Language-based Personality:
A New Approach to Personality in a Digital World

Ryan L. Boyd and James W. Pennebaker

The University of Texas at Austin, Department. of Psychology

Correspondence can be sent to the authors at 108 E. Dean Keeton St. Stop A8000, Dept. of Psychology, Austin, TX 78712-1043. e-mails: ryanboyd@utexas.edu, pennebaker@utexas.edu.
Preparation of this manuscript was aided by grants from the National Institute of Health [5R01GM112697-02], John Templeton Foundation [#48503], and the National Science Foundation [IIS-1344257].

Abstract

Personality is typically defined as the consistent set of traits, attitudes, emotions, and behaviors that people have. For several decades, a majority of researchers have tacitly agreed that the gold standard for measuring personality was with self-report questionnaires. Surveys are fast, inexpensive, and display beautiful psychometric properties. A considerable problem with this method, however, is that self-reports reflect only one aspect of personality – people's explicit theories of what they think they are like. We propose a complementary model that draws on a big data solution: the analysis of the words people use. Language use is relatively reliable over time, internally consistent, and differs considerably between people. Language-based measures of personality can be useful for capturing/modeling lower-level personality processes that are more closely associated with important objective behavioral outcomes than traditional personality measures. Additionally, the increasing availability of language data and advances in both statistical methods and technological power are rapidly creating new opportunities for the study of personality at "big data" scale. Such opportunities allow researchers to not only better understand the fundamental nature of personality, but at a scale never before imagined in psychological research.

keywords: language; personality; methods; self-report

Language-based Personality:
A New Approach to Personality in a Digital World

People differ dramatically in the ways they think, feel, and behave in general, forming the basis for what we refer to as personality. Going back to the ancient Greeks, formal thinking about personality has relied on different methods to measure and explain personality. Classically, Galen posited four general temperaments – sanguine, phlegmatic, melancholic, and choleric – based on his observations of biology and the theories of Hippocrates [1]. Freud [2] revolutionized the broader discussion about personality by arguing that inborn temperament and early experiences shaped what people were like later in life. Temperament researchers focused on the activity levels and emotionality of infants to posit the likely genetic and biological bases of individual differences [3]. Others, such as Gordon Allport [4] pointed to the enduring and stable behavioral styles that people possessed – including the ways they walked, gestured, or chewed gum. Even the most nuanced behaviors revealed people's basic characteristics.

Not until the advent of modern social science did psychologists begin to focus on the careful measurement of personality [5–7]. In the last quarter of the 20th century, the trait approach emerged that effectively defined modern personality theory, ushering in detailed factor models of the construct [8,9]. The new trait approach energized the field of personality research, in part because it leaned heavily on self-reports of participants' self-concepts for understanding their general personality characteristics. This was a profound development in personality research: widespread adoption of self-reports meant that it was now possible to have very large groups of people complete extensive personality scales rather than relying on more time- and resource-intensive approaches. Paired with advances in statistical and other computational methods, the adoption of self-report scales resulted in new ways of studying the domains and correlates of traits.

Self-report questionnaires can provide rich information about peoples' conscious self-concepts. However, most personality experts have harbored occasional doubts about the degree to which people's self-reported traits reflect who they really are [10]. For example, to what degree do self-theories map onto their actual behaviors? Across thousands of studies, we know that self-reports correlate nicely with other self-reports from the same people, yet often show lackluster overlap with more objective measures that presumably measure the same underlying trait. Researchers consistently find that widely-used and well-validated self-report measures are insufficient when it comes to forming an accurate understanding of even basic human patterns such as workplace behaviors [11], physical activity [12], expressions of happiness [13] or other emotional states [14].

Are we thinking about personality in the right way? Are people's self-theories the appropriate gold standard for assessing personality? If not self-reports, does a gold standard exist? As we outline below, we must move beyond the gold standard way of thinking. Self-reports reflect one dimension of personality, while nervous system activity may serve as another, genetic factors may be the basis of a third, and so on.

Beyond self-reports and biological markers, recent research has demonstrated that a powerful reflection of personality can be gleaned from the words people use in everyday life. As an increasing number of studies demonstrate, the ways in which people use words is reliable over time, internally consistent, predictive of a wide range of behaviors and even biological activity, and varies considerably from person to person. Language, then, is yet another fundamental dimension of personality. Of great benefit to researchers, and unlike other standard personality markers, people do not need to complete questionnaires or submit to invasive blood or genetic tests in order to provide useful personality data in the form of language.

Language and Personality in the Land of Big Data

Over half of the planet's population uses the internet, and over 80% of people in developed countries are internet users [15]. Every minute, more than 350,000 tweets are posted to Twitter, approximately 3 million Facebook posts are shared, 4 million Google queries are submitted, and over 170 million e-mails are sent [16,17]. In more human terms, the average office worker sees over 120 e-mails per day [18], the typical teen in the United States sends over 60 text messages per day from their mobile phones [19] and the average Facebook user writes 25 comments daily [20]. In short, the amount of language data generated by humans on a minute-by-minute basis around the world is nothing short of staggering.

As with the unprecedented availability of human-generated data, the field of psychology has witnessed a recent cascade in psychometric techniques that are well-suited to a big data research culture. Of the more recent psychological assessment methods, perhaps the most accessible and refined to date is that of automated language analysis, which is currently experiencing rapid adoption and growth across a wide range of academic fields. Historically, psychologists have long believed that a person's words can be revealing of deeper, meaningful psychological constructs [21–23]. For example, classical research on motivation found that the individual's personal strivings, such as the needs for affiliation and achievement, were manifest in their everyday words [24], and it has long been believed that linguistic cues can be used to identify different states of consciousness [25]. However, the modern rejuvenation of language research in the field of personality psychology has been primarily driven by the adoption of modern statistical methods and technological innovations, such as the boom of personal computing power and data accessibility [26].

Unlike most classical research on language and psychology, which typically treated linguistic measures as indicators of a person's transient mental state [27,14], several key studies were conducted early on in the current language analysis renaissance which demonstrated that the properties of language-based psychological measures behave in much the same way as traditional measures of personality. For example, Pennebaker and King [28] explored the psychometric properties of language as a psychological measure, finding that the majority of measures provided by the Linguistic Inquiry and Word Count method [29] exhibited all of the hallmarks of a standard individual differences measure: test-retest reliability, external validity, and internal consistency. A considerable amount research within the LIWC domain has expanded these initial findings, establishing the word-counting paradigm as a robust tool for measuring stable individual differences [30–32].

In the modern research world, where psychologically-relevant data is available in great abundance, psychometric techniques like language analysis allows researchers to indirectly probe and better understand how lower level psychological processes function and interact to manifest in the form of personality in the real world. In other words, techniques such as language analysis are particularly well-suited to the proximal measurement the lower level processes that cohere to form personality, especially in relation to traditional self-report measures. Countless patterns of attention, behaviors, and emotions are deeply embedded in a person's language [31], and psychologists now have access to an ever-growing number of methods to extract these patterns for deeper study.

Given the modern surge of language data as well as methods for extracting psychological information from such data, a logical next step for social scientists is to begin benefiting from the trait-like qualities of language-based measures in psychological research. In the current climate of the “Big Data” revolution, many of the logistical properties for which self-report measures are often lauded ring even truer for language-based measures of personality. While self-reports are relatively easy to collect compared to other measures such as physiological data, language analysis often relies on data that *already exists*. Moreover, pre-existing digital data from the web, smart phones, and social media are inherently ecologically valid, having originated from thoughts and behaviors that occur in the absence of researcher intervention.

It is vital to note that the analysis of language for personality research can be performed *at scale* in nearly any context where language data exists, bypassing the need to recruit and collect constrained self-report measures. While it is a harrowing and costly task to collect self-reported neuroticism from thousands of people, neuroticism’s underlying processes can be measured in millions of Reddit users’ language in an afternoon. As the number of people who use digital technology continues to increase around the world, along with the trails of psychologically actionable data that are left behind, it is imperative that new methods be adopted that are able to make good use of this data by capitalizing on the growing technological infrastructure (e.g., text messages, institutional databases, and social media). In failing to adapt to the new big data world, many personality researchers will be resigned solely to the study self-theories, and only in samples that are directly accessible and motivated to fill out questionnaires.

The Language-based Measurement of Personality

In contrast to most lexical theories of personality, which posit that descriptions of important personality traits are embedded within language *in general* [33–35], it is implicit to current psychological language analysis research that several characteristics of someone's personality are embedded in their unique patterns of language use. However, both approaches generally assume a taxonomical structure of personality – that is, personality as a broad, abstract construct is composed of lower-level psychological processes and behavioral tendencies [36].

The taxonomical structure of personality, both within a general personality psychology framework as well as within a language-based personality framework, is central to performing meaningful personality psychology research. For example, the underlying components of extraversion have been well-established to date across various methodologies: relative to introverts, extraverts generally engage in more social activity [37], experience greater positive affect and well-being [38], and are reactive to external stimulation [39–41]. Indeed, language-based personality research consistently and successfully identifies the same general underpinning processes of extraversion. Relative to their introverted counterparts, extraverts tend to use higher rates of social words, words indicative of positive emotions, and language that is representative of an external focused (i.e., fewer 1st person singular pronouns) [42].

The Two Dominant Modes of Language–Personality Research

Predicting self-report measures. Contemporary language analysis research typically adopts 1 of 2 overarching approaches. In the first approach, researchers seek to build language-based models of personality that approximate the data found in ubiquitous self-report based studies. In other words, one of the most common approaches to language–personality research involves using linguistic measures to *estimate* how people fill out personality self-report questionnaires. For example, Yarkoni [43] explored LIWC- and word-based statistical models of personality by using texts written by bloggers to predict their self-reported Big 5 scores (both overall scores as well as facet-level measures). Similarly, Schwartz et al. [44] adopted an "open-vocabulary" approach to predicting Big 5 self-report measures from Facebook status updates. Such an approach is currently the dominant paradigm in language–personality research and is primarily driven by research teams that lean heavily on a predictive modeling background, crossing boundaries from information sciences to social sciences [45–49].

Under the "estimate self-reports using language" model of study, researchers are ultimately seeking to maximize their account of variance in questionnaire scores via lexical features, and their studies often yield impressive results. Nevertheless, it is conceptually problematic to treat personality as measured by self-report questionnaires as "ground truth" scores for personality research. In part, well-established limitations of such measures, such as self-knowledge constraints and response biases [50], restrict these language-based models of personality to self-theories. More important is that aggregate measures of personality are distal *abstractions* of the very behaviors, feelings, and thoughts that we seek to understand. In estimating peoples' self-reported neuroticism from language, for example, questionnaire scores are treated as a "real" thing that can be objectively measured rather than a collection of supporting psychological processes. In other words, this paradigm treats self-reported personality as a "gold standard" while failing to acknowledge the flaws that they acquire as a part of the operationalization and data collection process.

Measuring personality processes. It is more consistent with modern theories of personality, then, when the use of language in personality research adopts a relatively more atomic demeanor to measuring personality *processes*, rather than traits as a generalized whole.

This alternative approach to language-based research in psychology, while not new, has begun to see increasing adoption among researchers in social and personality psychology.

Recent research has found that many basic cognitive tendencies that give rise to broader individual differences are deeply embedded in language use. For example, linguistic measures of various cognitive patterns are particularly predictive of objective outcomes such as college grades [51,52], life expectancy [53,54], and resilience to trauma [55,56]. Moreover, language-based measures of personality processes have reliable, trait-like properties [28,30]. Further still, such measures are often more predictive of specific, concrete behaviors than traditional self-report measures, providing both stronger and broader predictive coverage [57,58]. Finally, such low-level measures of personality processes may still be aggregated into higher-level abstractions for generalized predictive purposes, much like the work of Yarkoni [43], Schultheiss [59], Schwartz et al. [44], and others.

Particularly vital to Personality Psychology as a field, language-based measures of personality processes allow researchers to better *understand* the psychological features that underpin personality, thereby addressing classical criticisms of trait research being primarily descriptive rather than explanatory [60,61]. For example, Carey et al. [62] extensively debunked the widespread misconception that narcissists are prone to disproportionate self-focus by measuring rates of 1st person singular pronoun use, noting that other psychological processes related to a broader social orientations, including interaction style (e.g., disagreeable social behaviors) and disinhibition (e.g., impulsivity, sensation seeking), are more central pillars of the narcissistic personality [63,64]. Similarly, basic motivational processes that underpin traits such as political ideology, mindfulness, values, social personality, and motivation can be identified and integrated into theoretical understandings of the constructs [13,65–69] – something that is not possible with an approach that relies purely on self-report estimation.

Conclusions

While we have known for some time that self-report questionnaires suffer from critical limitations, personality psychologists have been slow to adopt alternatives. As personality and social psychology have become increasingly integrated [70], research from labs all over the world have found that a person's words say more than what meets the eyes (or ears). Thousands of published studies have demonstrated that language, a powerfully social component of human behavior, contains deeply embedded and hidden information about not just social processes, but also psychological functioning, attentional processes, and other important psychological constructs that are paramount to our understanding of personality. Moreover, new methods of quantifying psychological processes from language are constantly being created. The abundance of language-based methods designed to improve our understanding of psychological processes are particularly relevant and applicable to the modern digital age, where human-generated data is created a rate far beyond what we can currently process.

The future of personality research will continue to experiment with new methods to uncover the psychological processes that are embedded in the massive digital trail of human data. Language analysis for personality research is a low-hanging fruit that is ripe for the picking. In the coming years, the integration of objective multimodal data such as images, language, audio, mobile sensor data, and internet behaviors into more refined measures of personality and its supporting psychological processes are likely to occur. Given that the road has already begun to be paved in words, however, there has never been a better time to transition away from self-reports and towards language analysis as a foundational method in personality research.

References

1. Kagan J: *Galen's prophecy: Temperament in human nature*. Basic Books; 1998.
2. Freud S: *Three essays on the theory of sexuality*. Basic Books; 1905.
3. Thomas A, Chess S, Birch HG: **The origin of personality.** *Sci. Am.* 1970, **223**:102–109.
4. Allport GW: *Pattern and growth in personality.* Holt, Reinhart & Winston; 1961.
5. Allport FH, Allport GW: **Personality traits: Their classification and measurement**. *J. Abnorm. Soc. Psychol.* 1921, **16**:6–40.
6. Cattell RB: *The scientific analysis of personality*. Penguin Books; 1965.
7. Eysenck HJ: **The scientific study of personality**. *Br. J. Stat. Psychol.* 1953, **6**:44–52.
8. Costa PT, McCrae RR: *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. 1992.
9. Goldberg LR: **The structure of phenotypic personality traits**. *Am. Psychol.* 1993, **48**:26–34.
10. McCrae RR, Costa PT: **Self-concept and the stability of personality: Cross-sectional comparisons of self-reports and ratings**. *J. Pers. Soc. Psychol.* 1982, **43**:1282–1292.
11. Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N: **Reconsidering the use of personality tests in personnel selection contexts**. *Pers. Psychol.* 2007, **60**:683–729.
12. Rhodes RE, Smith NEI: **Personality correlates of physical activity: a review and meta-analysis**. *Br. J. Sports Med.* 2006, **40**:958–965.
13. Wojcik SP, Hovasapian A, Graham J, Motyl M, Ditto PH: **Conservatives report, but liberals display, greater happiness**. *Science (80-.).* 2015, **347**:1243–1246.
14. Stiles WB: *Describing talk: A taxonomy of verbal response modes*. 1992.

15. International Telecommunications Union: *ICT facts and figures 2016*. 2016.

16. Micro Focus: **How Much Data is Created on the Internet Each Day [Internet]**. [date unknown], [no volume].

17. Internet Live Stats: **Twitter usage statistics [Internet]**. [date unknown], [no volume].

18. Radicati Group: *Email statistics report, 2015-2019*. 2015.

19. Pew Internet Project: *Teens, smartphones, and texting*. 2012.

20. Leonard H: **This is what an average user does on Facebook [Internet]**. 2013, [no volume].

21. Freud S: *On aphasia*. International Universities Press; 1891.

22. Lasswell HD, Lerner D, de Sola Pool I: *The Comparative Study of Symbols: An Introduction*. Stanford University Press; 1952.

23. Stone PJ, Dunphy DC, Smith MS, Ogilvie DM: *The General Inquirer: A computer approach to content analysis*. M.I.T. Press; 1966.

24. McClelland DC, Atkinson JW, Clark RA, Lowell EL: *The achievement motive.* Irvington; 1953.

25. Martindale C: **The grammar of altered states of consciousness: A semiotic reinterpretation of aspects of psychoanalytic theory**. *Psychoanal. Contemp. Thought* 1975, **4**:331–354.

26. Boyd RL: **Psychological text analysis in the digital humanities**. In *Data analytics in the digital humanities*. Edited by Hai-Jew S. Springer International Publishing; 2017:161–189.

27. Gottschalk LA: **The unobtrusive measurement of psychological states and traits**. In *Text analysis for the social sciences: Methods for drawing statistical inferences from texts*

and transcripts. Edited by Roberts CW. Erlbaum; 1997:117–129.

28. Pennebaker JW, King LA: **Linguistic Styles: Language Use as an Individual Difference**. *J. Pers. Soc. Psychol.* 1999, **77**:1296–1312.

29. Pennebaker JW, Francis ME: *Linguistic Inquiry and Word Count (LIWC): A computer-based text analysis program*. 1999.

30. Boyd RL, Pennebaker JW: **Did Shakespeare write Double Falsehood? Identifying individuals by creating psychological signatures with text analysis.** *Psychol. Sci.* 2015, **26**:570–582.

31. Tausczik YR, Pennebaker JW: **The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods**. *J. Lang. Soc. Psychol.* 2010, **29**:24–54.

32. Pennebaker JW: **The Secret Life of Pronouns: What our Words Say About Us**. 2011, doi:10.1093/llc/fqt006.

33. Hogan R: **A socioanalytic perspective on the five-factor model**. In *The five-factor model of personality: Theoretical perspectives*. Edited by Wiggins JS. Guilford; 1996:163–179.

34. John OP, Angleitner A, Ostendorf F: **The lexical approach to personality: A historical review of trait taxonomic research**. *Eur. J. Pers.* 1988, **2**:171–203.

35. Golubkov S V.: **The language personality theory: An integrative approach to personality on the basis of its language phenomenology**. *Soc. Behav. Pers.* 2002, **30**:571–578.

36. Costa PT, McCrae RR: **Domains and Facets: Hierarchical Personality Assessment Using the Revised NEO Personality Inventory**. *J. Pers. Assess.* 1995, **64**:21–50.

37. Snyder M: **The influence of individuals on situations: Implications for understanding the links between personality and social behavior**. *J. Pers.* 1983, **51**:497–516.

38. Furnham A, Brewin CR: **Personality and happiness**. *Pers. Individ. Dif.* 1990, **11**:1093–1096.

39. Campbell JB: **Differential relationships of extraversion, impulsivity, and sociability to study habits**. *J. Res. Pers.* 1983, **17**:308–314.

40. Philipp RL, Wilde GJS: **Stimulation seeking behaviour and extraversion**. *Acta Psychol. (Amst).* 1970, **32**:269–280.

41. Smits DJM, Boeck PD: **From BIS/BAS to the big five**. *Eur. J. Pers.* 2006, **20**:255–270.

42. Mairesse F, Walker MA, Mehl MR, Moore RK: **Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text**. *J. Artif. Intell. Res.* 2007, **30**:457–500.

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

44. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, Shah A, Kosinski M, Stillwell D, Seligman MEP, et al.: **Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach**. *PLoS One* 2013, **8**:1–16.

45. Badenes H, Bengualid MN, Chen J, Gou L, Haber E, Mahmud J, Nichols JW, Pal A, Schoudt J, Smith BA, et al.: **System U: Automatically Deriving Personality Traits from Social Media for People Recommendation**. In *Proceedings of the 8th ACM Conference on Recommender Systems*. . ACM; 2014:373–374.

46. Chen J, Haber E, Kang R, Hsieh G, Mahmud J: **Making Use of Derived Personality: The Case of Social Media Ad Targeting**. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. . 2015:51–60.

47. Collins S, Sun Y, Kosinski M, Stillwell D, Markuzon N: **Are You Satisfied with Life?:**

Predicting Satisfaction with Life from Facebook. In *Social Computing, Behavioral-Cultural Modeling, and Prediction: 8th International Conference, SBP 2015, Washington, DC, USA, March 31-April 3, 2015. Proceedings*. Edited by Agarwal N, Xu K, Osgood N. Springer International Publishing; 2015:24–33.

48. Komisin M, Guinn C: **Identifying Personality Types Using Document Classification Methods**. In *Proceedings of the Twenty-Fifth International Florida Artificial Intelligence Research Society Conference*. . 2012:232–237.

49. Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, Ungar LH, Seligman MEP: **Automatic personality assessment through social media language.** *J. Pers. Soc. Psychol.* 2015, **108**:934–952.

50. Paulhus DL, Vazire S: **The self-report method**. In *Handbook of research methods in personality psychology*. . Guilford; 2007:224–239.

51. Pennebaker JW, Chung CK, Frazee J, Lavergne GM, Beaver DI: **When Small Words Foretell Academic Success: The Case of College Admissions Essays**. *PLoS One* 2015, **9**:e115844.

52. Robinson RL, Navea R, Ickes W: **Predicting Final Course Performance From Students' Written Self-Introductions**. *J. Lang. Soc. Psychol.* 2013, **32**:469–479.

53. Pressman SD, Cohen S: **Use of social words in autobiographies and longevity**. *Psychosom. Med.* 2007, **69**:262–269.

54. Penzel IB, Persich MR, Boyd RL, Robinson MD: **Linguistic Evidence for the Failure Mindset as a Predictor of Life Span Longevity**. *Ann. Behav. Med.* 2016, doi:10.1007/s12160-016-9857-x.

55. Pennebaker JW, Mayne TJ, Francis ME: **Linguistic predictors of adaptive**

bereavement. *J. Pers. Soc. Psychol.* 1997, **72**:863–871.

56. D’Andrea W, Chiu PH, Casas BR, Deldin P: **Linguistic predictors of post-traumatic stress disorder symptoms following 11 September 2001**. *Appl. Cogn. Psychol.* 2012, **26**:316–323.

57. Boyd R, Wilson S, Pennebaker J, Kosinski M, Stillwell D, Mihalcea R: **Values in Words: Using Language to Evaluate and Understand Personal Values**. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. . 2015:31–40.

58. Fast LA, Funder DC: **Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior.** *J. Pers. Soc. Psychol.* 2008, **94**:334–346.

59. Schultheiss O: **Are implicit motives revealed in mere words? Testing the marker-word hypothesis with computer-based text analysis**. *Front. Psychol.* 2013, **4**:748.

60. Pervin LA: **A Critical Analysis of Current Trait**. *Psychol. Inq.* 1994, **5**:103–113.

61. Wilson M, De Boeck P: **Descriptive and explanatory item response models**. In *Explanatory item response models: A generalized linear and nonlinear approach*. Edited by Wilson M, De Boeck P. Springer; 2004:43–74.

62. Carey AL, Brucks MS, Küfner ACP, Holtzman NS, große Deters F, Back MD, Donnellan MB, Pennebaker JW, Mehl MR: **Narcissism and the use of personal pronouns revisited.** *J. Pers. Soc. Psychol.* 2015, **109**:e1–e15.

63. Vazire S, Funder DC: **Impulsivity and the Self-Defeating Behavior of Narcissists**. *Personal. Soc. Psychol. Rev.* 2006, **10**:154–165.

64. Miller JD, Campbell WK, Young DL, Lakey CE, Reidy DE, Zeichner A, Goodie AS: **Examining the Relations Among Narcissism, Impulsivity, and Self-Defeating Behaviors**. *J. Pers.* 2009, **77**:761–794.

65. Collins SE, Chawla N, Hsu SH, Grow J, Otto JM, Marlatt GA: **Language-based measures of mindfulness: Initial validity and clinical utility.** *Psychol. Addict. Behav.* 2009, **23**:743–749.

66. Fetterman AK, Boyd RL, Robinson MD: **Power Versus Affiliation in Political Ideology**. *Personal. Soc. Psychol. Bull.* 2015, **41**:1195–1206.

67. Johnsen J-AK, Vambheim SM, Wynn R, Wangberg SC: **Language of motivation and emotion in an internet support group for smoking cessation: explorative use of automated content analysis to measure regulatory focus**. *Psychol. Res. Behav. Manag.* 2014, **7**:19–29.

68. Kacewicz E, Pennebaker JW, Davis M, Jeon M, Graesser AC: **Pronoun use reflects standings in social hierarchies**. *J. Lang. Soc. Psychol.* 2013, **33**:125–143.

69. Reysen S, Pierce L, Mazambani G, Mohebpour I, Puryear C, Snider JS, Gibson S, Blake ME: **Construction and Initial Validation of a Dictionary for Global Citizen Linguistic Markers**. *Int. J. Cyber Behav. Psychol. Learn.* 2014, **4**:1–15.

70. Snyder M, Deaux K: **Personality and social psychology: Crossing boundaries and integrating perspectives**. In *The Oxford handbook of personality and social psychology*. Edited by Snyder M, Deaux K. Oxford University Press; 2012:3–9.

Reference Annotations

13. Wojcik SP, Hovasapian A, Graham J, Motyl M, Ditto PH: **Conservatives report, but liberals display, greater happiness. *Science (80-.).* 2015, **347**:1243–1246.

A long-running debate in psychology is whether conservatives or liberals are more happy, in general. While past research has repeatedly found that conservatives *report* greater happiness in self-report paradigms, liberals actually *exhibit* greater happiness, as quantified in their language and other behavioral measures.

---

*30. Boyd RL, Pennebaker JW: **Did Shakespeare write Double Falsehood? Identifying individuals by creating psychological signatures with text analysis.** *Psychol. Sci.* 2015, **26**:570–582.

The authors used language-based measures of personality processes to successfully differentiate multiple people, ultimately determining that Shakespeare was the primary author of a disputed play. By using psychological language analysis, the differentiating linguistic measures were able to be interpreted in light of observer reports of different people, providing high convergence.

---

*46. Chen J, Haber E, Kang R, Hsieh G, Mahmud J: **Making Use of Derived Personality: The Case of Social Media Ad Targeting**. In *Proceedings of the Ninth International AAAI*

Conference on Web and Social Media. . 2015:51–60.

The authors used language samples to estimate self-report scores for the Big 5, then used these estimated scores to model responsiveness to targeted advertising. This work is an example of the many ways in which personality is often misconceptualized when studied from a predictive modeling viewpoint.

---

*49. Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell DJ, Ungar LH, Seligman MEP: **Automatic personality assessment through social media language.** *J. Pers. Soc. Psychol.* 2015, **108**:934–952.

One of several impressive studies where the stated goal is to maximize the variance accounted for in self-report personality questionnaires. The authors of this study demonstrated a new approach to estimating how people typically respond to self-reported measures of personality by using the language that people share on social media.

---

57. Boyd R, Wilson S, Pennebaker J, Kosinski M, Stillwell D, Mihalcea R: **Values in Words: Using Language to Evaluate and Understand Personal Values. In *Proceedings of the Ninth International AAAI Conference on Web and Social Media*. . 2015:31–40.

The authors introduced a new method for establishing language-based measures of core values. This research found that the language-based measures of values showed

poor convergence with self-reported values yet were vastly superior in terms of predictive strength and coverage when modeling the important relationship between values and behavior found in the real world.

---

62. Carey AL, Bucks MS, Küfner ACP, Holtzman NS, große Deters F, Back MD, Donnellan MB, Pennebaker JW, Mehl MR: **Narcissism and the use of personal pronouns revisited. *J. Pers. Soc. Psychol.* 2015, **109**:e1–e15.

The authors found that, contrary to both layperson and expert assumptions, narcissism is not associated with more self-focused language. This research is a prime example of how psychological language analysis can be extremely informative for personality theory and clarifying misguided assumptions.