Human language reveals a universal positivity bias

Peter Sheridan Dodds,[1,2,*] Eric M. Clark,[1,2] Suma Desu,[3] Morgan R. Frank,[1,2] Andrew J. Reagan,[1,2] Jake Ryland Williams,[1,2] Lewis Mitchell,[1,2] Kameron Decker Harris,[4] Isabel M. Kloumann,[5] James P. Bagrow,[1,2] Karine Megerdoomian,[6] Matthew T. McMahon,[6] Brian F. Tivnan,[6,2,†] and Christopher M. Danforth[1,2,‡]

[1] *Computational Story Lab, Vermont Advanced Computing Core, & the Department of Mathematics and Statistics, University of Vermont, Burlington, VT, 05401*
[2] *Vermont Complex Systems Center, University of Vermont, Burlington, VT, 05401*
[3] *Center for Computational Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139*
[4] *Applied Mathematics, University of Washington, Lewis Hall #202, Box 353925, Seattle, WA, 98195.*
[5] *Center for Applied Mathematics, Cornell University, Ithaca, NY, 14853.*
[6] *The MITRE Corporation, 7525 Colshire Drive, McLean, VA, 22102*

(Dated: June 17, 2014)

Using human evaluation of 100,000 words spread across 24 corpora in 10 languages diverse in origin and culture, we present evidence of a deep imprint of human sociality in language, observing that (1) the words of natural human language possess a universal positivity bias; (2) the estimated emotional content of words is consistent between languages under translation; and (3) this positivity bias is strongly independent of frequency of word usage. Alongside these general regularities, we describe inter-language variations in the emotional spectrum of languages which allow us to rank corpora. We also show how our word evaluations can be used to construct physical-like instruments for both real-time and offline measurement of the emotional content of large-scale texts.

Human language—our great social technology—reflects that which it describes through the stories it allows to be told, and us, the tellers of those stories. While language's shaping effect on thinking has long been controversial [1–3], we know that a rich array of metaphor encodes our conceptualizations [4], word choice reflects our internal motives and immediate social roles [5–7], and the way a language represents the present and future may condition economic choices [8].

In 1969, Boucher and Osgood framed the Pollyanna Hypothesis: a hypothetical, universal positivity bias in human communication [9]. From a selection of small-scale, cross-cultural studies, they marshaled evidence that positive words are likely more prevalent, more meaningful, more diversely used, and more readily learned. However, in being far from an exhaustive, data-driven analysis of language—the approach we take here—their findings could only be regarded as suggestive. Indeed, studies of the positivity of isolated words and word stems have produced conflicting results, some pointing toward a positivity bias [10], others the opposite [11, 12], though attempts to adjust for usage frequency tend to recover a positivity signal [13].

To deeply explore the positivity of human language, we constructed 24 corpora spread across 10 languages (see Supplementary Online Material). Our global coverage of linguistically and culturally diverse languages includes English, Spanish, French, German, Brazilian Portuguese, Korean, Chinese (Simplified), Russian, Indonesian, and Arabic. The sources of our corpora are similarly broad, spanning books [14], news outlets, social media, the web [15], television and movie subtitles, and music lyrics [16]. Our work here greatly expands upon our earlier study of English alone, where we found strong evidence for a usage-invariant positivity bias [17].

We address the social nature of language in two important ways: (1) we focus on the words people most commonly use, and (2) we measure how those same words are received by individuals. We take word usage frequency as the primary organizing measure of a word's importance. Such a data-driven approach is crucial for both understanding the structure of language and for creating linguistic instruments for principled measurements [18, 19]. By contrast, earlier studies focusing on meaning and emotion have used 'expert' generated word lists, and these fail to statistically match frequency distributions of natural language [10–12, 20], confounding attempts to make claims about language in general. For each of our corpora we selected between 5,000 to 10,000 of the most frequently used words, choosing the exact numbers so that we obtained approximately 10,000 words for each language.

We then paid native speakers to rate how they felt in response to individual words on a 9 point scale, with 1 corresponding to most negative or saddest, 5 to neutral, and 9 to most positive or happiest [10, 18] (see also Supplementary Online Material). This happy-sad semantic differential [20] functions as a coupling of two standard 5-point Likert scales. Participants were restricted to certain regions or countries (for example, Portuguese was rated by residents of Brazil). Overall, we collected 50 ratings per word for a total of around 5,000,000 individual human assessments, and we provide all data sets as part of the Supplementary Online Material.

In Fig. 1, we show distributions of the average happi-

* peter.dodds@uvm.edu
† btivnan@mitre.org
‡ chris.danforth@uvm.edu

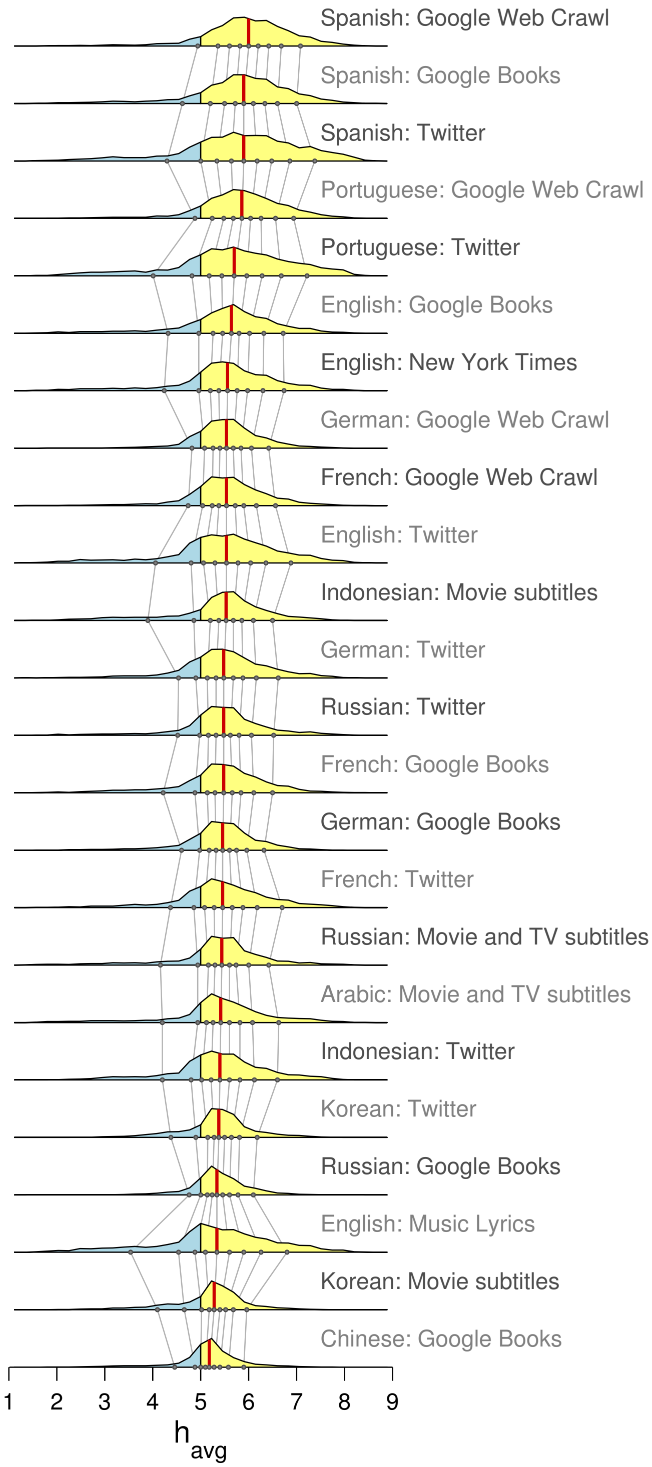

FIG. 1. Distributions of perceived average word happiness $h_{\rm avg}$ for 24 corpora in 10 languages. The histograms represent the 5000 most commonly used words in each corpora (see Supplementary Online Material for details), and native speakers scored words on a 1 to 9 double-Likert scale with 1 being extremely negative, 5 neutral, and 9 extremely positive. Yellow indicates positivity ($h_{\rm avg} > 5$) and blue negativity ($h_{\rm avg} < 5$), and distributions are ordered by increasing median (red vertical line). The background grey lines connect deciles of adjacent distributions. Fig. S1 shows the same distributions arranged according to increasing variance.

ness scores for all 24 corpora, leading to our most general observation of a clear positivity bias in natural language. We indicate the above neutral part of each distribution with yellow, below neutral with blue, and order the distributions moving upwards by increasing median (vertical red line). For all corpora, the median clearly exceeds the neutral score of 5. The background gray lines connect deciles for each distribution. In Fig. S1, we provide the same distributions ordered instead by increasing variance.

As is evident from the ordering in Figs. 1 and S1, while a positivity bias is the universal rule, there are minor differences between the happiness distributions of languages. For example, Latin American-evaluated corpora (Mexican Spanish and Brazilian Portuguese) exhibit relatively high medians and, to a lesser degree, higher variances. For other languages, we see those with multiple corpora have more variable medians, and specific corpora are not ordered by median in the same way across languages (e.g., Google Books has a lower median than Twitter for Russian, but the reverse is true for German and English). In terms of emotional variance, all four English corpora are among the highest, while Chinese and Russian Google Books seem especially constrained.

We now examine how individual words themselves vary in their average happiness score between languages. Owing to the scale of our corpora, we were compelled to use an online service, choosing Google Translate. For each of the 45 language pairs, we translated isolated words from one language to the other and then back. We then found all word pairs that (1) were translationally-stable, meaning the forward and back translation returns the original word, and (2) appeared in our corpora for each language.

We provide the resulting comparison between languages at the level of individual words in Fig. 2. We use the mean of each language's word happiness distribution derived from their merged corpora to generate a rough overall ordering, acknowledging that frequency of usage is no longer meaningful, and moreover is not relevant as we are now investigating the properties of individual words. Each cell shows a heat map comparison with word density increasing as shading moves from gray to white. The background colors reflect the ordering of each pair of languages, yellow if the row language had a higher average happiness than the column language, and blue for the reverse. In each cell, we display the number of translation-stable words between language pairs, N, along with the difference in average word happiness, Δ, where each word is equally weighted.

A linear relationship is clear for each language-language comparison, and is supported by Pearson's correlation coefficient r being in the range 0.73 to 0.89 (p-value $< 10^{-118}$ across all pairs; see Fig. 2 and Tabs. S3, S4, and S5). Overall, this strong agreement between languages, previously observed on a small scale for a Spanish-English translation [21], suggests that approximate estimates of word happiness for unscored languages

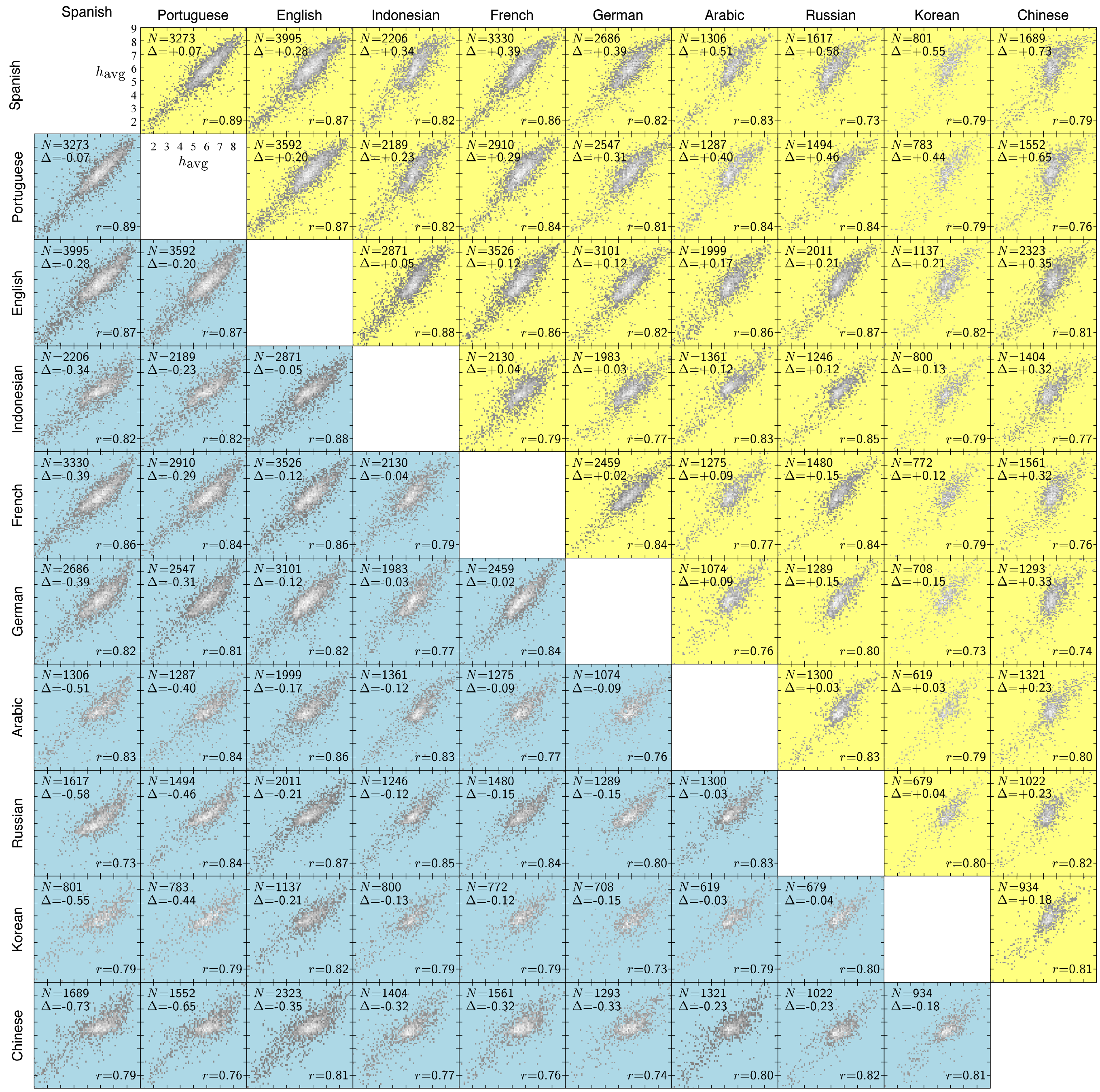

FIG. 2. Scatter plots of average happiness for words measured in different languages. We order languages from relatively most positive (Spanish) to relatively least positive (Chinese); a yellow background indicates the row language is more positive than the column language, and a blue background the converse. The overall plot matrix is symmetric about the leading diagonal, the redundancy allowing for easier comparison between languages. In each scatter plot, the key gives the number of translation-stable words for each language pair, N; the average difference in translation-stable word happiness between the row language and column language, Δ; and the Pearson correlation coefficient for the regression, r. All p-values are less than $10 and less than 10$ Fig. S2 shows histograms of differences in average happiness for translation-stable words.

could be generated with no expense from our existing data set. Some words will of course translate unsatisfactorily, with the dominant meaning changing between languages. For example 'lying' in English, most readily interpreted as speaking falsehoods by our participants, translates to 'acostado' in Spanish, meaning recumbent. Nevertheless, happiness scores obtained by translation will be serviceable for purposes where the effects of many

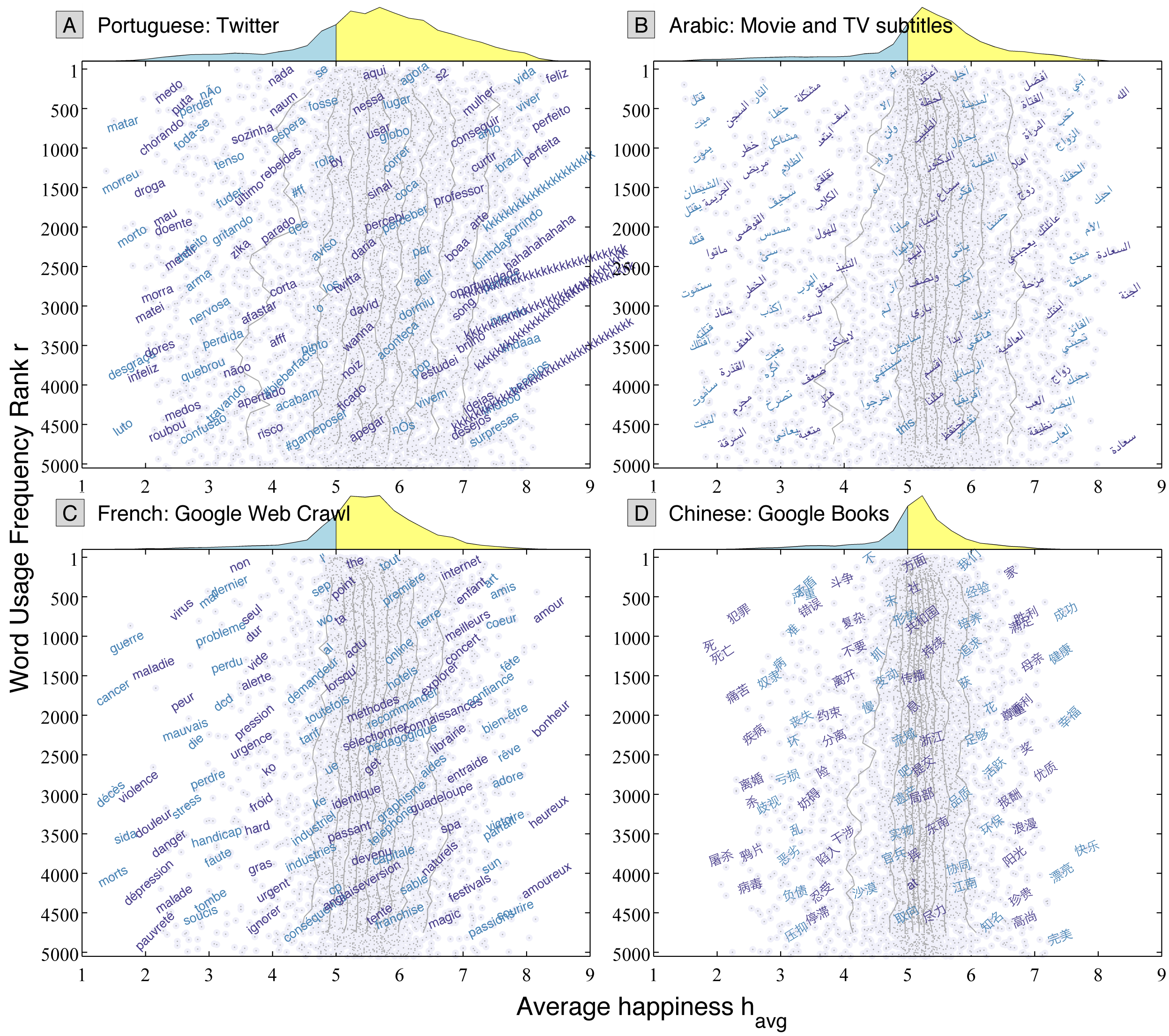

FIG. 3. Examples of how word happiness varies little with usage frequency. Above each plot is a histogram of average happiness h_{avg} for the 5000 most frequently used words in the given corpus, matching Fig. 1. Each point locates a word by its rank r and average happiness h_{avg}, and we show some regularly spaced example words. The descending gray curves of these jellyfish plots indicate deciles for windows of 500 words of contiguous usage rank, showing that the overall histogram's form is roughly maintained at all scales. The 'kkkkkk...' words represent laughter in Brazilian Portuguese, in the manner of 'hahaha...'. See Fig. S3 for an English translation, Figs. S4–S7 for all corpora, and Figs. S8–S11 for the equivalent plots for standard deviation of word happiness scores.

different words are incorporated. (See the Supplementary Online Material for links to an interactive visualization of Fig. 2.)

Stepping back from examining inter-language robustness, we return to a more detailed exploration of the rich structure of each corpus's happiness distribution. In Fig. 3, we show how average word happiness h_{avg} is largely independent of word usage frequency for four example corpora. We first plot usage frequency rank r of the 5000 most frequently used words as a function of their average happiness score, h_{avg} (background dots), along with some example evenly-spaced words. (We note that words at the extremes of the happiness scale are ones evaluators agreed upon strongly, while words near neutral range from being clearly neutral (e.g., $h_{\rm avg}$(‘the’)=4.98) to contentious with high standard deviation [17].) We then compute deciles for contiguous sets of 500 words, sliding this window through rank r. These deciles form the vertical strands. We overlay randomly chosen, equally-spaced example words to give a sense of each corpus's emotional texture.

We chose the four example corpora shown in Fig. 3 to be disparate in nature, covering diverse languages (French, Egyptian Arabic, Brazilian Portuguese, and Chinese), regions of the world (Europe, the Middle East, South America, and Asia), and texts (Twitter, movies and television, the Web [15], and books [14]). In the Supplementary Online Material, we show all 24 corpora yield similar plots (see Figs. S4–S7 and English translated versions, Figs. S12–S15). We also show how the standard deviation for word happiness exhibits an approximate self-similarity (Figs. S8–S11 and their translations, Figs. S16–S19).

Across all corpora, we observe visually that the deciles tend to stay fixed or move slightly toward the negative, with some expected fragility at the 10% and 90% levels (due to the distributions' tails), indicating that each corpus's overall happiness distribution approximately holds independent of word usage. In Fig. 3, for example, we see that both the Brazilian Portuguese and French examples show a small shift to the negative for increasingly rare words, while there is no visually clear trend for the Arabic and Chinese cases. Fitting $h_{\rm avg} = \alpha r + \beta$ typically returns α on the order of -1$\times 10^{-5}$ suggesting $h_{\rm avg}$ decreases 0.1 per 10,000 words. For standard deviations of happiness scores (Figs. S8–S11), we find a similarly weak drift toward higher values for increasingly rare words (see Tabs. S6 and S7 for correlations and linear fits for $h_{\rm avg}$ and $h_{\rm std}$ as a function of word rank r for all corpora). We thus find that, to first order, not just the positivity bias, but the happiness distribution itself applies for common words and rare words alike, revealing an unexpected addition to the many well known scalings found in natural language, famously exemplified by Zipf's law [22].

In constructing language-based instruments for measuring expressed happiness, such as our hedonometer [18], this frequency independence allows for a way to ‘increase the gain’ in a way resembling that of standard physical instruments. Moreover, we have earlier demonstrated the robustness of our hedonometer for the English language, showing, for example that measurements derived from Twitter correlate strongly with Gallup well-being polls and related indices at the state and city level for the United States [19].

Here, we provide an illustrative use of our hedonometer in the realm of literature, inspired by Vonnegut's shapes of stories [23, 24]. In Fig. 4, we show ‘happiness time series’ for three famous works of literature, evaluated in their original languages English, Russian, and French: **A.** Melville's Moby Dick [25], **B.** Dostoyevsky's Crime and Punishment [26], and **C.** Dumas' Count of Monte Cristo [25]. We slide a 10,000-word window through each work, computing the average happiness using a ‘lens’ for the hedonometer in the following manner. We capitalize on our instrument's tunablility to obtain a strong signal by excluding all words for which $3 < h_{\rm avg} < 7$, i.e., we keep words residing in the tails of each distribution [18]. Denoting a given lens by its corresponding set of allowed words L, we estimate the happiness score of any text T as $h_{\rm avg}(T) = \sum_{w\in L} f_w h_{\rm avg}(w) / \sum_{w\in L} f_w$ where f_w is the frequency of word w in T [27].

The three resulting happiness time series provide interesting, detailed views of each work's narrative trajectory revealing numerous peaks and troughs throughout, at times clearly dropping below neutral. Both Moby Dick and Crime and Punishment end on low notes, whereas the Count of Monte Cristo culminates with a rise in positivity, accurately reflecting the finishing arcs of all three. The ‘word shifts’ overlaying the time series compare two distinct regions of each work, showing how changes in word abundances lead to overall shifts in average happiness. Such word shifts are essential tests of any sentiment measurement, and are made possible by the linear form of our instrument [18, 27] (see pp. S25–S27 in the Supplementary Online Material for a full explanation). As one example, the third word shift for Moby Dick shows why the average happiness of the last 10% of the book is well below that of the first 25%. The major contribution is an increase in relatively negative words including ‘missing’, ‘shot’, ‘poor’, ‘die’, and ‘evil’. We include full diagnostic versions of all word shifts in Figs. S21–S34.

By adjusting the lens, many other related time series can be formed such as those produced by focusing on only positive or negative words. Emotional variance as a function of text position can also be readily extracted. In the Supplementary Online Material, we provide links to online, interactive versions of these graphs where different lenses and regions of comparisons may be easily explored. Beyond this example tool we have created here for the digital humanities and our hedonometer for measuring population well-being, the data sets we have generated for the present study may be useful in creating a great variety of language-based instruments for assessing emotional expression.

Overall, our major scientific finding is that when experienced in isolation and weighted properly according to usage, words—the atoms of human language—present an emotional spectrum with a universal, self-similar positive bias. We emphasize that this apparent linguistic encoding of our social nature is a system level property, and in no way asserts all natural texts will skew positive (as exemplified by certain passages of the three works in Fig. 4), or diminishes the salience of negative states [28]. Nevertheless, a general positive bias points towards a positive social evolution, and may be linked to the gradual if haphazard trajectory of modern civilization toward greater human rights and decreases in violence [29]. Going forward, our word happiness assessments should be periodically repeated, and carried out

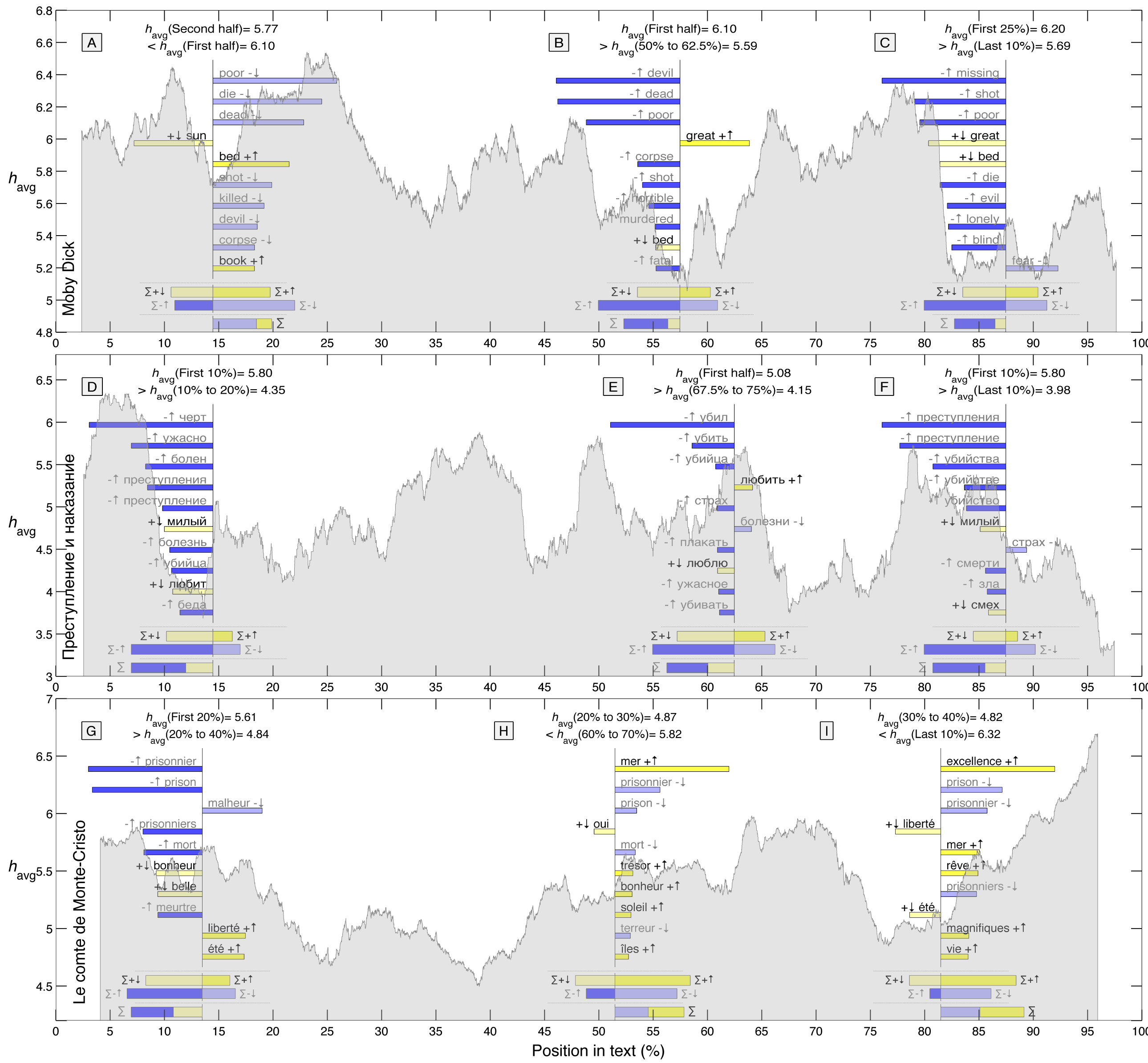

FIG. 4. Emotional time series for three great 19th century works of literature: Melville's Moby Dick, Dostoyevsky's Crime and Punishment, and Dumas' Count of Monte Cristo. Each point represents the language-specific happiness score for a window of 10,000 words (converted to lowercase), with the window translated throughout the work. The overlaid word shifts show example comparisons between different sections of each work. Word shifts indicate which words contribute the most toward and against the change in average happiness between two texts (see pp. S25–S27). While a robust instrument in general, we acknowledge the hedonometer's potential failure for individual words both due to language evolution and words possessing more than one meaning. While a robust instrument in general, we acknowledge the hedonometer's potential failure for individual words both due to language evolution and words possessing more than one meaning. For Moby Dick, we excluded 'cried' and 'cry' (to speak loudly rather than weep) and 'Coffin' (surname, still common on Nantucket). Such alterations, which can be done on a case by case basis, do not noticeably change the overall happiness curves while leaving the word shifts more informative. We provide links to online, interactive versions of these time series in the Supplementary Online Information.

for new languages, tested on different demographics, and expanded to phrases, both for the improvement of hedonometric instruments and to chart the dynamics of our collective social self.

[1] B. L. Whorf, *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, language (MIT Press, Cambridge, MA, 1956) edited by John B. Carroll.
[2] N. Chomsky, *Syntactic Structures*, language (Mouton, The Hague/Paris, 1957).
[3] S. Pinker, *The Language Instinct: How the Mind Creates Language*, language (William Morrow and Company, New York, NY, 1994).
[4] G. Lakoff and M. Johnson, *Metaphors We Live By*, language (University of Chicago Press, Chicago, IL, 1980).
[5] S. R. Campbell and J. W. Pennebaker, Psychological Science **14**, 60 (2003).
[6] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[7] J. W. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us* (Bloomsbury Press, New York, NY, 2011).
[8] M. K. Chen, American Economic Review 2013, 103(2): 690-731 **103**, 690 (2013).
[9] J. Boucher and C. E. Osgood, Journal of Verbal Learning and Verbal Behavior **8**, 1 (1969).
[10] M. M. Bradley and P. J. Lang, *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings*, Technical report C-1 (University of Florida, Gainesville, FL, 1999).
[11] P. J. Stone, D. C. Dunphy, and D. M. Smith, M. S.and Ogilvie, *The general inquirer: A computer approach to content analysis.* (MIT Press, Cambridge, Ma, 1966).
[12] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "Linguistic inquiry and word count: Liwc 2007," at `http://bit.ly/S1Dk2L`, accessed May 15, 2014. (2007).
[13] D. Jurafsky, V. Chahuneau, B. Routledge R., and N. A. Smith, First Monday **19** (2014).
[14] Google Labs ngram viewer. Available at `http://ngrams.googlelabs.com/`. Accessed May 15, 2014.
[15] (2006), Google Web 1T 5-gram Version 1, distributed by the Linguistic Data Consortium (LDC).
[16] P. S. Dodds and J. L. Payne, Phys. Rev. E **79**, 066115 (2009).
[17] I. M. Kloumann, C. M. Danforth, K. D. Harris, C. A. Bliss, and P. S. Dodds, PLoS ONE **7**, e29484 (2012).
[18] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth, PLoS ONE **6**, e26752 (2011).
[19] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth, PLoS ONE **8**, e64417 (2013).
[20] C. Osgood, G. Suci, and P. Tannenbaum, *The Measurement of Meaning* (University of Illinois, Urbana, IL, 1957).
[21] J. Redondo, I. Fraga, I. Padron, and M. Comesana, Behavior Research Methods **39**, 600 (August 2007).
[22] G. K. Zipf, *Human Behaviour and the Principle of Least-Effort*, patterns (Addison-Wesley, Cambridge, MA, 1949).
[23] K. Vonnegut, Jr., *A Man Without a Country*, stories (Seven Stories Press, New York, 2005).
[24] "Kurt Vonnegut on the shapes of stories," `https://www.youtube.com/watch?v=oP3c1h8v2ZQ`, accessed May 15, 2014.
[25] The Gutenberg Project: `http://www.gutenberg.org`; accessed November 15, 2013.
[26] F. Dostoyevsky, "Crime and punishment," Original Russian text. Obtained from `http://ilibrary.ru/text/69/p.1/index.html`, accessed December 15, 2013.
[27] P. S. Dodds and C. M. Danforth, Journal of Happiness Studies (2009), doi:10.1007/s10902-009-9150-9.
[28] J. P. Forgas, Current Directions in Psychological Science **22**, 225 (2013).
[29] S. Pinker, *The Better Angels of Our Nature: Why Violence Has Declined* (Viking Books, New York, 2011).
[30] Twitter API. Available at `http://dev.twitter.com/`. Accessed October 24, 2011.
[31] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, The Google Books Team, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. A. Lieberman, Science Magazine **331**, 176 (2011).
[32] E. Sandhaus, "The New York Times Annotated Corpus," Linguistic Data Consortium, Philadelphia (2008).

The authors acknowledge I. Ramiscal, C. Burke, P. Carrigan, M. Koehler, and Z. Henscheid, in part for their roles in developing hedonometer.org. The authors are also grateful for conversations with F. Henegan, A. Powers, and N. Atkins. PSD was supported by NSF CAREER Award # 0846668.