Social media-predicted personality traits and values can help match people to their ideal jobs

Margaret L. Kern[a,1], Paul X. McCarthy[b,c], Deepanjan Chakrabarty[b,c], and Marian-Andrei Rizoiu[d]

[a]Melbourne Graduate School of Education, The University of Melbourne, Parkville, VIC 3010, Australia; [b]Ribit.net, Data61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Eveleigh, NSW 2015, Australia; [c]Computer Science and Engineering, University of New South Wales, Kensington, NSW 2052, Australia; and [d]Faculty of Engineering and Information Technology, The University of Technology Sydney, Ultimo NSW 2007, Australia

Work is thought to be more enjoyable and beneficial to individuals and society when there is congruence between one's personality and one's occupation. We provide large-scale evidence that occupations have distinctive psychological profiles, which can successfully be predicted from linguistic information unobtrusively collected through social media. Based on 128,279 Twitter users representing 3,513 occupations, we automatically assess user personalities and visually map the personality profiles of different professions. Similar occupations cluster together, pointing to specific sets of jobs that one might be well suited for. Observations that contradict existing classifications may point to emerging occupations relevant to the 21st century workplace. Findings illustrate how social media can be used to match people to their ideal occupation.

personality | employment | linguistic analysis | social media | 21st century workplace

Imagine that you are a young adult looking for work. You want a job that not only pays the bills, but also one that you will succeed at and enjoy—after all, it will consume most of your waking hours. How do you find the right profession?

The US Bureau of Labor Statistics (1) classifies occupations into 867 categories, which encompass tens of thousands of specific job titles. Yet many occupations that will be needed in the coming decades do not yet exist, and many existing categories are becoming obsolete (2, 3). Organizations are increasingly concerned that employee skills are mismatched with industry requirements, with 1 in 3 people being underqualified and 1 in 4 overqualified for their current positions (4). Many employees also desire meaningful careers, such that their work contributes not only to their financial wellbeing but also to their psychological wellbeing (5). Yet only 20% to 30% of workers globally report feeling engaged in their work, and 18% of workers are actively disengaged (6).

Scholars and practitioners have long suggested that work is more likely to be enjoyable and beneficial to the individual and society when there is congruence between the person and the occupation (7, 8). Since the 1960s, psychologists have suggested that one's personality provides an important clue toward the occupations that one will succeed at (8). "Personality" refers to the biopsychosocial characteristics that distinguish a person, which include dispositional traits, contextualized features of the person (e.g., values, goals, motivations), and integrative life narratives (9). Here, we specifically focus on traits and values.

"Traits" refer to relatively consistent ways of thinking, behaving, and feeling across situations (10). "Values" represent the things in life that are most important to a person (9, 11). A number of measurable schema of traits and values exist; here we focus on "the Big 5" (10), which classify traits into 5 broad factors (extraversion, agreeableness, conscientiousness, emotional stability, and openness), and 5 of Schwartz's "basic values" (11), which identify personal values that are generally recognized across cultures (helping others, tradition, taking pleasure in life, achieving success, excitement).

Distinctive personality profiles appear across a range of occupations (12, 13). A study of 8,458 employed individuals found that individuals who held a job that fitted their personality were more likely to earn up to 10% greater income (14). Studies also find that the Big 5 predict meaningful life outcomes, including physical and mental health, longevity, social relationships, health-related behaviors, antisocial behavior, and social contribution, at levels on par with intelligence and socioeconomic status (15–17). Values are closely tied to the self, express motivational goals, and distally impact behavior (18).

As people engage with social media, they leave behind digital fingerprints—behavioral traces of their personality—which can be detected at a large scale (19–22). Linguistic analyses of social media information have been used to predict an array of outcomes, including age, gender, political orientation, physical and mental illness, and unemployment (22–25). However, associations between these factors and career success across a broad range of occupations are unknown.

Here, we present a 21st century approach for matching one's personality with congruent occupations by applying machine-learning approaches to linguistic information publicly available through online social media (i.e., Twitter), based on 128,279 users representing 3,513 occupations.

Matching Personality Digital Fingerprints with Occupations

As a proof of concept, we first used a select set of occupations among a small number of users to test whether different personality digital fingerprints—based on Big 5 scores derived from linguistic information available from Twitter—could be linked

www.manaraa.com

to specific occupations. We hypothesized that each occupation would have a distinctive profile and that similar occupations (e.g., computer programmers and scientists) would have similar digital fingerprints, whereas dissimilar occupations (e.g., computer programmers and athletes) would have distinctive digital fingerprints.

Fig. 1A provides a "dot painting" of the Big 5 digital fingerprints for 1,035 users across 9 occupations. Individuals' scores
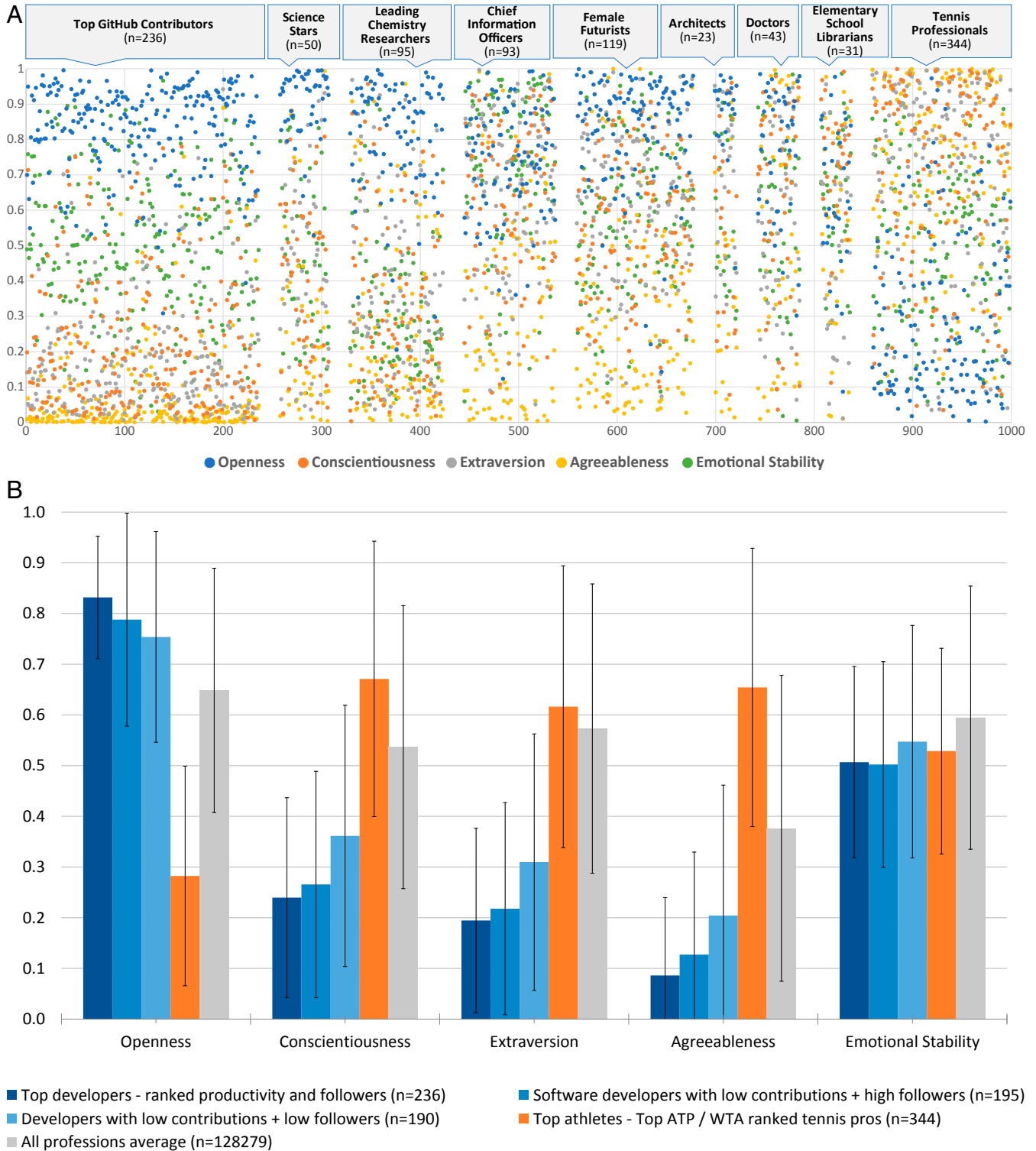
Fig. 1. (*A*) Big 5 dot painting, providing digital fingerprints of 1,035 individuals across 9 occupations. Each dot corresponds to a user, with people grouped within their self-identified occupation. (*B*) Big 5 profile comparison. Shown are the Big 5 personality profiles for 621 software developers with varying levels of success (based on productivity and peer influence: dark blue bars, top GitHub contributors; medium blue bars, influential GitHub contributors; light blue bars, mainstream GitHub contributors), those for professional tennis players (orange bars), and mean values for the sample of 128,279 users (gray bars). The error bars show 1 SD for each sample. ATP = Association of Tennis Professionals; WTA = Women's Tennis Association.

for each of the Big 5 traits are visualized, with higher scores at the top of the graph. Software programmers, science stars, and top chemistry researchers appeared to be more open (indicated by dark blue dots high on the graph) and less agreeable and conscientious (indicated by yellow and orange dots low on the graph), whereas tennis players were less open and more conscientious and agreeable. Architects, female futurists, and chief information officers tended toward greater openness and emotional stability and less agreeableness, whereas librarians and doctors presented mixed profiles.

To further explore evidence for similarities within occupations, we drew a set of 621 open source software developers with active profiles on the GitHub repository and classified them as being top GitHub contributors, influential GitHub contributors, or mainstream GitHub contributors. Fig. 1B illustrates the median Big 5 profile for these 3 sets of GitHub contributors, along with the median profiles of the professional tennis players and the median of all 128,279 users in our dataset for comparison. For all but emotional stability, the GitHub contributors' profiles (blue bars) and tennis players profiles (orange bars) were opposite, with contributors being relatively high on openness and low on conscientiousness, agreeableness, and extraversion and tennis players being relatively low on openness and high on conscientiousness, agreeableness, and extraversion. Patterns were more distinctive for top GitHub contributors (dark blue bars), whereas mainstream contributors were similar to the full sample.

Aligned with prior studies that have used linguistic information on social media as indicators of personality (19, 21, 22), we observed that distinctive digital fingerprints occurred across users, which could be detected from their Twitter language. These fingerprints aligned with different occupations, with

greater alignment for similar occupations (in terms of the cognitive and noncognitive skills required by the occupation) and greater differentiation for individuals who were most successful within an occupation (as shown by the top contributors compared to mainstream contributors and by successful tennis professionals compared to amateur players that likely exist within the full sample of Twitter users).

Mapping Vocations Based on Psychological Profiles

Replicating these similarities and differences at a large scale, we used the psychological profiles of more than 100,000 users to build a vocations map—a 2D visualization that clustered occupations based on their personality digital fingerprints. From our dataset of 128,279 users, we selected occupations that had a minimum of 50 users within a given occupation, resulting in 101,152 users representing 1,227 professions. We included both Big 5 and 5 basic value scores, resulting in a 10-dimensional numerical vector representing the personality digital fingerprints of each user. We then computed occupation profiles by aggregating all individuals with the same occupation and automatically clustered occupations based on profile similarity. We expected that occupations that are classified within the same categories within the US Standard Occupation Classification (1) would cluster together.

The vocations map (Fig. 2) visually illustrates the distances among 20 medoids (i.e., the occupation at the middle of the cluster), automatically discovered from the data, with the other occupations clustered around these medoids (see http://bit.ly/vocation-map-interactive for an interactive version). Fig. 2, *Insets* zoom into 2 clusters (concert manager and software programmer), illustrating occupations that clustered within each one.

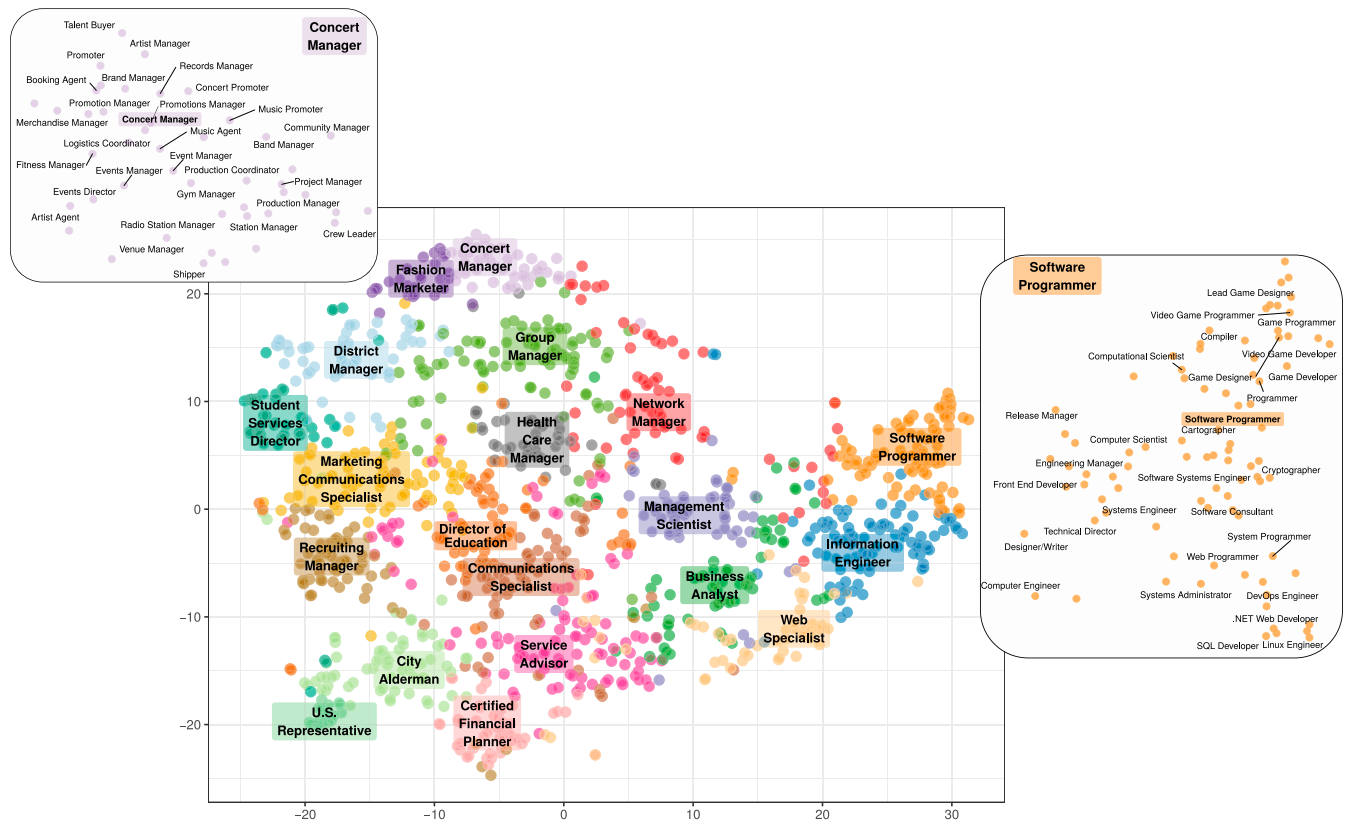**Fig. 2.** The vocations map. Vocations are clustered by the predicted personality digital fingerprints of 101,152 Twitter users, across 1,227 occupations. *Insets* illustrate specific job titles that are part of the software programmer (*Right*) and concert manager (*Upper Left*) clusters. An interactive version of this map is at http://bit.ly/vocation-map-interactive.

Kern et al.

www.manaraa.com

Clear clusters emerged around technology (with software and science roles in Fig. 2, *Right Inset*) and music, fashion, arts, and education (Fig. 2, *Upper Left Inset*). The bottom part of the map in Fig. 2 includes managers, advisers, and politicians.

While many of the combinations align with existing categories in the US Standard Occupation Classification (supporting the validity of the map), some jobs appeared in alternative clusters. For instance, nurse managers clustered with campaigners and box office managers, rather than being part of a medical cluster. This alignment makes sense based on the skills required for the jobs; similar to campaigners and box office managers, nurse managers must work with a number of internal and external people, manage customer relationships, and deal with intense periods of high stress.

Differences between a priori occupational categories based on the Standard Occupation Classification and those arising from the automatic clustering may also capture an evolution of occupations. For instance, traditional forms of cartography, although a common occupation in the past, are becoming a lost art (26). Alternatives, evident in the software programmer cluster, include DevOps—a fast-growing occupation that combines software development and information technology operations (27).

Predicting Occupation from Personality Digital Fingerprints

The vocations map suggests that personality digital fingerprints cluster into specific occupational clusters, supporting the use of linguistic information from social media to identify good-fitting jobs based on one's personality, both for existing and for future occupations. However, the map's utility depends on how accurately one's occupation can be determined. We selected 10 professions with the largest number of users, resulting in a balanced subset of 9,550 individuals (955 in each class). We trained a machine-learning algorithm and tested how accurately an individual's occupation could be predicted, based on 5 classifiers, using 10-fold cross-validation. We compared the predictions with the observed profession using the accuracy measure, which can be interpreted as the probability that each prediction is correct (note that the prediction for each user can be made using only the Big 5, only the 5 basic values, or all 10 features).

Fig. 3*A* plots the performance for each classifier, using only the 5 traits, only the 5 values, or all 10 features. Each barplot shows the mean accuracy over the 10-folds, with the error bars indicating the SD. All classifiers obtained an accuracy higher than 70%, with the best performance obtained by eXtreme Gradient Boosting (XGBoost). This suggests that user occupations could indeed be successfully predicted from their personality digital fingerprints. Predictions using the Big 5 yielded slightly more accurate results than predictions using the basic values. Predictions using both sets of features boosted accuracy by almost 10%, indicating that the traits and values are complementary in predicting user occupations.

We also investigated cases where prediction failed. Fig. 3*B* shows the confusion matrix for XGBoost, which contains 10 rows (indicating the predicted value) and 10 columns (indicating the actual occupation) corresponding to 10 professions. Cells indicate the confusion rate or how many times the observed occupation differs from the predicted occupation; darker shades indicate greater confusion (greater error). Rows and columns are ordered based on the confusion rate (indicated by dendrograms).

Two pairs of occupations were often mistaken for each other: school principal and superintendent and data scientist and software engineer. Both pairs require similar skill sets, and indeed one might precede the other. Interestingly, the confusion rates were not symmetrical: School principals were more often confused with teachers than the other way around—which makes
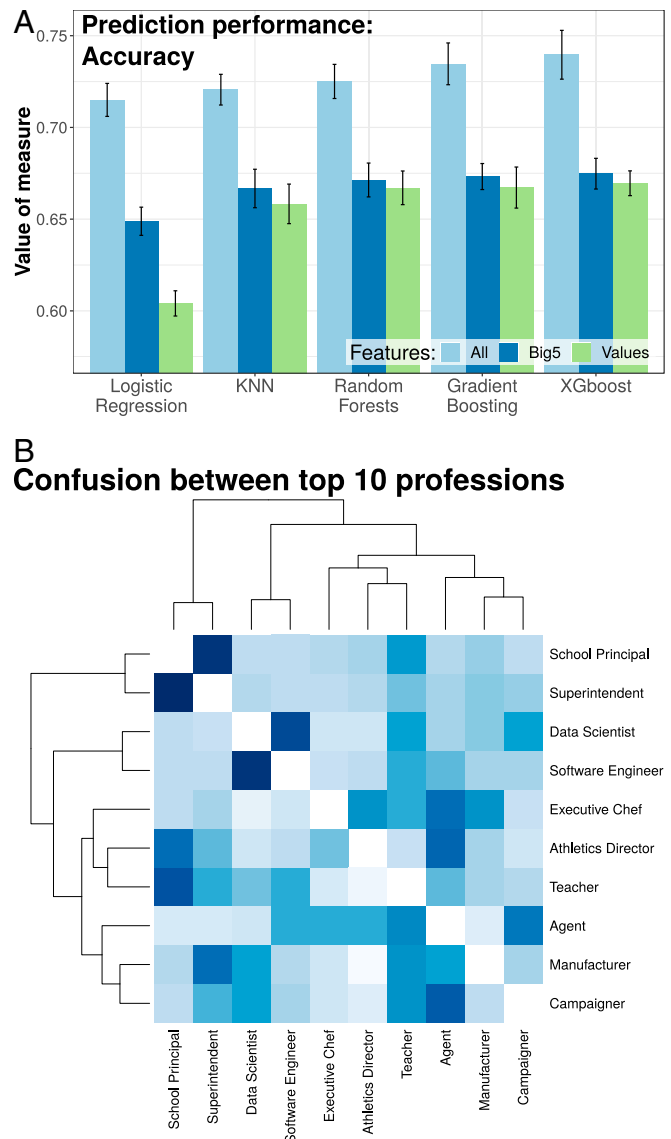

A Prediction performance: Accuracy

B Confusion between top 10 professions

Fig. 3. (*A*) Prediction accuracy (mean and SD) for the top 10 professions. The traits and values are complementary features; using them jointly boosted prediction accuracy by almost 10%. (*B*) Confusion heat map illustrates which of the top 10 professions are most often mistaken for one another in the machine-learning model predictions, with errors indicated by a darker blue color.

sense, as most principals are at some point teachers, but only some teachers become principals.

These results suggest that user occupations are predictable based on their psychological profiles. When the classifier was mistaken, it predicted occupations with similar skill sets. This is reassuring in considering applications of automatic recommendations, suggesting that the recommended occupation would not stray too far away from a person's "ideal match."

Discussion and Conclusion

Using a large dataset, information unobtrusively available online (i.e., Twitter language), and a combination of Big 5 traits and 5 basic values, our study suggests that personality digital fingerprints relate to distinctive occupations. Our analytic approach potentially provides an alternative for identifying occupations which might interest a person, as opposed to relying upon extensive self-report assessments. Notably, while many of the

www.manaraa.com

occupations that clustered together are intuitively related, occupations that rely on similar skill sets and interests that are not traditionally part of an occupational category may point to alternative vocations that might provide good matches for a person.

Our results demonstrate the potential to create an atlas of career aptitude, based on noncognitive personality traits and values. We anticipate that this could have significant applications in career guidance for new graduates, disengaged employees, career changers, and the unemployed.

Occupations that clustered together also may provide an indication of up-and-coming jobs that might play an important role in the 21st century workplace. For jobs that are disappearing due to automation, a data-driven atlas could reveal which emerging occupations are aligned with those that are disappearing, based on one's personality.

The sample used here consisted of English-speaking Twitter users who included their occupation on their profile and with sufficient linguistic data, such that the pattern of results may not generalize to broader populations. Still, our results illustrate the value of applying data analytic approaches to social media data for practical applications. A similar approach potentially could be applied to other platforms. For instance, a service could be developed where posts across a range of sites could be compiled, and the methods provided here could be used to identify potential suitable occupations.

Work is a core part of human life; comprises most of our waking hours; and impacts the physical, mental, social, and economic wellbeing of individuals and communities (28). Many people desire an occupation that aligns with who they are as an individual. As people broadcast their lives online, they create digital fingerprints, creating the possibility for a modern approach to matching one's personality and occupation and ultimately supporting the wellbeing and success of individuals, organizations, and society.

Materials and Methods

We began with 15,000 job titles from the US Bureau of Labor Statistics (1). Using the Twitter Application Programming Interface (API), we selected 1.5 million English-speaking Twitter users who self-identified these job titles in their Twitter profile field and obtained their latest 200 tweets. We then used IBM Watson's system to obtain normalized trait and value scores for each user. Sufficient linguistic data were available to determine the digital fingerprints for 128,279 users, representing 3,513 occupations.

Creating Personality Digital Fingerprints. To automatically determine each user's personality digital fingerprint, we used the IBM Watson Personality Insights system (29), which is a commercial service that, among other services, uses linguistic data available through digital sources (such as social media) to infer personality characteristics of users (30). IBM Watson provides an API that gathers linguistic information from digital sources such as Twitter. An open-vocabulary machine-learning approach computes raw trait and value scores for each user. These raw scores are then compared to a reference population to determine percentiles corresponding to the user's raw values. For example, a percentile of 0.649 for extraversion indicates that the user's extraversion score is in the 65th percentile compared to the reference population. The percentiles scores are normalized scores, representing a percentile ranking for each characteristic as inferred from the input text.

The mean absolute error provides an indication of the estimated difference between the predicted scores (e.g., a person's estimated extraversion score) and the actual score (e.g., their true extraversion score). Compared to self-reported surveys, IBM Watson estimates error rates of 12% for the Big 5 and 11% for the 5 basic values.

To create personality digital fingerprints, we first used the 5 traits as a proof of concept. Then, to provide a more robust fingerprint for the vocation map and occupation predictions, we added the 5 basic values. Each user's personality digital fingerprint can thus be represented by a 5-dimensional numerical vector, representing the Big 5 traits or the 5 basic values, or by a 10-dimensional numerical vector, representing both traits and values.

Aligning Personalities and Occupations. As a proof of concept, we began with the Big 5 traits. We hand curated a dataset of 1,035 users across 9 occupations. We selected occupations for which existed readily available public lists of people in these roles, such as the majority of top-ranked tennis professionals and GitHub's most productive open source software contributors. For other categories, such as science stars and futurists, we used publicly available lists of people with a common job title, which we mapped to their Twitter user ID. (See *SI Appendix* for additional details, including rationale, sources, and number of users selected from each occupation.) We visually created the Big 5 dot painting (Fig. 1A), which provides a scatterplot of the Big 5 traits across the 9 occupations, with users in the same profession grouped together.

To further explore evidence for similarities within occupations, we drew an additional set of 621 open source software developers with active profiles on the GitHub repository (http://www.github.com), representing varying levels of impact as a programmer. Open source software developers have data readily available in terms of their productivity (indicated by the number of posts and commits to GitHub) and their peer influence within the GitHub community (indicated by the number of their followers). Based on productivity and peer influence, we created 3 groups: top GitHub contributors (n = 236), each with over 500 posts and over 1,000 followers; influential GitHub contributors (n = 190) with 200 to 500 contributions and over 1,000 followers; and mainstream GitHub contributors (n = 195), with fewer than 200 posts and fewer than 1,000 followers. We visually compared median Big 5 profiles for each programmer group, tennis professionals (n = 170), and the full set of 128,279 users (Fig. 1B).

Developing the Vocations Map. We returned to the user dataset and selected occupations that had a minimum of 50 users within a given occupation. This resulted in 101,152 users representing 1,227 occupations. To provide a more robust indication of one's digital fingerprint, we included both the Big 5 traits and 5 basic values, resulting in a 10-dimensional numerical vector for each user. For each occupation with a minimum of 50 users, we computed the median values for each of the 10 traits and values for users with that occupation.

Given the profiles of 2 professions $u = [u_i; i = 1..10]$ and $v = [v_i; i = 1..10]$, we computed their similarity using the Euclidean distance:

$$dist(u, v) = \sqrt{\sum_{i=1}^{10} (u_i - v_i)^2}. \tag{1}$$

We also tested the cosine distance but found it achieved lower performances for the clustering of the occupations.

We employed Partitioning Around Medoids (PAM) (31), an unsupervised machine-learning algorithm that automatically partitions the dataset into nonoverlapping groups, specifying 20 clusters (see *SI Appendix* for details). PAM aims to automatically uncover the "optimal" partition, in which occupations within one cluster are as similar as possible and as dissimilar to occupations in other clusters as possible. This ensures that occupations in one cluster are coherent in term of their similarity, based on the trait and value median scores for each occupation.

PAM chooses existing points in the dataset to serve as centers or medoids. The medoid is the object of a cluster whose average dissimilarity to all objects within the cluster is minimized (i.e., it is the most centrally located point in the cluster within the 10-dimensional space). Each occupation is assigned to a single cluster based on the minimal distance between that occupation and the medoid, compared to other medoids. PAM automatically discovers the clusters and the medoids simultaneously. Note that the clustering is performed on occupation profiles (i.e., the aggregates of individuals within an occupation), rather than on individuals themselves.

We then used the t-distributed stochastic neighbor embedding (t-SNE) (32) to visualize the 10-dimensional space of the profession profiles in 2D space, which we call the vocations map (Fig. 2).

Occupation Prediction. Intuitively, for given users, we could see where their profile fits within the 10-dimensional space and identify the closest occupations. In practice, we trained a machine-learning algorithm to learn a nonlinear map between user profiles and occupations on one set of data and then tested how accurately one's occupation could be predicted in a second set of data.

We selected 10 of the largest occupations: agent, athletics director, campaigner, data scientist, executive chef, manufacturer, school principal, software engineer, superintendent, and teacher. Of these 10 occupations, the smallest one included 955 individuals. For balance, we randomly

sampled 955 individuals from each occupation, resulting in a subset of 9,550 individuals.

We trained and tested 5 off-the-shelf machine-learning classifiers: k nearest neighbor (KNN), logistic regression, random forests (33), gradient boosted decision trees (34), and XGBoost (35). Each of the 5 classifiers has hyperparameters (i.e., parameters that impact performance but are not learned from the data), which we tuned using randomized-search 3-fold cross-validation each time they were learned. On each tuning, we performed 40 random search iterations (i.e., 40 combinations of hyperparameters were tried).

The results were obtained through 10-fold cross-validation, in which the dataset was divided into 10-folds, and a prediction model was developed based on 9-folds and then tested on the 10th fold. This was repeated, such that each fold served as the test set once, resulting in a final prediction for each individual in the dataset. We compared the prediction with the observed (ground truth) profession and we computed 4 standard performance measures: accuracy, precision, recall, and f1. The results for accuracy are shown in Fig. 3*A* (see *SI Appendix* for the others). We repeated the training and testing of the models 3 times, with only the Big 5, only the 5 basic values, or all 10 features. The results obtained for each setup are shown as bars of different colors in Fig. 3*A*.

Data Availability. The codes for reproducing the vocation map and the user profession predictions are available at https://github.com/behavioral-ds/VocationMap. The Twitter user data will be made available on demand on a case basis only, as per the Twitter Terms of Service.

1. US Bureau of Labor Statistics, *Standard Occupation Classification Manual* (US Bureau of Labor Statistics, Washington, DC, 2016).
2. W. F. Cascio, R. Montealegre, How technology is changing work and organizations. *Annu. Rev. Organ. Psychol. Organ. Behav.* **3**, 349–375 (2016).
3. World Economic Forum, The Future of Jobs: Employment, skills and workforce strategy for the fourth industrial revolution. http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf. Accessed 12 October 2018.
4. G. Quintini, "Over-qualified or under-skilled: A review of existing literature" (Tech. Rep. 121, OECD Publishing, Paris, France, 2011; https://doi.org/10.1787/5kg58j9d7b6d-en).
5. A Hurst *et al.*, Purpose at work: 2016 workforce purpose index. https://cdn.imperative.com/media/public/Global_Purpose_Index_2016.pdf. Accessed 27 November 2018.
6. Gallup, Inc., *State of the Global Workplace* (Gallup Press, New York, NY, ed. 1, 2017).
7. M. R. Barrick, M. K. Mount, The big five personality dimensions and job performance: A meta-analysis. *Pers. Psychol.* **44**, 1–26 (1991).
8. J. L. Holland, *The Psychology of Vocational Choice: A Theory of Personality Types and Model Environments* (Blaisdell, Oxford, England, 1966).
9. D. P. McAdams, B. D. Olson, Personality development: Continuity and change over the life course. *Annu. Rev. Psychol.* **61**, 517–542 (2010).
10. O. P. John, S. Srivastava, "The big five trait taxonomy: History, measurement, and theoretical perspectives" in *Handbook of Personality: Theory and Research*, L. A. Pervin, O. P. John, Eds. (Guildford Press, New York, NY, 1999), pp. 102–138.
11. S. H. Schwartz, An overview of the Schwartz theory of basic values. *Online Read. Psychol. Cult.* **2** (2012). https://scholarworks.gvsu.edu/orpc/vol2/iss1/11. Accessed 12 October 2018.
12. J. W. Lounsbury, R. P. Steel, L. W. Gibson, A. W. Drost, Personality traits and career satisfaction of human resource professionals. *Hum. Resour. Dev. Int.* **11**, 351–366 (2008).
13. J. W. Lounsbury *et al.*, An investigation of the personality traits of scientists versus nonscientists and their relationship with career satisfaction: Relationship of personality traits and career satisfaction of scientists and nonscientists. *R&D Manage.* **42**, 47–59 (2012).
14. J. J. A. Denissen *et al.*, Uncovering the power of personality to shape income. *Psychol. Sci.* **29**, 3–13 (2018).
15. H. S. Friedman, M. L. Kern, Personality, well-being, and health. *Annu. Rev. Psychol.* **65**, 719–742 (2014).
16. D. J. Ozer, V. Benet-Martínez, Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.* **57**, 401–421 (2006).
17. B. W. Roberts, N. R. Kuncel, R. Shiner, A. Caspi, L. R. Goldberg, The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspect. Psychol. Sci.* **2**, 313–345 (2007).
18. S. Hitlin, J. A. Piliavin, Values: Reviving a dormant concept. *Annu. Rev. Sociol.* **30**, 359–393 (2004).
19. M. L. Kern *et al.*, The online social self: An open vocabulary approach to personality. *Assessment* **21**, 158–169 (2014).
20. M. Kosinski, D. Stillwell, T. Graepel, Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5802–5805 (2013).
21. G. Park *et al.*, Automatic personality assessment through social media language. *J. Personal. Soc. Psychol.* **108**, 934–952 (2015).
22. H. A. Schwartz *et al.*, Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One* **8**, e73791 (2013).
23. S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, J. C. Eichstaedt, Detecting depression and mental illness on social media: An integrative review. *Curr. Opin. Behav. Sci.* **18**, 43–49 (2017).
24. A. Llorente, M. Garcia-Herranz, M. Cebrian, E. Moro, Social media fingerprints of unemployment. *PLoS One* **10**, e0128692 (2015).
25. M. Settanni, D. Azucar, D. Marengo, Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychol. Behav. Soc. Netw.* **21**, 217–228 (2018).
26. "Cartographer jobs: Are they still relevant in the 21st century?" *GISGeography* (2018). https://gisgeography.com/cartographer-job-salary/. Accessed 27 November 2018.
27. K. Weins, "New DevOps trends: 2016 state of the cloud survey" *Rightscale Cloud Management Blog* (2016). https://web.archive.org/web/20161125060511/http://www.rightscale.com/blog/cloud-industry-insights/new-devops-trends-2016-state-cloud-survey. Accessed 11 May 2016.
28. H. De Witte, J. Pienaar, N. De Cuyper, Review of 30 years of longitudinal studies on the association between job insecurity and health and well-being: Is there causal evidence?: Review of longitudinal studies on job insecurity. *Aust. Psychol.* **51**, 18–31 (2016).
29. IBM, *Watson Personality Insights* (IBM, Armonk, NY, 2016). https://www.ibm.com/watson/services/personality-insights/. Accessed 1 October 2019.
30. IBM, *IBM Cloud Docs Documentation* (IBM, Armonk, NY, 2018). https://console.bluemix.net/docs/services/personality-insights/models.html. Accessed 1 October 2019.
31. A. P. Reynolds, G. Richards, B. de la Iglesia, V. J. Rayward-Smith, Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *J. Math. Model. Algorithms* **5**, 475–504 (2006).
32. L. J. P. van der Maaten, G. E. Hinton, Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
33. T. K. Ho, "Random decision forests" in *3rd International Conference on Document Analysis and Recognition* (IEEE, Piscataway, NJ, 1995), pp. 278–282.
34. J. H. Friedman, Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
35. T. Chen, C. Guestrin, "XGBoost: A scalable tree boosting system" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, B. Krishnapuram, Ed. (ACM Press, New York, NY, 2016), pp. 785–794.

Kern et al.

www.manaraa.com