

Birds of a feather do flock together: behavior and language-based personality  
assessment reveal personality homophily among couples and friends

Youyou Wu

Department of Psychology, University of Cambridge

David Stillwell

Judge Business School, University of Cambridge

H. Andrew Schwartz

Department of Computer Science, Stony Brook University

Michal Kosinski

Graduate School of Business, Stanford University

## Abstract

Friends and spouses tend to be similar in a broad range of characteristics (i.e. *homophily*), such as age, educational level, attitudes, values, and general intelligence. Surprisingly, little evidence has been found for similarity in personality—one of the most fundamental psychological constructs. We argue that the lack of evidence for personality homophily derives from the tendency of individuals to make personality judgments in relation to a salient comparison group rather than in absolute terms when responding to the self-report and peer-report questionnaires commonly used in personality research (i.e. *reference-group effect*). We address this limitation by employing personality measures based on behavior and language that are resistant to this bias. Results based on a large sample of  $n = 231,707$  participants provide evidence for a strong personality homophily between romantic partners ( $r_s = .20$  to  $.47$ ) and between friends ( $r_s = .11$  to  $.32$ ). We discuss the practical and methodological implications of these findings.

*Keywords:* homophily, personality, personality assessment, reference-group effect, social network, close relationships

**Birds of a feather do flock together: behavior and language-based personality assessment reveal personality homophily among couples and friends**

It is well established that people prefer to interact with similar others (McPherson, Smith-Lovin, & Cook, 2001). This tendency, referred to as *homophily*, was found for traits such as age, education, attitudes, values, and general intelligence (Rushton & Bons, 2005). Surprisingly, however, little evidence was present for homophily in personality—a fundamental psychological construct that underpins much of the variation in human behaviors and preferences. Past studies showed no or only weak similarity between romantic partners and between friends (Altmann, Sierau, & Roth, 2013; Botwin, Buss, & Shackelford, 1997; Buss, 1984a; Funder, Kolar, & Blackman, 1995; Rushton & Bons, 2005; Watson et al., 2004; Watson, Beer, & McDade-Montez, 2014), with occasional findings indicating moderate homophily in Openness and Conscientiousness between romantic partners (McCrae & Martin, 2008; Rammstedt & Schupp, 2008; Watson, Hubbard, & Wiese, 2000). This has led researchers to endorse the conclusion of early theorists that “mating is essentially random for personality differences” (Eysenck, 1990).

We argue that the weak evidence for personality homophily stems from the reliance on self-report (e.g. Kolar, Funder, & Colvin, 1996; David Watson et al., 2014) and peer-report (e.g. Altmann et al., 2013; Buss, 1984a) in a majority of previous studies. Such assessment approaches are inappropriate when investigating similarity between friends and between romantic partners, as they are affected by a tendency of the respondents to judge themselves in relation to a salient comparison group, rather than in absolute terms (*the reference-group effect*; Heine, Lehman, Peng, & Greenholtz, 2002). For instance, an objectively introverted engineer might perceive herself as an extrovert, if she was surrounded by a peer-group of even more introverted engineers. Similar bias affects peer-report as well; the introverted

friends of the said engineer might also see her as extroverted, as the most salient reference group is their own. In fact, some personality questionnaires specifically instruct people to describe themselves “in relation to other people you know of” (e.g. Goldberg et al., 2006), and several studies have shown that the reference-group effect pertains to questionnaire-based personality judgments (Credé, Bashshur, & Niehorster, 2010; Wood, Brown, Maltby, & Watkinson, 2012). Consequently, self and peer-report of personality amplify the differences in actual personality and obscure the similarities between peers, rendering them inappropriate to study homophily.

Indeed, rare evidence of homophily emerges from a few studies relying on alternative ways of measuring personality that are less susceptible to the reference-group effect. For example, Botwin et al. (1997) and Buss (1984a) provided evidence for homophily between spouses by using independent interviewers’ ratings of personality.<sup>1</sup> Also, Buss (1984b) showed relatively strong homophily between romantic partners by comparing the self and peer-reported frequencies of certain personality-related behaviors (the *act frequency approach*; Buss & Craik, 1983). Introversion, for example, was measured by asking participants to judge whether in last three months they “watched the soap opera on TV” or “went for a long walk alone”. This approach focuses on concrete behaviors and thus leaves less room for subjective comparisons.

This paper thus aims to study personality homophily using behavior-based and language-based personality measures that do not involve self-report questionnaire, bypassing the reference-group effect. The behavior-based personality assessment approach used here relies on a common digital behavioral residue: *Facebook Likes*.<sup>2</sup> Likes were previously

---

<sup>1</sup> Such judgments are still susceptible to a reference-group effect—the interviewer’s reference group—but it pertains equally to both dyad members, and thus does not obscure the similarity between them.

<sup>2</sup> Facebook users generate Likes by clicking a Like button accompanying others’ posts and pictures, or Facebook pages related to products, famous people, books, etc. Likes are also generated outside of the Facebook

shown to enable accurate assessment of individuals' personality (Kosinski, Stillwell, & Graepel, 2013; Youyou, Kosinski, & Stillwell, 2015) and to be linked with personality in a theory-consistent fashion (Kosinski et al., 2013). For example, participants with high Openness to Experience tend to like content related to art (e.g. "Salvador Dalí"), spirituality (e.g. "meditation"), or science (e.g. "TED talks"); while Extroverts express their preference for "partying", "dancing", and celebrities (e.g. "Snookie").<sup>3</sup>

The language-based personality assessment used in this work is based on participants' language used in Facebook status updates<sup>4</sup> (Park et al., 2014; Schwartz et al., 2013). Previous research has well documented the links between personality and language use (e.g. Cohen, Minor, Baillie, & Dahir, 2008; Fast & Funder, 2008; Hirsh & Peterson, 2009; Mehl, Gosling, & Pennebaker, 2006; Pennebaker, Mehl, & Niederhoffer, 2003; Tausczik & Pennebaker, 2010). The things we talk about in everyday life reflect our behavior, feelings, and thoughts (Tausczik & Pennebaker, 2010). Neurotic individuals, for example, tend to use first-person singulars (e.g. "I", "me", or "mine"), while extroverts use more words describing positive emotions (e.g. "great", "happy", or "amazing"; Schwartz et al., 2013). Several studies illustrated the feasibility of assessing personality based on language use in social media in general (Golbeck, Robles, Edmondson, & Turner, 2011; Sumner, Byers, Boochever, & Park, 2012; Sumner, Byers, & Shearing, 2011) and Facebook status updates in particular (Park et al., 2014; Schwartz et al., 2013).

Both behavior-based and language-based assessment approaches are applied in three steps. First, we obtain a sample of participants that contains their Facebook profiles and scores on a traditional self-report questionnaire (Goldberg et al., 2006) measuring the widely-

---

environment; Like buttons can be found across many websites and even in the offline world (where they can be "clicked" by photographing a QR code with a smartphone).

<sup>3</sup> See Kosinski et al. (2013) for more examples of Likes strongly associated with each personality trait.

<sup>4</sup> Facebook users can use status updates to share their thoughts, feelings, and life events with their Facebook friends.

accepted Five Factor Model (FFM; Costa & McCrae, 1992). Second, we use this sample to train cross-validated predictive models linking self-report of personality scores with behaviors and language use. Third, we apply those models to a *separate sample* of participants to estimate their personality scores based on their Likes and status updates. The resulting personality scores are compared between romantic partners and between friends to investigate the degree of personality homophily.

Importantly, although both language and behavior-based models were trained on participants' self-report personality scores, they do not inherit the reference-group effect. Those models aim at minimizing squared errors across the entire sample, therefore averaging out the effect of participants overestimating or underestimating their true scores in the context of their reference points.

## Methods

### *Participants*

This study relies on three, partially overlapping, samples obtained from the myPersonality project (<http://mypersonality.org>). MyPersonality was a popular Facebook application that offered users psychometric tests and feedback on their scores. All participants provided an opt-in consent to record their personality scores and the contents of their Facebook profile. The average age was 24.7. Females constituted 57.8% of the sample.

The first sample was used to develop behavior-based personality assessment approach. It contained 236,251 US participants who completed personality questionnaire and had at least 20 Likes on their Facebook profile.

The second sample was used to develop language-based personality assessment. It included 59,547 US participants who completed a personality questionnaire and wrote at least 500 words across all status updates.

The third sample was used to investigate the existence of personality homophily. It included 231,707 US participants who had at least one friend or romantic partner in the sample. The friendship dyads between those participants were identified using Facebook friend list; romantic couples were identified using the “relationship status” field. Their personality scores were estimated using self-report questionnaire, Facebook Likes, or Facebook status updates. Questionnaire scores were available for 123,273 friendship dyads and 4,348 heterosexual romantic couples. Behavior-based personality scores were available for 49,920 friendship dyads and 983 heterosexual romantic couples. The language-based scores were available for 6,274 friendship dyads and 283 heterosexual romantic couples. For individuals belonging to multiple friendship dyads, we randomly include only one dyad in the sample to ensure that all friendship dyads are independent from each other.

#### *Questionnaire-based personality measure*

Participants completed the 20 to 100-item-long International Personality Item Pool (IPIP) FFM questionnaire (Goldberg et al., 2006). The Cronbach’s alpha reliability coefficients for the 100-item-long questionnaire completed by 29% of participants are .84, .92, .93, .88, and .93 for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (respectively). Corresponding values for 20-item-long version are .48, .67, .73, .58, and .65.

#### *Behavior-based personality measure*

The behavior-based personality assessment model was built using a *separate* training sample and following the procedure described in detail in Youyou, Kosinski, and Stillwell (2015). First, we used records of participants’ Likes to build a participant-Like matrix. As some of the participants had very few Likes, and some Likes were associated with very few participants, the matrix was trimmed so that each participant was associated with at least 20 Likes, and each Like was associated with at least 20 participants. The resulting matrix

included 236,251 participants (rows) and 128,787 unique Likes (columns). Second, we used participants' personality scores to build five models predicting each trait from Likes. We used a 10-fold, cross-validated linear regression model with a combination of L1 (LASSO; Least Absolute Shrinkage and Selection Operator, or LASSO; Tibshirani, 1996) and L2 (ridge) penalty. The resulting five linear regression models were applied to participants' Likes to generate behavior-based personality scores in this study.

The predictive accuracy of these prediction models, indicated by the correlations between Likes-predicted and participants' self-reported personality scores, was  $r_s = .51$  for Openness,  $.40$  for Conscientiousness,  $.43$  for Extraversion,  $.39$  for Agreeableness, and  $.39$  for Neuroticism. As shown in Figure S1 in the online Supplemental Material, the assessment accuracy increased with number of Likes available on given participants' profiles. Note that some Likes were diagnostic of multiple personality traits; for example, liking "Kindle" was associated with high Openness, high Agreeableness, and low Neuroticism.

To calculate the reliability of the models, we randomly split the Likes into two halves and developed two separate models following the procedure described above. The split-half reliability, indicated by the correlations between the scores predicted using two separate models, was  $r_s = .81$  for Openness,  $.78$  for Conscientiousness,  $.76$  for Extraversion,  $.75$  for Agreeableness, and  $.77$  for Neuroticism. As shown in Figure S2, the split-half reliability also increased with number of Likes available.

The resulting models were applied to dyads where both members provided us with a list of their Likes and had at least 20 Likes after the Likes shared between dyad members are removed. Shared Likes were removed before assessing dyad members' personality to ensure that the overlap in Likes does not artificially inflate personality homophily. The overlap in Likes between dyad members was relatively low in our sample: friends shared 5.8 Likes on average or 1.4% of their joint Likes, while romantic partners shared 14.5 Likes or 3.3% of



their joint Likes. Participants in the current sample had 150 Likes on average, which corresponds to an expected accuracy of around .49.

#### *Language-based personality measure*

The language-based personality assessment model was developed using an open-vocabulary approach similar to one employed by Park et al. (2014). First, we obtained a training sample of 59,547 myPersonality participants who completed a 20 to 100-item-long IPIP FFM questionnaire and provided us with a list of their status updates with at least 500 words across all status updates. The sample included 8.7 million status updates: 147 updates per participant on average. Second, we extracted words and phrases (two and three-word sequences such as *happy birthday*, or *I love you*) from each participant's status updates, and kept those mentioned by at least 10% of the participants. We created binary representations (0 or 1) of every language feature, indicating whether the word or phrase was ever used by each participant, as well as continuous representations, encoding the percentage of a participant's total words that were the given word or phrase.

Third, before proceeding to training predictive models, we controlled for overlap in language between dyad members by assessing their personality using separate sets of models. Specifically, we created two datasets by randomly splitting the available words and phrases into two halves. This enabled us to train *two sets* of predictive models relying on different features and ensure that the use of the exact same words or phrases (e.g. due to repeating phrases used by one's friends) did not artificially inflate personality homophily; dyad members' personality could be estimated using separate models based on different words and phrases.

Finally, we built two sets of five models predicting each of the personality traits from language features. We used a 10-fold, cross-validated linear regression model with regularization (Tikhonov, 1963) to avoid overfit from an L2 penalty (ridge), univariate

feature selection, and randomized principal components analysis exactly as done by Park et al., (2014).

The predictive accuracy of those models, indicated by the correlations between language and questionnaire-based personality scores, was  $r_s = .37$  for Openness,  $.33$  for Conscientiousness,  $.35$  for Extraversion,  $.27$  for Agreeableness, and  $.29$  for Neuroticism. As shown in Figure S1, accuracy of the prediction grows with the number of words a given participant writes. The split-half reliability, indicated by the correlations between the scores predicted using two separate models, was  $r_s = .71$  for Openness,  $.75$  for Conscientiousness,  $.80$  for Extraversion,  $.72$  for Agreeableness, and  $.76$  for Neuroticism. As shown in Figure S2, the split-half reliability also increased with number of words a given participant writes. Participants in the current sample wrote 3,141 words on average, which corresponds to an expected accuracy of around  $.33$ . The resulting two sets of models applied to dyads where both members had at least 500 words across all status updates to estimate language-based personality scores.

### *Estimating homophily*

Personality homophily was estimated by correlating dyad members' scores on a given personality trait across all dyads. For romantic couples, personality scores were aligned by gender and Pearson product-moment correlations between male and female scores were calculated. For friendship dyads, the assignment of dyad members as Person A or Person B is arbitrary and thus intraclass correlation was used (see Watson et al., 2000).<sup>5</sup>

## **Results**

The goal of this study was to verify whether individuals tend to interact with others characterized by similar personality traits. The estimates of personality homophily—based on

---

<sup>5</sup> We also tested an alternative approach: randomly assigning friendship dyad members as person A or person B, and conducting regular Pearson product-moment correlations. The results were very similar to those obtained in the intraclass correlation (within  $\Delta r = +/- .02$  of the original values).

three different personality measures used in this study—are presented in Figure 1. Consistent with previous research, questionnaire-based scores show weak to negligible personality homophily for both romantic couples  $\bar{r}(4,346) = .09$ , and friends  $\bar{r}(246,544) = .06$ .<sup>6</sup> In contrast, behavior-based scores show significant positive correlations across all five personality traits:  $\bar{r}(981) = .24$  on average for romantic couples, and  $\bar{r}(99,838) = .12$  on average for friends. Even stronger homophily can be seen for the language-based scores:  $\bar{r}(281) = .36$  for romantic couples and  $\bar{r}(12,546) = .24$  for friends. Neither of the scores were significantly correlated ( $|rs| < .01$ ) when individuals were paired randomly.

Personality scores (and resulting homophily) were estimated *after* controlling for shared Likes or shared language features, as described in the Methods section. Although the overlap in Likes or language features between dyad members might be driven by factors such as a shared environment, culture, or interpersonal influence, they can also stem from actual personality similarity. Thus, controlling the effect by removing shared Likes or applying disjoint language models results in a *lower-bound* estimate of homophily—some of the effect is lost. To estimate the *upper-bound* of the homophily, we re-estimated our results *without* controlling for shared Likes or language features. As expected, the results presented in Figure S3 show even stronger homophily. The behavior-based similarity correlations were  $\bar{r}(1,063) = .33$  for romantic couples and  $\bar{r}(102,276) = .17$  for friends; the language-based ones were  $\bar{r}(281) = .41$  for romantic couples and  $\bar{r}(12,546) = .26$  for friends.

Additionally, we investigated whether personality homophily might be a by-product of similarity in other traits (Buss, 1984a). To this end, we conduct partial correlations of personality scores, controlling for dyad members' age and education.<sup>7</sup>

---

<sup>6</sup> The average correlations of the five traits were calculated using Fisher's *r*-to-*z* transformation. Due to large sample sizes, all correlations reported in the paper are significant at  $p < .01$ .

<sup>7</sup> We used key words such as “university” or “college” (excluding “community college”) in the school names to identify a sample of users with higher education degrees.

Controlling for education level did not considerably change the level of homophily between friends: the results were within  $\Delta r_s(6,292) = \pm .02$  and  $\Delta r_s(2,012) = \pm .04$  of the original values for behavior-based and language-based scores respectively. Similarly, little change was observed when controlling for age. Results controlled for age were within  $\Delta r_s = \pm .03$  of the original values for both romantic couples and friends, with an exception of Conscientiousness, where homophily decreased by  $\Delta r(753) = -.09$  (behavior-based scores) and  $\Delta r(250) = -.11$  (language-based scores) for romantic couples.

In other words, homophily in personality seems to be independent from the homophily in age and education. The only exception was romantic partners' similarity in Conscientiousness, which was partially due to their similarity in age. However, the similarity correlations remained significant at  $p < .001$  even after removing the effect of age:  $r(753) = .17$  (behavior-based scores) and  $r(250) = .36$  (language-based scores).

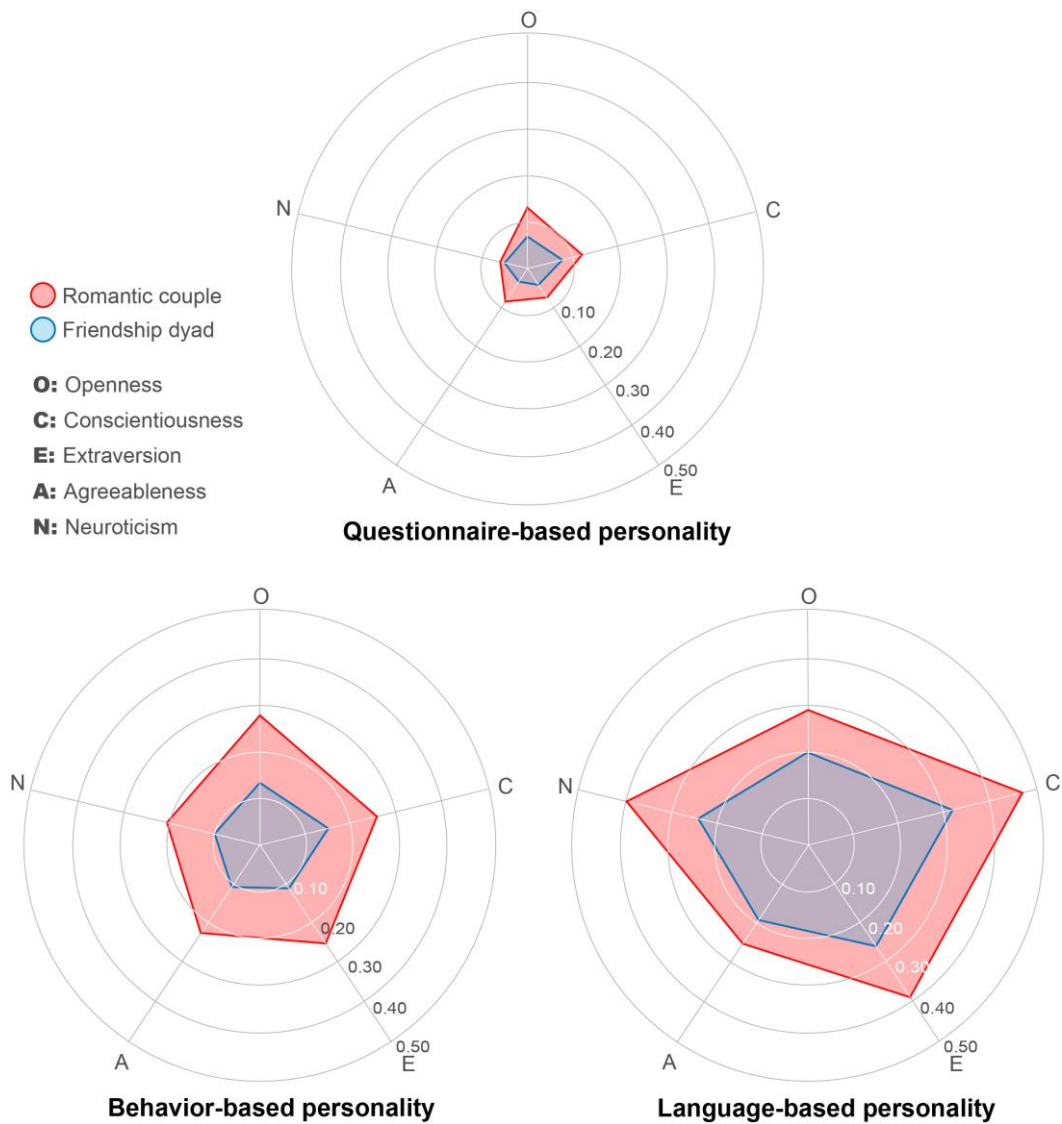


Fig. 1. Personality similarity between romantic partners (red) and between friends (blue). Panels show the similarity correlations for the personality measures based on questionnaire, behavior, or language.

### Discussion

Our findings provide evidence for personality homophily: friends and romantic couples are characterized by similar personality profiles. We employ personality measures relying on three different sources of data: traditional self-report items, behavioral residues, and language use. Consistent with previous research, self-report of personality scores were only weakly correlated between romantic partners or between friends. In contrast, relatively

strong homophily was detected when using behavior and language-based measures. Stronger personality homophily was detected for romantic couples than for friends.

We also showed that dyadic similarity in most personality traits was unlikely due to similarity in age and education. The only exception was romantic partners' similarity in Conscientiousness. Compared with other traits, conscientiousness is most strongly associated with age increase, especially before the age of 30 (Allemand, Zimprich, & Hendriks, 2008; Donnellan & Lucas, 2008; Soto, John, Gosling, & Potter, 2011; Srivastava, John, Gosling, & Potter, 2003). Since 88% of the participants are between 18 and 30 years of age, couples' similarity in Conscientiousness is expected to be partially explained by their similarity in age.

Together, these results challenge the widely-accepted notion that personality homophily does not exist. The scarcity of the evidence for personality homophily is likely to be driven by the reference group effect characteristic of questionnaire-based measures. Notably, our results are consistent with those obtained in (rare) previous studies that relied on personality assessment resistant to the reference-group effect (Botwin et al., 1997; Buss, 1984a, 1984b). The findings imply that personality matching could be used to suggest friends and partners on social networks and dating websites.

Our findings also highlight the limitations pertinent to questionnaire measure (Heine et al., 2002). While self-report in the form of personality questionnaires provides excellent reliability and validity in most applications (Costa & McCrae, 1992; Goldberg et al., 2006; Ozer & Benet-Martinez, 2006; Vazire, 2006), it fails for the purpose of assessing personality similarity between individuals due to its subjective nature. As illustrated in this study, personality assessment based on digital records of behavior and language, while still relatively new and unproven, has the potential to address this issue.

Both behavior and language-based personality assessment are unlikely to be affected by the reference-group effect: people do not use words or do things just because their friends

did *not* do so. In fact, the reverse effect is likely: shared environment, culture, or interpersonal influence may inflate the similarity in behavior and language between dyad members. For example, people from the same cohort are likely to like similar pop stars of their age; two friends might both use the word “explosion” in their status updates because they lived in the same area and an explosion just happened nearby. We addressed this limitation by controlling for age and education, removing Likes shared between dyad members, and applying two disjoint sets of language-based models to dyad members.

We hope that future research will replicate our findings using other personality measures. Additionally, it would be interesting to investigate via longitudinal studies whether personality homophily stems from convergence (the tendency for dyads to become more alike over time) or active assortment (initial preference for similar friends or partners). Research could also explore the association between similarity in personality and various relational outcomes, such as relationship satisfaction.

### References

- Allemand, M., Zimprich, D., & Hendriks, a a J. (2008). Age differences in five personality domains across the life span. *Developmental Psychology, 44*(3), 758–770.
- Altmann, T., Sierau, S., & Roth, M. (2013). I guess you’re just not my type. *Journal of Individual Differences, 34*(2), 105–117.
- Botwin, M. D., Buss, D. M., & Shackelford, T. K. (1997). Personality and mate preferences: five factors in mate selection and marital satisfaction. *Journal of Personality, 65*(1), 107–136.
- Buss, D. M. (1984a). Marital assortment for personality dispositions: assessment with three different data sources. *Behavior Genetics, 14*(2), 111–23.
- Buss, D. M. (1984b). Toward a psychology of person-environment (PE) correlation: The role of spouse selection. *Journal of Personality and Social Psychology, 47*(2), 361–377.
- Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review, 90*(2), 105–126.

- Cohen, A. S., Minor, K. S., Baillie, L. E., & Dahir, A. M. (2008). Clarifying the linguistic signature: measuring personality from natural speech. *Journal of Personality Assessment, 90*(6), 559–563.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual*. Psychological Assessment (Vol. 4). Odessa, FL: Psychological Assessment Resources.
- Credé, M., Bashshur, M., & Niehorster, S. (2010). Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment, 92*(5), 390–399.
- Donnellan, M. B., & Lucas, R. E. (2008). Age differences in the big five across the life span: Evidence from two national samples. *Psychology and Aging, 23*(3), 558–566. doi:10.1037/a0012897
- Eysenck, H. J. (1990). Genetic and environmental contributions to individual differences: the three major dimensions of personality. *Journal of Personality, 58*(1), 245–261.
- Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology, 94*(2), 334–346.
- Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*(4), 656–672.
- Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. In *2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing* (pp. 149–156).
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84–96.
- Heine, S. J., Lehman, D. R., Peng, K., Greenholtz, J., & Of, D. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82*(6), 903–918.
- Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43*(3), 524–527.
- Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality, 64*(2), 311–37.
- Kosinski, M., Stillwell, D., & 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.



- McCrae, R. R., Martin, T. A., Hrebícková, M., Urbánek, T., Boomsma, D. I., Willemsen, G., & Costa, P. T. (2008). Personality trait similarity between spouses in four cultures. *Journal of Personality, 76*(5), 1137–1164.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology, 27*, 415–444.
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*(5), 862–877.
- Ozer, D. J., & Benet-Martinez, V. (2006). Personality And The Prediction Of Consequential Outcomes. *Annual Review Of Psychology, 57*, 401–421.
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2014). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology, 108*(6), 934–952.
- Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: our words, our selves. *Annual Review of Psychology, 54*(1), 547–577.
- Rammstedt, B., & Schupp, J. (2008). Only the congruent survive – Personality similarities in couples. *Personality and Individual Differences, 45*(6), 533–535.
- Rushton, J. P., & Bons, T. A. (2005). Mate choice and friendship in twins: evidence for genetic similarity. *Psychological Science, 16*(7), 555–559.
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... others. (2013). Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PloS One, 8*(9), e73791.
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology, 100*(2), 330–348.
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*(5), 1041–1053.
- Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012* (Vol. 2, pp. 386–393).
- Sumner, C., Byers, A., & Shearing, M. (2011). Determining personality traits & privacy concerns from Facebook activity. In *Black hat briefings '11*.

- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 58(1), 267–288.
- Tikhonov, A. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 5, 1035–1038.
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, 40(5), 472–481.
- Watson, D., Beer, A., & McDade-Montez, E. (2014). The role of active assortment in spousal similarity. *Journal of Personality*, 82(2), 116–129.
- Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: the role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, 78(3), 546–558.
- Watson, D., Klohnen, E. C., Casillas, A., Simms, E. N., Haig, J., & Berry, D. S. (2004). Match makers and deal breakers: analyses of assortative mating in newlywed couples. *Journal of Personality*, 72(5), 1029–1068.
- Wood, A. M., Brown, G. D. a, Maltby, J., & Watkinson, P. (2012). How are personality judgments made? A cognitive model of reference group effects, personality scale responses, and behavioral reactions. *Journal of Personality*, 80(5), 1275–1311.
- Youyou, W., Kosinski, M., & Stillwell, D. J. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 112–116.

**Author Contributions**

W.Y. designed the research; W.Y., A.S., and M.K. performed the analysis; M.K. and D.S. collected and cleaned the data; W.Y., D.S., A.S., and M.K. wrote the paper.

**Acknowledgments:**

We thank John Rust, Patrick Morse, Sandra Matz, Jason Rentfrow, Josh Sacco, and Vesselin Popov for their critical reading of the manuscript.

## Supplemental Materials

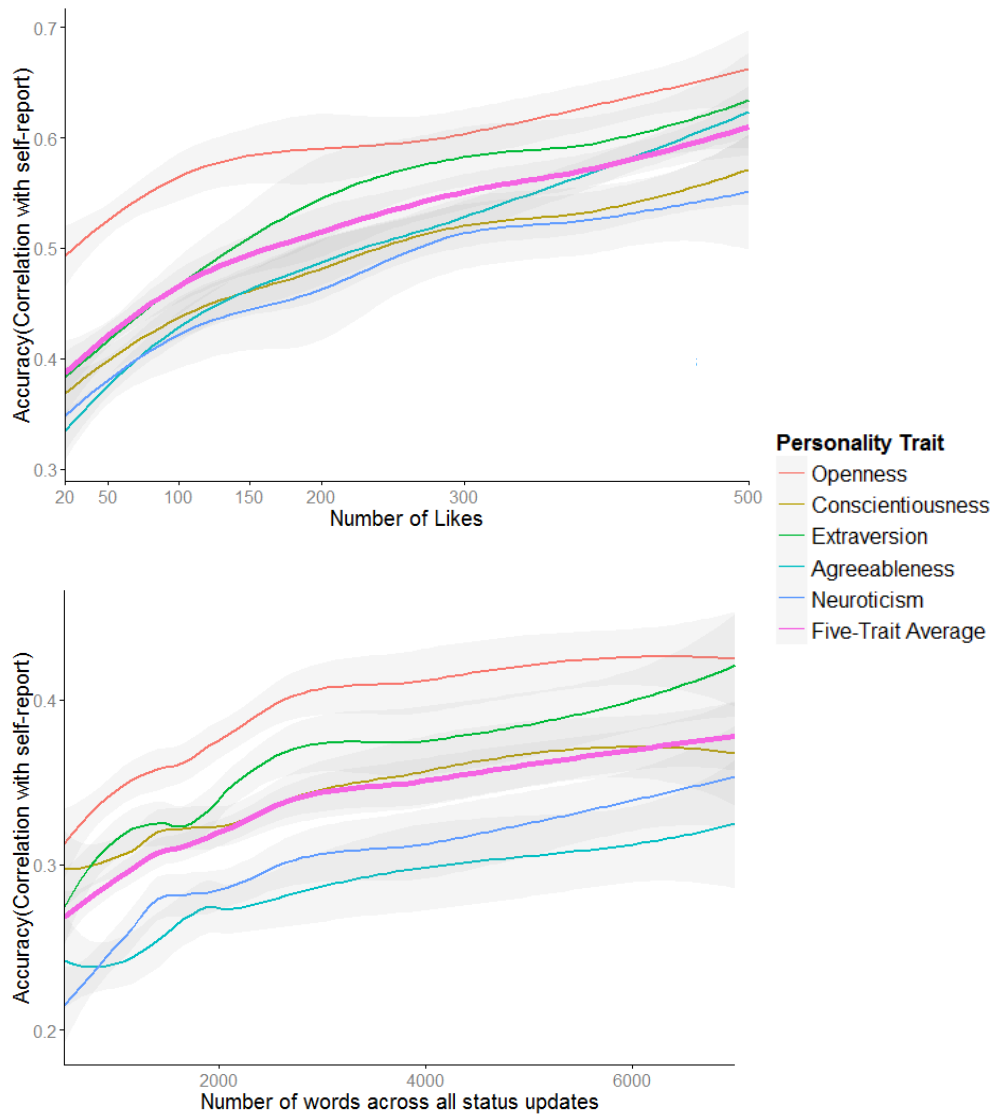


Fig. S1. Accuracy (correlation with self-report) of behavior-based (upper panel) and language-based (lower panel) models, plotted against the number of Likes or the number of words available for predictions. The accuracy curves were smoothed using a LOWESS approach. Grey ribbons represent the 95% confidence interval.

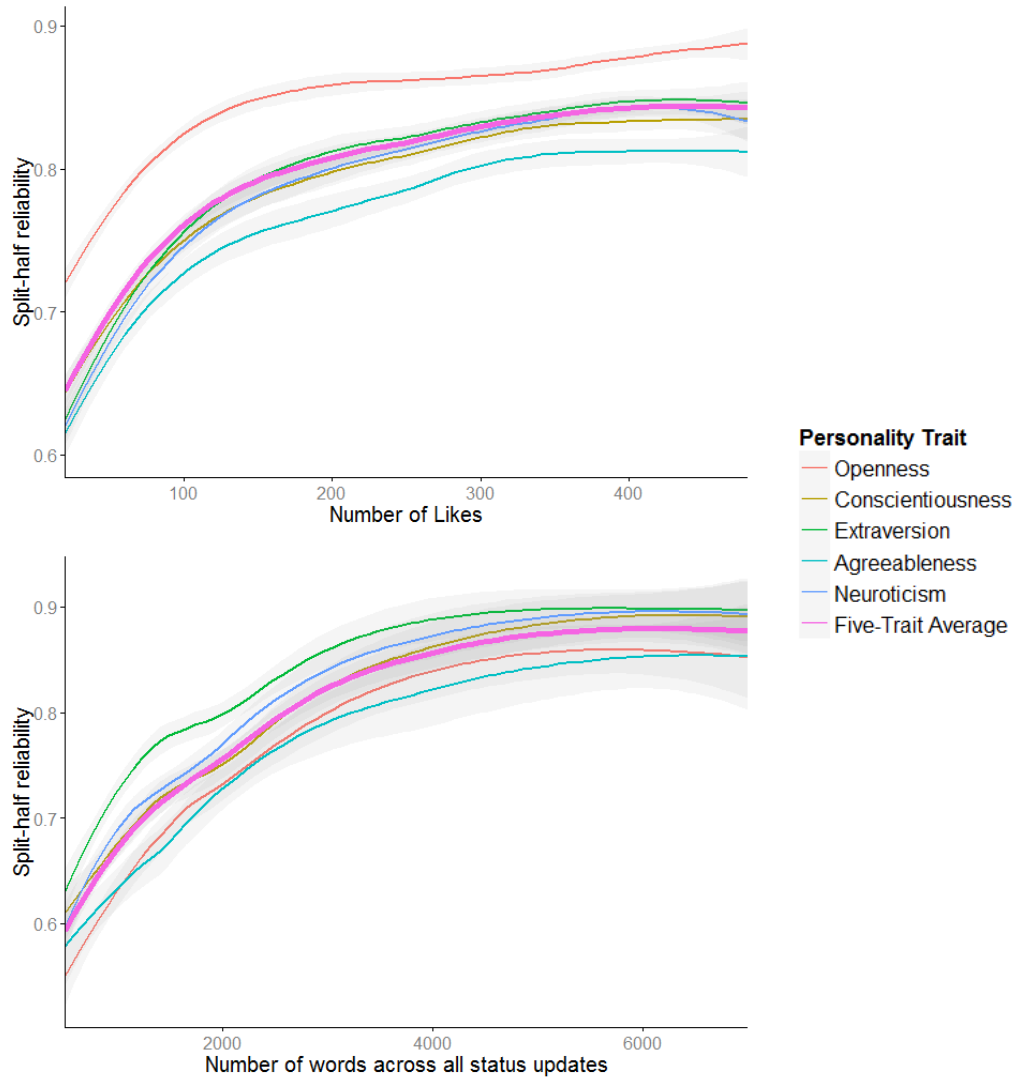


Fig. S2. Split-half reliability of behavior-based (upper panel) and language-based (lower panel) models, plotted against the number of Likes or the number of words available for predictions. The accuracy curves were smoothed using a LOWESS approach. Grey ribbons represent the 95% confidence interval.



Fig. S3. Personality homophily between romantic partners (red) and between friends (blue) *without* controlling for shared Likes or shared language features. Panels show the similarity correlations for the personality measures based on behavior or language.