, a user-level model, and then a cascaded model whereby message predictions inform the user-level predictions.

As the first work attempting the prediction of user-level *SWL* using lexical features, we explore a moderately sized and consistent feature space for both user-level and message-level models:

ngrams. We used both unigrams and bigrams as features in this task, which we extracted using an informal text tokenizer ** that handles social media content and markup such as emoticons. Trigrams were not included in order to keep the number of features smaller.

topics. We used the 2000 topics released by Ref. 59, created by running *latent dirichlet allocation* (*LDA*) over a set of 18 million Facebook status updates from the MyPersonality application.[60] These topics were derived from the same domain, and thus add a more coarse-grained Facebook lexical feature to our models. A user, u's, usage of a topic, t, was calculated as: $p(t|u) = \sum_{w \in words_u} p(t|w) * p(w|u)$, where $p(t|w)$ is the probability of a topic given a word (a value provided by the generated topic model) and $p(w|u)$ is a user's probability of mentioning word w. Additionally, beyond simply prediction utility, topics provide insight into the latent categories of language that characterize well-being.

lexica. We also included the manually developed categories of words from Linguistic Inquiry and Word Count (LIWC)[61] as well as the weighted lexica from Dodd's Hedonometer.[12] While LDA

** http://wwbp.org/data.html

topics provide a data-driven set of categorical features, these lexica provide features grounded in psychological and linguistic theory and human judgment. LIWC, in particular, was developed over decades with many iterations[61] and its *positive emotion* and *negative emotion* categories are widely used, including being used by Ref. 11 in his measure of "gross national happiness" (*GNH*) and by Ref. 62 to track diurnal mood variation. *GNH* along with Dodd's *Hedonometer* – both lexica – also function as baselines for our predictive models.

Each feature is included as binary (1 if mentioned at least once, 0 otherwise) as well as in relative frequency over a its message or user ($\frac{freq(feature)}{\sum_{word \in doc} freq(w)}$). This results in hundreds of thousands of features. Thus, to reduce the change of overfitting, we filtered out infrequent features defined as those used by less than 10% of users or in less than 0.1% of messages.

4.1. *Message-level Models.*

We explore models for finding expressions of both *SWL* and positive and negative expressions of the five *PERMA* components. The features described above are aggregated at the message-level, where *n-grams* are encoded as booleans (i.e., whether they exist or not in the message) and the others are encoded as frequencies (probability over all words in message). We then use Randomized Principal Component Analysis[63] (*RPCA*) to transform the space to a more manageable size for ridge regression. Specifically, we reduce the feature matrix to $\frac{1}{4} * train\_size$ components for all models utilizing more than that many features. The projection matrix from *RPCA* is stored as part of the model such that prediction / test data is transformed based on a projection matrix fit to training data. Over the training data, we tested other prediction algorithms, such as Lasso (L1 penalized) regression,[64] which works well with sparse data, but ridge regression with RPCA performed better. Given that we have annotations for SWL and both negative and positive aspects of *PERMA*, we train a model for each outcome, resulting in 11 regression models total.

We have released a version of the PERMA language model without RPCA in the form of a weighted lexicon, extracted using the method described in Ref. 65 (due to Facebook policy restrictions, we are not able to release the annotated messages).

4.2. *User-Level Models.*

Our basic user-level model fits ridge regression[66] of the ngram, topic, and lexicon features to *SWL* scores. Just like with message level models, we use *RPCA* to reduce the dimension of the ngram feature space. We also tried Lasso (L1 penalized) regression,[64] which works well with sparse data (such as ngrams), but *RPCA* and combined with Ridge Regression yielded better results.

4.3. *Cascaded Message-to-User Level Well-Being Prediction*

Although well-being is attributed to people, we believe it might be possible to capture expressions of it at the message level and to pass this information along to the user level. Our cascaded model aggregates predictions of all message-level attributes across all of each user's messages and incorporates them as the mean prediction across a user's messages. For instance, $SWL_{user} = \frac{1}{\#msgs} \sum_{msg \in user} SWL_{msg}$. This in turn becomes a feature supplied to the user-level model. For example, if we train message-level predictors for both polarities for each of the five domains of

PERMA, this results in 10 features within the cascaded model: 5 domains $*$ 2 polarities. User-level attributes were not distributed to the messages because our annotated messages did not have any user-level attributes. This same cascading concept could be used in reverse, when the goal is message-level prediction, rather than user-level attributes. All algorithms were carried out using their SciKit-Learn implementation.[67]

5. Data Acquisition

Message-Level Data. We used Amazon's Mechanical Turk (MTurk) to acquire *PERMA* and *SWL* annotations for 5,100 public Facebook status updates. The status updates were randomly selected from among 230 million public Facebook messages that contained at least 50% English words according to the ASpell Official English dictionary.[††]

On MTurk, the largest online task-based labor market, tasks are completed by an on-demand labor force composed of "Turkers".[68] We set up a MTurk task where workers received $0.01 per annotation. Upon entry, turkers completed a research consent and were shown a video that explained the well-being category they were rating. Upon passing a quiz testing their understanding of the concepts, they were qualified to annotate the particular component of well-being for which they were trained.

For each of the 10 *PERMA* components, turkers indicated the "extent to which [a message] expresses" the particular component, by using a slider with rating scale that ranged from "none" (0) to "very strongly" (6). For *SWL*, workers indicated their agreement that the message indicates life satisfaction (0 = strongly disagree, 3 = neutral, 6 = strongly agree). Some examples statuses can be seen in Tables 1 and 2.

We decided to opt for getting more messages rather than getting more ratings per messages, utilizing two ratings for most messages. A third rating was brought in for disagreements (defined as those outside 1 standard deviation, 2 points, of each other). Between the two initial ratings, intra-class correlations[69] were in the "moderate" to "substantial"[‡‡] agreement range (i.e., $.4$ to $.8$) for all but two domains; positive engagement and negative accomplishment were in the "fair" agreement range (i.e., $.2$ to $.4$) suggesting these categories were more difficult to annotate and thus necessitating a third rating to improve accuracy (less accurate ratings only make our task of improving user-level predictions more difficult). In the end we used the mean rating for each message as the gold-standard.

We found the PERMA categories each contain different but related information. Considering all positive PERMA domains and then all negative domains, intercorrelations ranged from 0.36 to 0.68 (Pearson's r). Consistent with the psychological literature,[71] the highest correlations were between positive relationships ($R+$) and positive emotion ($P+$). There was an inverse but weak relationship between the positive and negative dimensions of PERMA ($r = -0.04$ to -0.39), supporting the idea of the two polarities being orthogonal. Given this, it is possible that each of the 10 PERMA categories will contribute independent information for predicting well being.

[††] http://misc.aspell.net/wiki/English_Dictionaries; we plan to make this data available upon ethics board approval for public sharing.
[‡‡] range labels provided by Ref. 70

Table 1. Examples of statuses with contrastive values for each PERMA category.

Status Update	P^+	P^-	E^+	E^-	R^+	R^-	M^+	M^-	A^+	A^-
Celebrating this amazing day.. lmao.. first of many	6.0	0.0	3.0	0.0	4.5	0.0	3.5	0.5	5.5	0.0
I wanna thank GOD for letting me see another BDAY...LOVE YA BIG MAMA I KNO U SMILING DOWN ON ME!!!!	3.0	1.7	4.0	0.0	5.0	0.0	5.5	0.0	3.0	0.0
Goin to laundry mat got hella laundry to do uuuhhhh.......just did a major clean up take him out...take him out of the game already..	0.0	5.5	3.0	0.0	4.5	1.7	0.0	0.0	0.0	0.0
I have such an amazing bf he took good care of me at the hospital.which he always takes good care of me.Im so blessed to have him.	5.0	1.0	1.7	0.0	6.0	1.0	2.0	0.5	0.0	0.0

Table 2. Example statuses expressing both polarities for categories engagement, relationships, and meaning.

Status Update	Cont. Category			
I hate wen you watch a movie and the ending is sooo predictable :/	E^+	3.0	E^-	2.7
Just when I thought my whole world had crumbled into a million pieces, you came along and brought me crazy glue and band-aids.	R^+	2.7	R^-	2.0
Another FUN bright photoshoot coming up in my future! Cant WAIT! :)	M^+	2.5	M^-	1.3

User-Level Data. Our user-level data was acquired through the MyPersonality Facebook App[60] from users who agreed to share their status updates for research purposes. We focused on users who took the Satisfaction with Life scale[20] (SWL), which previous research has shown has high internal consistency (reported alphas range from .79 to .89), and moderate temporal stability (reported test-retest correlations over two month intervals range from .50-.82).

Mean *SWL* in our sample was 4.3 (on a 7-point scale), consistent with the mean *SWL* reported in North American college and adult samples (typically 4.8). A subset of our sample ($N = 157$) retook the *SWL* scale six months later, and the resulting test-retest correlation ($r = .62$) was similar to those reported in past studies.[72] Test-retest correlation forms an upper-bound on the predictive accuracy we should expect to get from our models.

We further refined the data to only include users' status updates made in the 6 months prior to their taking the *SWL* questionnaire and only from those who wrote at least 500 words in that time. This resulted in a dataset of 2,198 individuals, having collectively written 260,840 messages. The messages in this dataset and those from the message-level dataset are completely disjoint. Thus, when creating the cascaded model, the message-level models will not have observed any of the users' messages before, avoiding overfitting issues. Though the message-level data set is smaller in number of messages (5,100 versus 260,000 for the user-level one), our hope is that since it has more labels (5,100 messages versus 2,200 users) can supplement to user-level data to improve accuracy. Furthermore, the benefit of a cascaded model might become even greater if fewer labeled users are available.

6. Evaluation

We evaluated our message-level, user-level, and cascaded message-to-user predictive models over a corpora of Facebook status updates. The second and third evaluation are over a user-level corpus in which volunteers from Facebook took the *SWL* questionnaire and shared their status updates. The

SWL questionnaire, an accepted metric of well-being, is used as a gold-standard for evaluating models at the user-level. Each corpus was divided randomly into 80% training/development instances and 20% test. We then test whether our models do better than a baseline of happy/hedonic lexica,[11,12] and whether a cascaded message-to-user level model improves upon the pure user-level model.

Message-level Results. Table 3 shows the results of each message-level regression model, reported using the Pearson correlation coefficient (r) between the predicted score and the annotated score. The annotated messages were randomly divided into 80% training and development (4080 messages) and a 20% test set (1020 messages). Interestingly, for some categories, *topics* were better than *ngrams*, while for others *ngrams* were better than *topics*. We also noticed that positive relationships seems to be the easiest component to predict, higher than both positive emotion and *SWL*.

Table 3. Message-level prediction scores as the Pearson correlation coefficient (r) across the PERMA and SWL categories.

features	$\mathbf{P}^{+}$	$\mathbf{P}^{-}$	$\mathbf{E}^{+}$	$\mathbf{E}^{-}$	$\mathbf{R}^{+}$	$\mathbf{R}^{-}$	$\mathbf{M}^{+}$	$\mathbf{M}^{-}$	$\mathbf{A}^{+}$	$\mathbf{A}^{-}$	**SWL**
ngrams	0.563	0.412	0.349	0.347	0.604	0.365	0.405	0.279	0.363	0.285	0.479
topics	0.500	0.430	0.252	0.405	0.500	0.417	0.400	0.286	0.260	0.322	0.495
lexica	0.413	0.442	0.205	0.292	0.445	0.376	0.217	0.205	0.147	0.292	0.430
ngrams+topics	0.598	0.492	0.369	0.421	0.641	0.401	**0.451**	0.314	0.380	**0.398**	0.550
ngrams+lexica	0.613	0.526	0.346	0.375	0.621	0.446	0.435	**0.331**	**0.414**	0.361	0.543
topics+lexica	0.509	0.464	0.247	0.376	0.525	0.423	0.372	0.260	0.307	0.268	0.505
ngrams+topics+lexica	**0.617**	**0.504**	**0.374**	**0.422**	**0.655**	**0.427**	0.441	0.311	0.402	0.352	**0.566**

User-level and Cascaded Results. Both the user-level and cascaded message-to-user-level models were evaluated over the same user-level corpus, divided such that 80% (1758 users) was used for training and development while 20% was held out for testing (440 users). The scores represent the correlation (Pearson r) between predictions over the test data and the users' scores from the *SWL* scale. Since we are studying well-being, Pearson correlation allows us to frame the results with a metric used widely in social sciences. To put our results in perspective, subjective (user-level) psychological variables typically have a "correlational upper-bound" in the range of $r = 0.3$ to $r = 0.4$ with human behaviors such as language use.[73]

Table 4 shows the user-level predictions for all combinations of features as well as the cascaded models. We see that an ngram model out-performs baselines of methods used previously for happiness prediction,[11,12,39] while our best user-level results come from a combination of *ngrams*, *topics*, and the *lexica* (though they are not significantly better than *ngrams* and *topics* alone).

Cascading models significantly boost performance, increasing it from $.301$ with the user-level language features alone to $.333$ with cascaded models. This is quite surprising, considering the message-level models were only based on 5,100 messages, but there were over 200,000 messages across all the users.

Analyses of individual message-level predictors for each component of PERMA showed all message-level predictors add to the prediction, with out-of-sample correlations ranging from $r = .15$ to $r = .247$. All domains of *PERMA*, as modeled through the language of annotated messages, had an impact on user-level *SWL*.

Table 4. User-level prediction scores as the Pearson correlation coefficient (r) . *message predictions*: message-level regression feeding cascaded model; *user features*: all user level language features (*ngrams + topics + lexica*. bold: significant ($p < .05$; p is Bonferroni corrected for multiple comparisons[74]) improvement over user-level features alone.

user-level models	*ngrams*	.262
	topics	.254
	lexica	.198
	ngrams + topics	.299
	ngrams + lexica	.269
	topics + lexica	.252
	ngrams + topics + lexica	.301
baselines	(mean)	.000
	lexica: GNH	.210
	lexica: Hedonometer	.108
cascaded models	*message predictions* alone	.236
	user features alone	.301
	message predictions + user features	**.333**

7. Discussion: Well-Being Insights

The LDA topics most highly correlated with user-level SWL shed light on the mechanisms which may contribute to a person's satisfaction with their life in a way that is in line with the psychological literature. Figure 1 shows four of the top ten positively correlated topics, and the two leading negative ones. Several of the topics tap various aspects of engagement. *excited, super, tomorrow* references

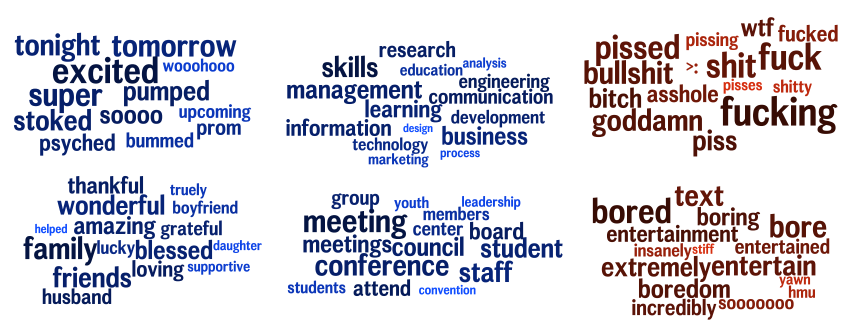

Fig. 1. The top 4 topics positively correlated (blue) and 2 topics negatively correlated (red) with *SWL*.

affective and psychological states which suggest that people are happily engrossed in activities of life.[75] *meeting, conference, staff* hints at communal engagement, which overlaps with involvement, dedication, and organizational citizenship behavior.[76] The *bored, bore, text* topic reflects the converse; disengagement merges as one of the strongest negative predictors of SWL. Engagement — a core component of $PERMA$[8] — is considered a key part of healthy aging.[77]

The *skills, management, business* topic corroborates theories that county level SWL is linked to employment in the "professional" occupation sector.[78] Theoretical psychology[79] suggests that people in high value-creation occupations, in which continuous learning, roles of responsibility and skill development are valued, would be more satisfied with their lives. The *family, friends, wonderful* topic supports the well-established idea that good relationships are a strong predictor of well-being.[29]The swearing topic emerges as the single strongest (negative) topic predictor of SWL.

8. Conclusions

We presented the task of predicting well-being, a multidimensional construct, based on natural language use. We developed predictive models of well-being, as measured through the *satisfaction with life*(*SWL*) scale, over Facebook volunteers. Our models significantly out-predict baselines of popular happy and hedonistic lexica.[11,12]

We created both message-level models as well as user-level models, and found a cascaded model, in which message-level predictions inform user-level predictions, gave the best performance. Additionally, we introduced corpora with annotated well-being data, and show that for such human-level information, language analysis can go beyond prediction and demonstrate insight into what leads to ”the good life”.

Well-being prediction is a worthwhile task for the social media mining community as the construct is gaining popularity and it is known to be linked with health, economics, and longevity. People and governments have started to recognize that economic measures alone do not capture the welfare of societies. Methodologically, there is much to explore, such as more sophisticated joint models of user and message-level information or the use of syntactic structure as features. Additionally, we suggest that language-based analyses of well-being need not end with *prediction*. Links between *SWL* and the everyday language in social-media enriches our *understanding* of well-being and its determinants, indicators and consequences.

Acknowledgment. We thank the Templeton Religion Trust for their generous support of this study.

References

1. M. Easton, *BBC News* **2** (2006).
2. E. Diener, *American psychologist* **55**, p. 34 (2000).
3. E. Diener and M. Y. Chan, *Applied Psychology: Health and Well-Being* **3**, 1 (2011).
4. R. T. Howell, M. L. Kern and S. Lyubomirsky, *Health Psychology Review* **1**, 83 (2007).
5. S. Lyubomirsky, L. King and E. Diener, *Psychological Bulletin* **131**, p. 803 (2005).
6. S. D. Pressman and S. Cohen, *Psychological bulletin* **131**, p. 925 (2005).
7. H. S. Friedman and M. L. Kern, *Psychology* **65**, p. 719 (2014).
8. M. E. P. Seligman, *Flourish: A Visionary New Understand of Happiness and Well-being* (Free Press, 2011).
9. C. Alm, D. Roth and R. Sproat, Emotions from text: machine learning for text-based emotion prediction, in *Proceedings of the conference on Empirical Methods in Natural Language Processing*, 2005.
10. R. Mihalcea and H. Liu, A corpus-based approach to finding happiness, in *Proceedings of the AAAI Spring Symposium on Computational Approaches to Weblogs*, 2006.
11. A. Kramer, An unobtrusive behavioral model of gross national happiness, in *Proc of the 28th int conf on Human factors in comp sys*, 2010.
12. P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss and C. M. Danforth, *Diversity* , p. 26 (2011).
13. J. Bollen, H. Mao and A. Pepe, Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena., in *ICWSM*, 2011.
14. S. M. Mohammad, #emotional tweets, in *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, 2012.
15. C. Peterson and N. Park, *Psychological Inquiry* **14**, 143 (2003).
16. J. E. Stiglitz, A. Sen and J.-P. Fitoussi, *The Commission on the Measurement of Economic Performance and Social Progress website. Available online at: http://www. stiglitz-sen-fitoussi* (2009).

17. M. Abramovitz, T. Scitovsky and A. Inkeles, *Bulletin of the American Academy of Arts and Sciences* , 11 (1973).
18. R. Layard, *Economics and Happiness* , 147 (2005).
19. E. Diener, R. Inglehart and L. Tay, *Social Indicators Research* , 1 (2012).
20. E. Diener, R. A. Emmons, R. J. Larsen and S. Griffin, *Journal of personality assessment* **49**, 71 (1985).
21. E. Diener, W. Ng, J. Harter and R. Arora, *Journal of Personality and Social Psychology* **99**, p. 52 (2010).
22. C. D. Ryff and C. L. M. Keyes, *Journal of personality and social psychology* **69**, p. 719 (1995).
23. J. E. Stiglitz, A. Sen and J.-P. Fitoussi, *The Commission on the Measurement of Economic Performance and Social Progress website. Available online at: http://www. stiglitz-sen-fitoussi* (2009).
24. B. S. Frey and A. Stutzer, *Happiness and economics: How the economy and institutions affect human well-being* (Princeton University Press, 2010).
25. M. J. C. Forgeard, E. Jayawickreme, M. Kern and M. E. P. Seligman, *International Journal of Wellbeing* **1**, 49 (2011).
26. J. J. Appleton, S. L. Christenson and M. J. Furlong, *Psychology in the Schools* **45**, 369 (2008).
27. M. Csikszentmihalyi, *Creativity: Flow and the psychology of discovery and invention* (Harper Perennial, 1997).
28. S. E. Taylor, *The Oxford Handbook of Health Psychology* , 189 (2011).
29. L. Tay, K. Tan, E. Diener and E. Gonzalez, *Applied Psychology: Health and Well-Being* (2012).
30. M. F. Steger, T. B. Kashdan, B. A. Sullivan and D. Lorentz, *Journal of Personality* **76**, 199 (2008).
31. V. E. Frankl, *Man's search for ultimate meaning* (Insight Books/Plenum Press, 1997).
32. S. M. Schueller and M. E. Seligman, *The Journal of Positive Psychology* **5**, 253 (2010).
33. K. A. Ericsson, *The pursuit of excellence through education* , 21 (2002).
34. L. I. Pearlin and C. Schooler, *Journal of health and social behavior* , 2 (1978).
35. R. Likert, *Archives of psychology* (1932).
36. P. Turney, Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews, in *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, (Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, July 2002).
37. B. Pang, L. Lee and S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.
38. C. Strapparava and R. Mihalcea, Semeval-2007 task 14: Affective text, in *Proceedings of the 4th International Workshop on Semantic Evaluations*, 2007.
39. S. M. Mohammad, S. Kiritchenko and X. Zhu, *arXiv preprint arXiv:1308.6242* (2013).
40. B. Pang and L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts, in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 2004.
41. R. McDonald, K. Hannan, T. Neylon, M. Wells and J. Reynar, Structured models for fine-to-coarse sentiment analysis, in *Annual Meeting-Association For Computational Linguistics*, (1)2007.
42. J. Read, Using emoticons to reduce dependency in machine learning techniques for sentiment classification, in *Proceedings of the ACL Student Research Workshop*, (Association for Computational Linguistics, Ann Arbor, Michigan, June 2005).
43. D. Davidov, O. Tsur and A. Rappoport, Enhanced sentiment learning using twitter hashtags and smileys, in *Coling 2010: Posters*, (Coling 2010 Organizing Committee, Beijing, China, August 2010).
44. Y.-j. Tang and H.-H. Chen, *Sentiment Analysis where AI meets Psychology (SAAIP)* , p. 11 (2011).
45. S. Aman and S. Szpakowicz, Identifying expressions of emotion in text, in *Text, Speech and Dialogue*, 2007.
46. C. Strapparava and R. Mihalcea, Learning to identify emotions in text, in *Proceedings of the 2008 ACM symposium on Applied computing*, 2008.
47. R. R. McCrae and O. P. John, *Journal of Personality* **60**, 175 (1992).

48. Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli and D. Stillwell, *Web Science* (2012).
49. T. Yarkoni, *Journal of Research in Personality* **44**, 363 (June 2010).
50. J. W. Pennebaker, M. E. Francis and R. J. Booth, *Word J. Of The Int Ling Assoc* (2001).
51. F. Mairesse, M. Walker, M. Mehl and R. Moore, *Journal of Artificial Intelligence Research* **30**, 457 (2007).
52. A. Gill, S. Nowson and J. Oberlander, *Proc. of AAAI ICWSM* (2009).
53. S. Nowson, Identifying more bloggers: Towards large scale personality classification of personal weblogs, in *In Proceedings of the International Conference on Weblogs and Social*, 2007.
54. S. Saleem, R. Prasad, S. Vitaladevuni, M. Pacula, M. Crystal, B. Marx, D. Sloan, J. Vasterling and T. Speroff, Automatic detection of psychological distress indicators and severity assessment from online forum posts, in *Proceedings of COLING 2012*, (The COLING 2012 Organizing Committee, Mumbai, India, December 2012).
55. M. De Choudhury, M. Gamon, S. Counts and E. Horvitz, Predicting depression via social media, in *ICWSM*, 2013.
56. D. Rao, D. Yarowsky, A. Shreevats and M. Gupta, Classifying latent user attributes in twitter, in *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 2010.
57. M. Pennacchiotti and A.-M. Popescu, *ICWSM* **11**, 281 (2011).
58. G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead and M. Mitchell, Clpsych 2015 shared task: Depression and ptsd on twitter, in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver, Colorado, USA, June. North American Chapter of the Association for Computational Linguistics*, 2015.
59. H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Seligman *et al.*, *PloS one* **8**, p. e73791 (2013).
60. M. Kosinski and D. Stillwell, mypersonality project, in *http://www.mypersonality.org/wiki/*, 2012.
61. J. W. Pennebaker, C. Chung, M. Ireland, A. Gonzales and R. Booth, *Austin, TX, LIWC. Net* (2007).
62. S. Golder and M. Macy, *Science* **333**, 1878 (2011).
63. N. Halko, P.-G. Martinsson and J. A. Tropp, *SIAM review* **53**, 217 (2011).
64. R. Tibshirani, *Journal of the Royal Statistical Society. Series B (Methodological)* , 267 (1996).
65. M. Sap, G. Park, J. Eichstaedt, M. Kern, L. Ungar and H. A. Schwartz, Developing age and gender predictive lexica over social media, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP*, 2014.
66. A. Hoerl and R. Kennard, *Technometrics* **12**, 55 (1970).
67. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, *The Journal of Machine Learning Research* **12**, 2825 (2011).
68. D. Chandler and A. Kapelner, *Arxiv* (2010).
69. P. E. Shrout and J. L. Fleiss, *Psychological bulletin* **86**, p. 420 (1979).
70. J. R. Landis and G. G. Koch, *Biometrics* , 159 (1977).
71. E. Diener and M. E. Seligman, *Psychological Science* **13**, 81 (2002).
72. W. Pavot and E. Diener, *Psychological Assessment* **5**, p. 164 (1993).
73. G. Meyer, S. Finn, L. Eyde, G. Kay, K. Moreland, R. Dies, E. Eisman, T. Kubiszyn and G. Read, *American psychologist* **56**, p. 128 (2001).
74. O. J. Dunn, *Journal of the American Statistical Association* **56**, 52 (1961).
75. W. B. Schaufeli, A. B. Bakker and M. Salanova, *Educational and psychological Measurement* **66**, 701 (2006).
76. J. J. Appleton, S. L. Christenson, D. Kim and A. L. Reschly, *Journal of School Psychology* **44**, 427 (2006).
77. J. W. Rowe and R. L. Kahn, *Science* (1987).
78. N. M. Lawless and R. E. Lucas, *Social Indicators Research* **101**, 341 (2011).
79. R. M. Ryan and E. L. Deci, *American psychologist* **55**, p. 68 (2000).