Tracking Fluctuations in Psychological States Using Social Media Language:

A Case Study of Weekly Emotion

Online Supplemental Materials

European Journal of Personality, *Eur. J. Pers.* (2020)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.2261

Tracking Fluctuations in Psychological States Using Social Media Language: A Case Study of Weekly Emotion

JOHANNES C. EICHSTAEDT[1]* and AARON C. WEIDMAN[2]*
[1]*Stanford University, USA*
[2]*University of Michigan, USA*

Abstract: Personality psychologists are increasingly documenting dynamic, within-person processes. Big data methodologies can augment this endeavour by allowing for the collection of naturalistic and personality-relevant digital traces from online environments. Whereas big data methods have primarily been used to catalogue static personality dimensions, here we present a case study in how they can be used to track dynamic fluctuations in psychological states. We apply a text-based, machine learning prediction model to Facebook status updates to compute weekly trajectories of emotional valence and arousal. We train this model on 2895 human-annotated Facebook statuses and apply the resulting model to 303 575 Facebook statuses posted by 640 US Facebook users who had previously self-reported their Big Five traits, yielding an average of 28 weekly estimates per user. We examine the correlations between model-predicted emotion and self-reported personality, providing a test of the robustness of these links when using weekly aggregated data, rather than momentary data as in prior work. We further present dynamic visualizations of weekly valence and arousal for every user, while making the final data set of 17 937 weeks openly available. We discuss the strengths and drawbacks of this method in the context of personality psychology's evolution into a dynamic science. © *2020 European Association of Personality Psychology*

Key words: big data; digital footprints; experience sampling; emotion; personality

Model Fitting: Examining the Link Between Valence and Arousal

We sought to model the relation between valence and arousal at the message level (in the calibration sample) and the week level (in the validation sample). Following Kuppens, Tuerlinckx, Barrett, and Russell (2013), to formally probe this relation, for each sample we fit a series of models to the data, each of which represented a distinct hypothesized relation between valence and arousal. To aid in parameter interpretation, prior to model fitting, we grand-mean centered valence and arousal annotations in the calibration sample, and in the validation sample we person-mean-centered valence and arousal ratings (which themselves were nested within-persons; see Enders & Tofighi, 2007). Given that weekly ratings of valence and arousal were nested within users in the validation sample, we used multilevel modeling to fit these models.

The seven models were as follows. Model 1 represented arousal as orthogonal to valence, Model 2 represented arousal as a symmetric positive linear function of valence, Model 3 represented arousal as a symmetric *v*-shaped function of valence (i.e., by predicting arousal from the absolute value of valence), and Models 4-6 represented arousal as an asymmetric *v*-shaped function of valence. Specifically, Model 4 included a parameter allowing for the positive and negative valence slopes to have different intercepts (i.e., a positivity/negativity offset), Model 5 included a parameter allowing for the positive and negative slopes to have different steepness (i.e., a positivity/negativity bias), and Model 6 included both an offset and a bias parameter. Finally, Model 7 represented arousal as an asymmetric, positive linear function of valence, in that it included parameters for both positivity/negativity offset and positivity/negativity bias.

We assessed model fit using approximate fit indices for each sample, given that these seven models are not hierarchically nested. In the calibration sample, we used the intuitive coefficient of determination (i.e., R^2). In the validation sample, we used the Akaike Information

Criterion (AIC) and the Bayesian Information Criterion (BIC) to index relative fit of each model; these indices penalize complex models with more freely estimated parameters, and lower values indicate better fit. Tables S1 and S2 show complete model details and fit statistics for the calibration and validation samples, respectively.

Results: Links Between Emotion and User Gender and Age

We found that women tended to show higher average levels of valence and arousal than men (*r*s = .22 and .23, respectively, *p*s < .001). Given that conveying emotion through Facebook status updates can be considered a form of expression, this finding is broadly consistent with previous work showing that women tend to be more emotionally expressive than men, particularly for positive emotions (e.g., Kring & Gordon, 1998; LaFrance, Hecht, & Levy Paluck, 2003). Women also showed greater variability in valence than men (β = .19, p < .001) but not in arousal (β = -.01, p = .65); the link with valence may be explained by prior work showing that women may, in some circumstances, show more emotional reactivity than men (Gard & Kring, 2007; Grossman & Wood, 1993). Age showed a relatively weak association with average levels of valence and arousal (*r*s = .06 and .02, respectively, *p*s > .14); although prior work has shown that people tend to experience less negative affect as they age (e.g., Charles, Reynolds, & Gatz, 2001), this effect may not have emerged as strongly here given that our sample was primarily comprised of young adults. Younger adults also showed greater variability in valence than older adults (β = -.12, p < .01) but arousal variability showed no link with age (β = -.04, p = .17); the link with valence is in line with some prior evidence showing that affect variability decreases with age (Röcke, Li, & Smith, 2009).

Table S1: Model fit statistics for relation between valence and arousal in calibration sample

Model	Proposed Relation	R^2
1	Independence	.000
2	Linear Relation	.049
3	V-shaped: Symmetric	.163
4	V-shaped: Positivity offset	.172
5	V-shaped: Negativity bias	.180
6	V-shaped: Positivity offset and negativity bias	.181
7	Linear Relation: Positivity offset and negativity bias	.181

N = 2,985 emotion observations

Table S2: Model fit statistics for relation between valence and arousal in validation sample

Model	Proposed Relation	AIC	BIC
1	Independence	27520	27543
2	Linear Relation	25571	25618
3	V-shaped: Symmetric	27356	27402
4	V-shaped: Positivity offset	26542	26622
5	V-shaped: Negativity bias	26114	26192
6	V-shaped: Positivity offset and negativity bias	26096	26212
7	Linear Relation: Positivity offset and negativity bias	25426	25542

N = 303,576 emotion observations

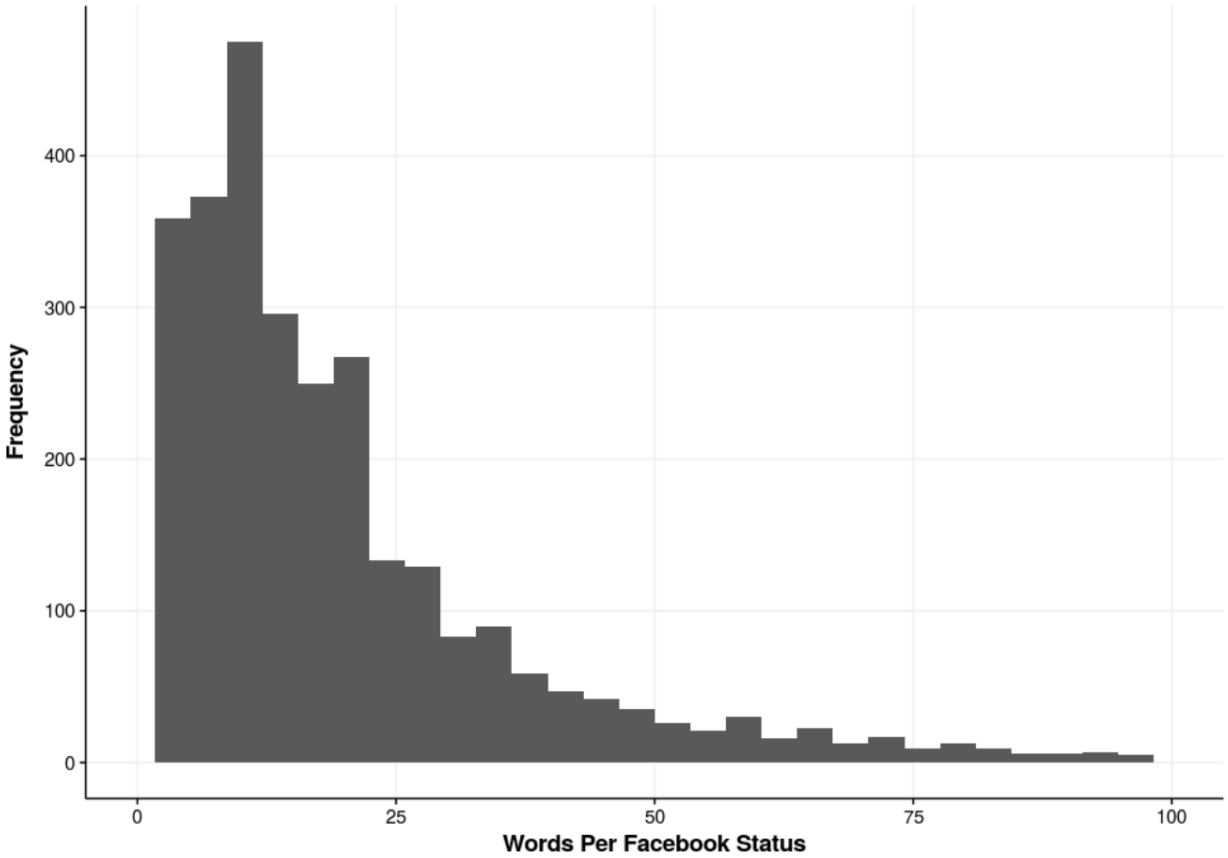

Figure S1: Histogram of words per Facebook status in the calibration sample.

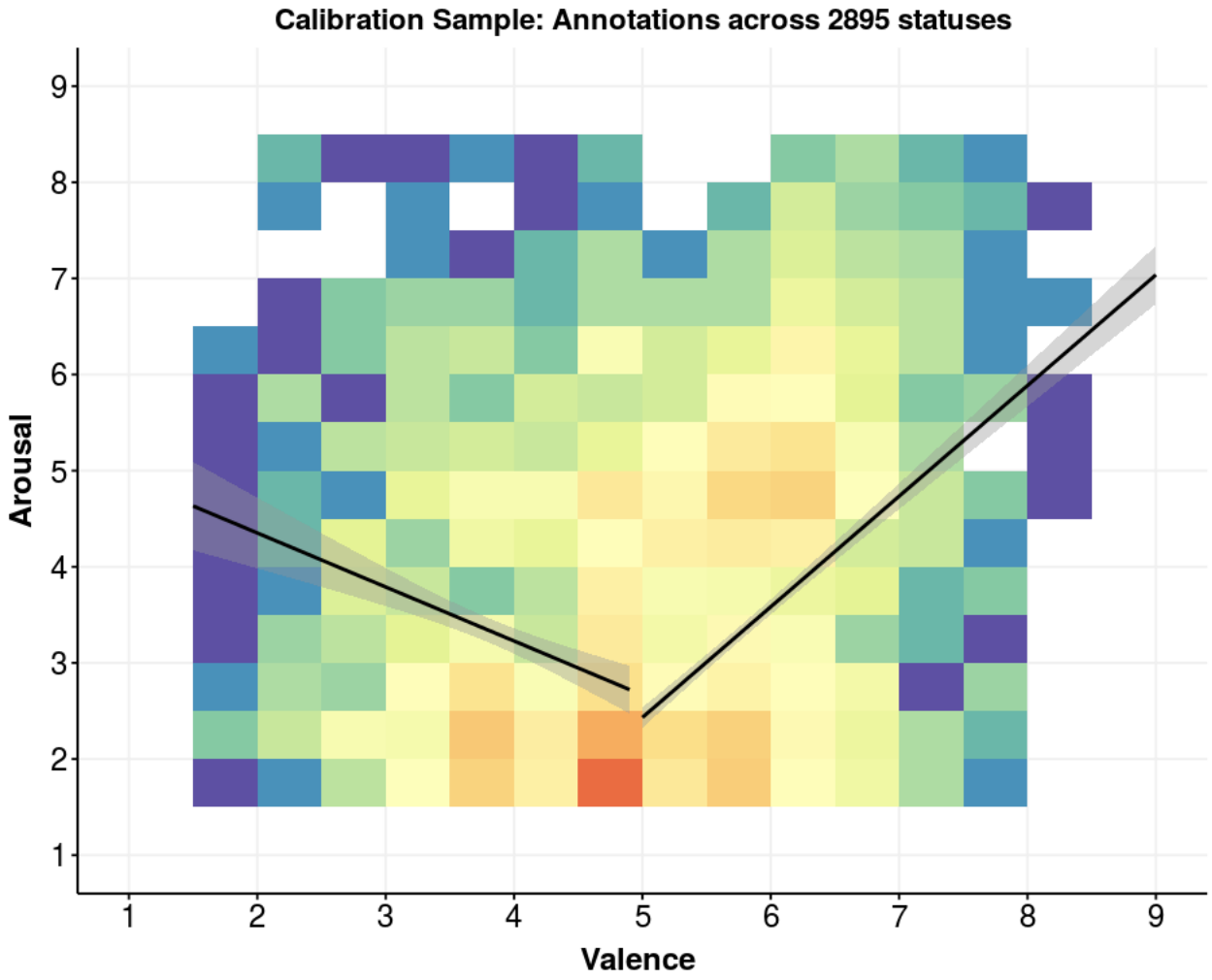

Figure S2: Valence as a function of arousal in the calibration sample. Each point represents an annotated estimate of valence and arousal for one Facebook status (*N* = 2,985 statuses).

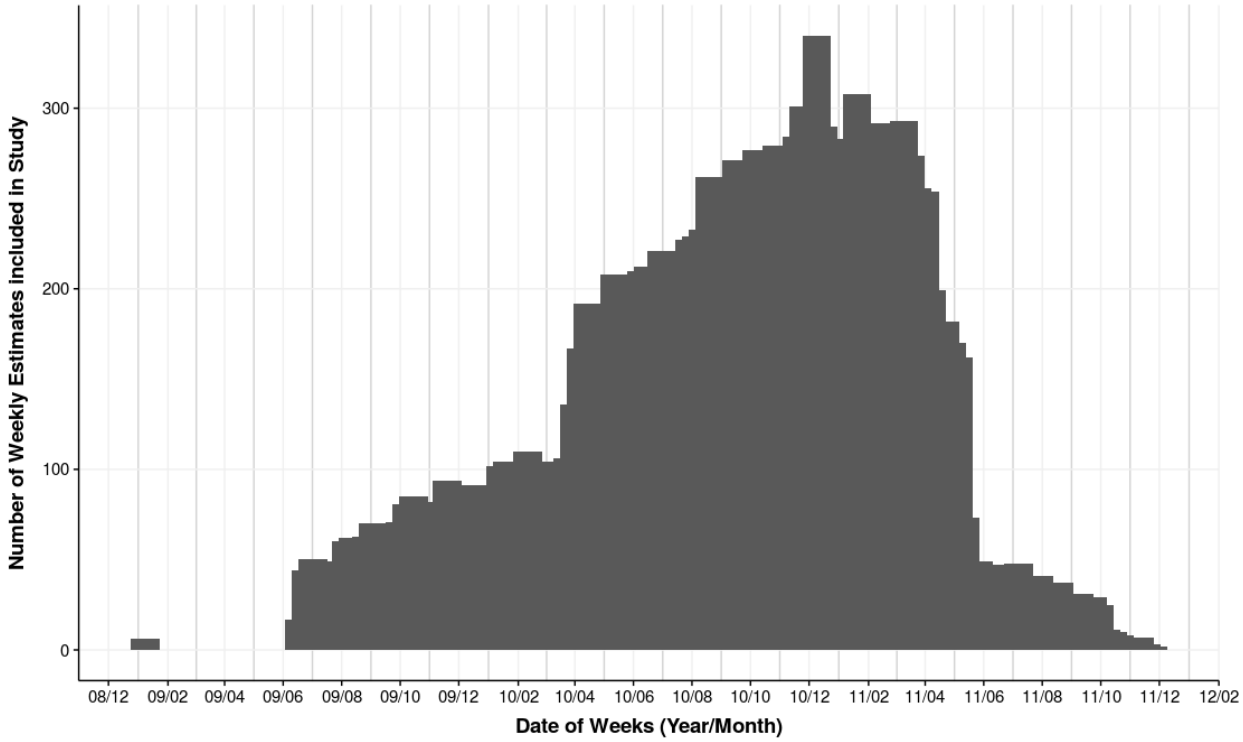

Figure S3: Temporal distribution of weeks included in the final validation sample.

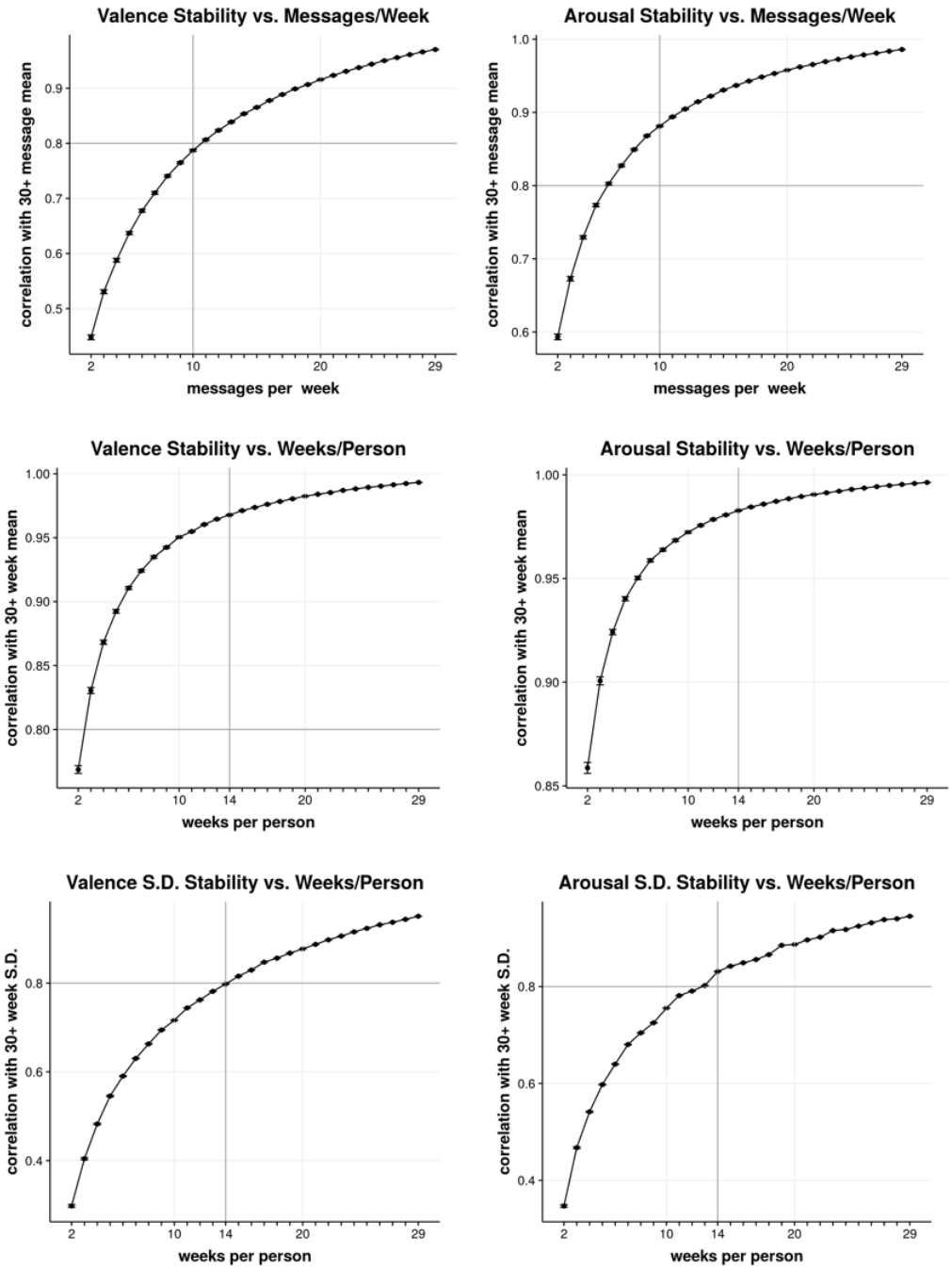

Figure S4: Stability of valence and arousal mean and standard deviations as a function of different thresholds for (a) 2 to 29 messages per week (top row), and (b) 2 to 29 weeks per person (bottom two rows). Vertical axis values represent correlations for sub-samples of different sizes (2 to 29) (a) between weekly emotion estimates derived in with a "ground truth" estimate using > 30 messages per week, and (b) between weekly emotion estimates and "ground truth" estimates using > 30 weeks per person (with at least 10 messages per week). In each panel, we sought to determine the subsample size for which this correlation exceeded .80 (i.e., a threshold for adequate reliability, shown as a horizontal line).

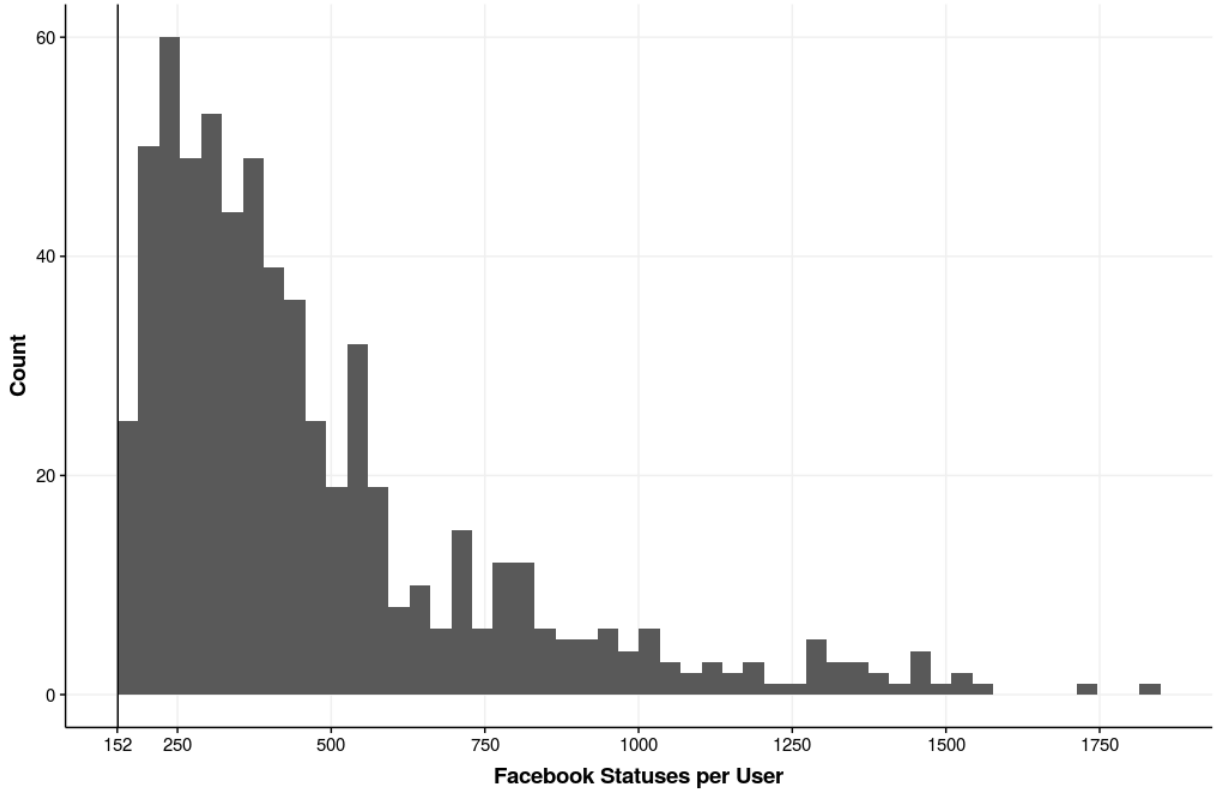

Figure S5: Histogram of Facebook statuses per user (across 640 users) included in the validation sample. Every user has at least 14 weeks with at least 10 statuses each in them for a minimum of at least 140 statuses (actual empirical minimum: 152).

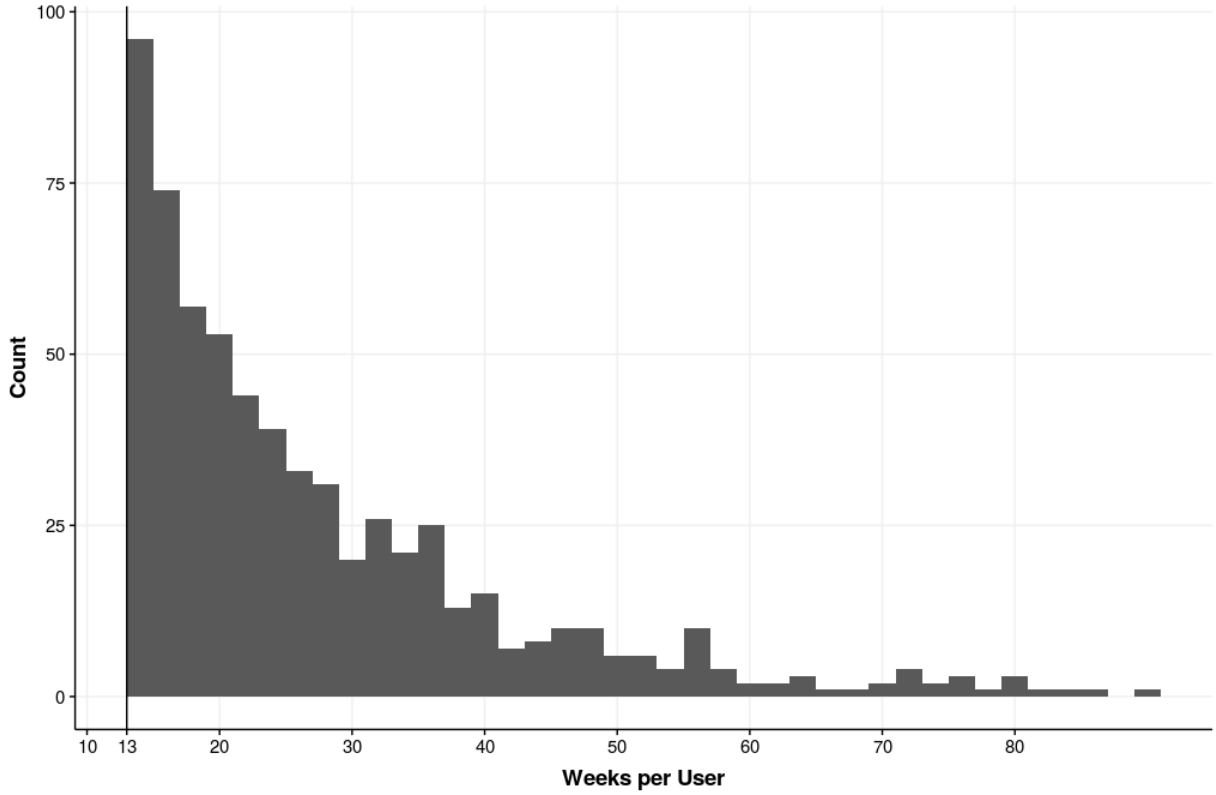

Figure S6: Histogram of weeks per user (across 640 users) included in the validation sample, for a total of 17,936 weeks. Every user has at least 14 weeks of data.

Table S3: Relations between demographics, personality and weekly valence and arousal with no age or gender controls

	Mean levels				Variabilty (controlled for mean levels)			
	Valence		Arousal		Valence		Arousal	
	r [95% CI]	p	r	p	β [95% CI]	p	β [95% CI]	p
Demographics								
Age	.07 [-.01, .15]	.073	.04 [-.04, .11]	.612	**-.14 [-.22, -.07]**	.000	-.05 [-.10, .01]	.115
Gender	**.22 [.14, .29]**	.000	**.23 [.16, .30]**	.000	**.19 [.12, .27]**	.000	-.01 [-.07, .04]	.648
Personality								
Agreeableness	**.21 [.13, .28]**	.000	**.12 [.05, .20]**	.005	.01 [-.06, .09]	.750	.03 [-.03, .08]	.363
Extraversion	**.32 [.25, .39]**	.000	**.18 [.11, .26]**	.006	.02 [-.06, .10]	.634	-.02 [-.08, .04]	.533
Conscientiousness	**.19 [.12, .27]**	.000	.11 [.03, .19]	.242	-.04 [-.12, .03]	.262	-.03 [-.09, .03]	.313
Openness	-.05 [-.12, .03]	.231	**-.11 [-.18, -.03]**	.040	-.05 [-.13, .02]	.164	-.01 [-.07, .05]	.786
Neuroticism	**-.16 [-.23, -.08**	.000	-.04 [-.12, .04]	.549	**.08 [.01, .16]**	.029	.00 [-.05, .06]	.863

Note: Valence and arousal standardized coefficients for users' mean-levels across weeks or standard deviation across weeks (variability) regressed against personality and demographics. Coefficients for variability are adjusted for mean levels.

Table S4: Descriptive statistics for primary variables in validation data set

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Valence mean	5.10	0.20													
2. Arousal mean	3.09	0.59	.53** [.47, .58]												
3. Valence SD	0.24	0.07	.26** [.19, .33]	.26** [.19, .33]											
4. Arousal SD	0.45	0.19	.33** [.26, .40]	.68** [.64, .72]	.28** [.21, .35]										
5. Age	22.48	7.08	.07 [-.01, .15]	.04 [-.04, .11]	-.12** [-.20, -.05]	-.02 [-.10, .06]									
6. Gender	0.50	0.50	.22** [.14, .29]	.23** [.16, .30]	.24** [.17, .31]	.14** [.07, .22]	.01 [-.07, .09]								
7. Agreeableness	3.52	0.60	.21** [.13, .28]	.12** [.05, .20]	.07 [-.01, .14]	.11** [.03, .19]	.05 [-.03, .12]	-.05 [-.13, .03]							
8. Conscientiousness	3.30	0.63	.19** [.12, .27]	.11** [.03, .19]	.01 [-.07, .09]	.05 [-.03, .12]	.10** [.03, .18]	-.02 [-.10, .05]	.19** [.12, .27]						
9. Extraversion	3.45	0.76	.32** [.25, .39]	.18** [.11, .26]	.10* [.02, .18]	.11** [.03, .18]	.06 [-.02, .13]	-.00 [-.08, .08]	.20** [.13, .28]	.22** [.14, .29]					
10. Neuroticism	2.91	0.82	-.16** [-.23, -.08]	-.04 [-.12, .04]	.04 [-.04, .12]	-.02 [-.10, .05]	-.05 [-.12, .03]	.33** [.26, .40]	-.41** [-.47, -.34]	-.36** [-.42, -.29]	-.41** [-.47, -.34]				
11. Openness	4.01	0.49	-.05 [-.12, .03]	-.11** [-.18, -.03]	-.07 [-.14, .01]	-.08* [-.16, -.00]	.05 [-.03, .13]	-.08* [-.16, -.00]	.05 [-.03, .13]	.03 [-.05, .10]	.18** [.11, .26]	-.03 [-.10, .05]			
12. Valence autocorrelation (lag 1)	0.05	0.26	.19** [.12, .27]	.13** [.05, .20]	.22** [.14, .29]	.11** [.03, .19]	-.02 [-.10, .06]	.09* [.01, .16]	.05 [-.03, .12]	.11** [.03, .18]	.02 [-.05, .10]	-.03 [-.11, .05]	.01 [-.07, .09]		
13. Arousal autocorrelation (lag 1)	0.13	0.27	.15** [.07, .22]	.21** [.13, .28]	.08* [.00, .15]	.25** [.17, .32]	.08* [.00, .16]	.09* [.01, .16]	.02 [-.06, .09]	.08* [.00, .15]	.05 [-.03, .12]	-.02 [-.10, .06]	-.02 [-.10, .06]	.20** [.13, .27]	
14. Positivity Bias	3.62	98.08	.06 [-.02, .14]	.01 [-.07, .09]	-.00 [-.08, .08]	.02 [-.06, .10]	.00 [-.08, .08]	-.03 [-.11, .05]	.04 [-.04, .12]	-.02 [-.10, .06]	.02 [-.07, .10]	-.04 [-.12, .04]	-.01 [-.09, .07]	-.04 [-.12, .04]	-.01 [-.10, .07]

Note. M and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. * indicates $p < .05$. ** indicates $p < .01$.

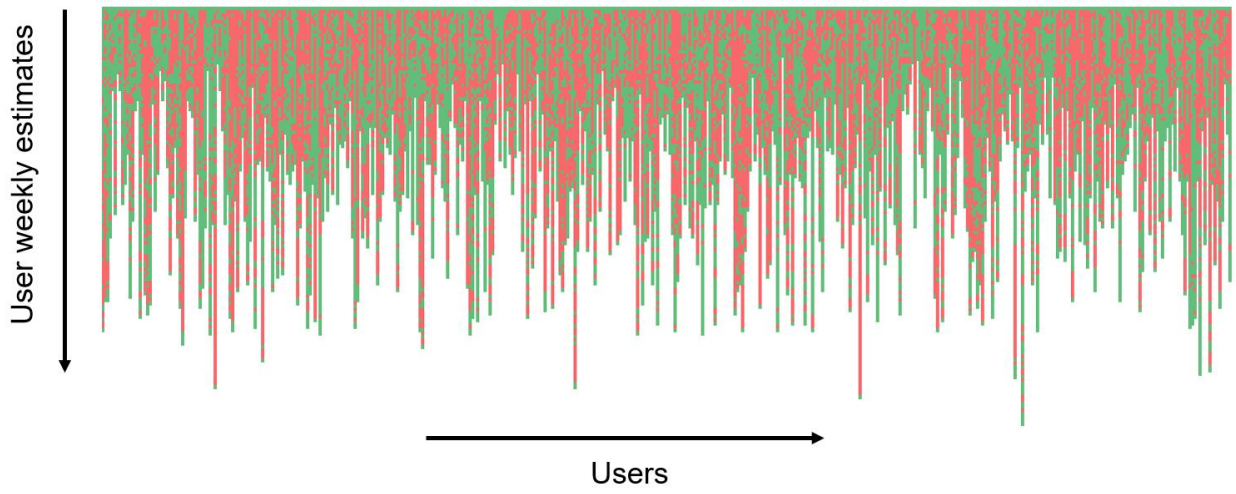

Figure S7. User time series in validation data set. Green dots indicate weeks for which estimates are available, red dots signify insufficient data. All weeks shown relative to user, with the first row being the first week for which a weekly estimate is available for a given user, and all subsequent weeks in that column are weeks subsequent to that week for that user. Total range of weeks (from first week with estimate to last week with estimate) is 15 to 125 per user (mean = 57.62 , SD = 24.09).

Table S5. Average autocorrelations for lags of 1 to 7 weeks for valence and arousal in validation data set

Valence	Average	SD	SE
Lag 1	0.05	0.26	0.01
Lag 2	0.03	0.23	0.01
Lag 3	0.02	0.22	0.01
Lag 4	-0.01	0.20	0.01
Lag 5	-0.01	0.18	0.01
Lag 6	-0.01	0.17	0.01
Lag 7	-0.01	0.16	0.01

Arousal	Average	SD	SE
Lag 1	0.13	0.27	0.01
Lag 2	0.07	0.24	0.01
Lag 3	0.05	0.22	0.01
Lag 4	0.03	0.20	0.01
Lag 5	0.02	0.20	0.01
Lag 6	0.01	0.18	0.01
Lag 7	0.01	0.17	0.01

Table S6. Associations of lag 1 autocorrelation coefficients across users in validation data set

	Autocorrelation, lag 1				Autocorrelation, lag 1 (ctrld. for age, gender)			
	Valence Lag 1		Arousal Lag 1		Valence Lag 1		Arousal Lag 1	
	r [95% CI]	*p*	r [95% CI]	*p*	β [95% CI]	*p*	β [95% CI]	*p*
Demographics								
Age	-0.02 [-0.10, 0.06]	0.58	**0.08 [0.00, 0.16]**	0.04	-0.02 [-0.03, -0.02]	0.57	**0.08 [0.08, 0.08]**	0.04
Gender	**0.09 [0.01, 0.16]**	0.03	**0.09 [0.01, 0.16]**	0.03	**0.09 [0.05, 0.13]**	0.03	**0.09 [0.04, 0.13]**	0.03
Personality								
Agreeableness	0.05 [-0.03, 0.12]	0.24	0.02 [-0.06, 0.09]	0.68	0.05 [0.02, 0.09]	0.18	0.02 [-0.02, 0.05]	0.66
Extraversion	0.02 [-0.05, 0.10]	0.54	0.05 [-0.03, 0.12]	0.23	0.03 [-0.00, 0.05]	0.51	0.04 [0.02, 0.07]	0.27
Conscientiousness	**0.11 [0.03, 0.18]**	0.01	**0.08 [0.00, 0.15]**	0.05	**0.11 [0.08, 0.14]**	0.00	0.07 [0.04, 0.11]	0.07
Openness	-0.03 [-0.11, 0.05]	0.44	-0.02 [-0.10, 0.06]	0.65	-0.07 [-0.09, -0.04]	0.10	-0.05 [-0.08, -0.02]	0.25
Neuroticism	0.01 [-0.07, 0.09]	0.83	-0.02 [-0.10, 0.06]	0.62	0.02 [-0.02, 0.06]	0.67	-0.02 [-0.06, 0.03]	0.67
Usage statistics								
Total words	**0.13 [0.06, 0.21]**	0.00	**0.16 [0.09, 0.24]**	0.00	**0.13 [0.13, 0.13]**	0.00	**0.15 [0.15, 0.15]**	0.00
Total messages	**0.19 [0.11, 0.26]**	0.00	**0.23 [0.15, 0.30]**	0.00	**0.18 [0.18, 0.18]**	0.00	**0.21 [0.21, 0.21]**	0.00
Weeks per user	**0.17 [0.09, 0.24]**	0.00	**0.19 [0.12, 0.27]**	0.00	**0.16 [0.16, 0.16]**	0.00	**0.18 [0.18, 0.19]**	0.00

Supplemental References

Charles, S. T., Reynolds, C. A., & Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology*, *80*, 136-151.

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multi-level models: A new look at an old issue. *Psychological Methods*, *12*, 121-138.

Gard, M. G., & Kring, A. M. (2007). Sex differences in the time course of emotion. Emotion, 7, 429-43.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology*, *65*, 1010-1022.

Kring, A. M. & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, *74*, 686-703.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*, 917-940.

LaFrance, M., Hecht, M. A., & Paluck, E. L. (2003). The contingent smile: a meta-analysis of sex differences in smiling. *Psychological Bulletin*, *129*, 305-334.

Röcke, C., Li, S. C., & Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults? *Psychology and Aging*, 24, 863.