Computers in Human Behavior 92 (2019) 428–445

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Predicting demographics, moral foundations, and human values from digital behaviours

Kyriaki Kalimeri[a,*], Mariano G. Beiró[b,1], Matteo Delfino[a], Robert Raleigh[c], Ciro Cattuto[a]

[a] Data Science Laboratory, ISI Foundation, Turin, Italy
[b] Universidad de Buenos Aires, Facultad de Ingeniería, INTECIN (CONICET), CABA, Argentina
[c] PathSight Predictive Science Inc., New York City, USA

ARTICLE INFO

Keywords:
Moral foundations
Machine learning
Smartphone data
Computational social science
Psychological profiles
Demographics

ABSTRACT

Personal electronic devices including smartphones give access to behavioural signals that can be used to learn about the characteristics and preferences of individuals. In this study, we explore the connection between demographic and psychological attributes and the digital behavioural records, for a cohort of 7633 people, closely representative of the US population with respect to gender, age, geographical distribution, education, and income. Along with the demographic data, we collected self-reported assessments on validated psychometric questionnaires for moral traits and basic human values, and combined this information with passively collected multi-modal digital data from web browsing behaviour and smartphone usage. A machine learning framework was then designed to infer both the demographic and psychological attributes from the behavioural data. In a cross-validated setting, our models predicted demographic attributes with good accuracy as measured by the weighted AUROC score (Area Under the Receiver Operating Characteristic), but were less performant for the moral traits and human values. These results call for further investigation, since they are still far from unveiling individuals' psychological fabric. This connection, along with the most predictive features that we provide for each attribute, might prove useful for designing personalised services, communication strategies, and interventions, and can be used to sketch a portrait of people with similar worldview.

1. Introduction

Strategic demographic attributes, namely gender and age, have been traditionally employed to gain insights into the population; today scientists, policymakers, and practitioners from diverse disciplines are paying an ever-growing attention to digital data as they offer an alternative, complementary view of the society. Computational social science and digital humanities fields emerge exactly from this interdisciplinary interest to address societal questions timely and in greater scale. The research questions asked in these fields are complex; cultural, psychological, and other human factors are entangled within the social phenomena under investigation.

Seeing society through the lens of the digital data provides a unique opportunity to capture accurately the complexity of the human psychometric attributes, shedding light to deeper, socio-culturally relevant descriptors, which yield a more useful understanding of the human behaviour. Digital data act efficiently as a proxy to demographic attributes with important applications ranging from the gender gap tracking (Fatehkia, Kashyap, & Weber, 2018) to monitoring (Zagheni, Weber, & Gummadi, 2017) and assimilation of migrants (Dubois, Zagheni, Garimella, & Weber, 2018), to unemployment (Bonanomi, Rosina, Cattuto, & Kalimeri, 2017; Burke & Kraut, 2013; Llorente, Garcia-Herranz, Cebrian, & Moro, 2015), and health monitoring (Ginsberg et al., 2009). The need for understanding human factors via digital data sparkled an ever-growing interest in automatic recognition of personality traits due to its association with important life aspects, including "happiness, physical and psychological health, occupational choice, satisfaction, and performance, community involvement, criminal activity, and political ideology" (Ozer & Benet-Martinez, 2006). Over the last years, researchers proposed a series of computational personality recognition methods based on a varied set of digital data (see Vinciarelli & Mohammadi, 2014; Finnerty, Lepri, & Pianesi, 2016; Farnadi et al., 2016 for a review) demonstrating the feasibility of the approach.

In this study, we aim to advance an integrative view of the person, attempting to infer the moral traits as described by the Moral Foundations Theory (MFT) (Haidt, 2003) and personal values by the

* Corresponding author.
E-mail address: kalimeri@ieee.org (K. Kalimeri).
[1] Work done when in Data Science Laboratory, ISI Foundation, Turin, Italy.

https://doi.org/10.1016/j.chb.2018.11.024
Received 5 December 2017; Received in revised form 29 October 2018; Accepted 12 November 2018
Available online 21 November 2018
0747-5632/

Schwartz theory of human values (SHV) (Schwartz, 1992). Such psychological constructs are associated with attitudes and behaviours in complex situations like politics (Miles & Vaisey, 2015), charitable donations (Hoover, Johnson, Boghrati, Graham, & Dehghani, 2018; Winterich, Zhang, & Mittal, 2012), climate change (Wolsko, Ariceaga, & Seiden, 2016), poverty (Low & Wui, 2015), vaccination (Amin et al., 2017), or even violent protests (Mooijman, Hoover, Lin, Ji, & Dehghani, 2018); where personality dispositions alone do not suffice to explain our judgements. But how much do shallow digital records reveal us? Which data sources are the most informative? What ethical implications might emerge?

Attempting to tackle these questions, we present a large-scale study consisting of spontaneous, "in-the-wild", observations of digital behaviours. A cohort of 7633 participants, nearly representative to the US census with respect to several demographic variables, was recruited. Our participants allowed access to their digital data - either desktop web browsing or smartphone data - for one month. Often in computational social science, large-scale studies are based only on observational data; here, we obtained from all participants self-assessments on a series of validated psychometric questionnaires regarding their moral views, human values, and demographic information. We employed the individual digital traces to train classification models for predicting moral traits, human values and demographic attributes. The reported results are cross-validated on the information provided directly by the participants.

Such experimental design allows us to provide a clear view of the attributes, both demographic and psychological, that are easier to predict from proxy digital data while shedding light to the ones that instead remain still challenging. At the same time, we present a comparative study on the predictive power of each digital data source and modality (web browsing, mobile browsing, application usage) as well as their combination. The quantitative comparison of performance among different modalities demonstrates the informative power of the mobile web browsing, whose role is essential in bridging online and offline worlds (Golder & Macy, 2014a).

Additionally to the quantitative performance of the predictive models for demographic and psychometric attributes, we take a deeper look into the models, and in particular at the top predictors per attribute, providing a qualitative analysis of the behavioural proxies associated with each attribute. We provide interesting insights on the individuals, sketching cyber-cultural profiles based on their interests, cultural elements, and habitual actions, essential for an in-depth understanding of the perception and attitudes of people towards important social issues.

We tackle discrimination and privacy-preserving issues, pointing out the exact digital proxies employed to differentiate between two individuals both for the demographic and psychological attributes. Even though further investigation is required in this domain, we highlight how well attributes can be inferred and by which digital proxies, contributing to the transparency of the predictions and raising awareness around ethical implications (Chen, Fraiberger, Moakler, & Provost, 2017).

Undoubtedly, disclosing the connection between digital human behaviour and psychometrics gives not only the possibility of endowing machines with a notion of social intelligence but also paves the way to a fairer personalised user experience. With the important penetration of smartphones in society, the effects of automatically predicting demographics and moral values at scale may help to better and timely understand attitudes and judgements towards current or upcoming issues, preventing phenomena like gender or ethnicity discrimination, while improving the efficiency of communication campaigns in the digital humanities field.

2. Literature review on psychometric and demographic attribute inference

Researchers across various disciplines including sociology, demography and public health have always been keen on examining how society functions observing populations at scale. Here, we present the most influential studies that tackled the issue of demographic and psychometric attribute prediction employing data sources similar to ours.

2.1. Inferring demographics

There is a huge body of literature regarding demographic attribute prediction from digital data. Early studies performed by means of traditional interviews revealed differences in the way men and women use the Internet, showing for example that women were more prone to use email while males preferred web browsing (Jackson, Ervin, Gardner, & Schmitt, 2001). Attributes like gender and age of Internet users can be predicted from their Web browsing behaviours. Among the most influential studies on the topic, Hu, Zeng, Li, Niu, and Chen (2007) treated web browsing data as a hidden variable propagating demographic information between different users, Weber and Castillo (2010) employed a Yahoo! query database to infer gender, income, and ethnicity using web browsing query data from a closely representative sample of the US population, while Weber and Jaimes (2011) analysed the queries submitted from different ZIP codes enriched by US census data and exploited them to highlight differences in user behaviour and search patterns among several demographic groups.

Gender and age can also be inferred using call detail records from smartphone devices over large populations (Dong, Yang, Tang, Yang, & Chawla, 2014; Felbo, Sundsoy, Pentland, Lehmann, & de Montjoye, 2015; Ying, Chang, Huang, & Tseng, 2012), exploiting features as the number of unique contacts, the number of calls, text messages, and the total duration of calls for incoming and outgoing interactions. Moreover, several demographic attributes can be predicted from the installed applications on smartphones (Malmi & Weber, 2016). Seneviratne, Seneviratne, Mohapatra, and Mahanti (2015) predicted the user's gender based only on the list of apps installed by 218 users. The automatic prediction of demographic characteristics allows for a deeper understanding of online communication and behavioural patterns, and has various applications: in particular, understanding gender differences when interacting with sponsored search results (Jansen & Solomon, 2010) can help build better recommendation systems for e-commerce websites, improving user experience and increasing sales (Wang, Guo, Lan, Xu, & Cheng, 2016; Zhao et al., 2014).

Social media are increasingly used for demographic attribute prediction (see Cesare, Grant, & Nsoesie, 2017 for a review). Recently, Zhong, Yuan, Zhong, Zhang, and Xie (2015) investigated the predictive power of location check-ins for inferring users' demographics on the Chinese social network Sina Weibo. Mislove, Lehmann, Ahn, Onnela, and Rosenquist (2011) assessed the demographics of Twitter users in terms of gender and ethnicity, while others predicted the demographics of users from their Facebook data (Bi, Shokouhi, Kosinski, & Graepel, 2013; Kosinski, Stillwell, & Graepel, 2013; Youyou, Kosinski, & Stillwell, 2015).

2.2. Inferring personality traits

Over the last years, researchers attempted to infer personality traits from diverse data sources including text, video, audio, mobile and wearable devices, social media, web queries and others. The most dominant paradigm in the computational social science community is the Big-Five (BF) or Five-Factor Model (FFM) (Costa & McCrae, 1992). Based on this model, many recent studies exploited the relationship between automatically extracted behavioural characteristics derived from digital data and the Big-Five personality traits (see Vinciarelli & Mohammadi, 2014; Finnerty et al., 2016 for literature reviews). Butt and Phillips (2008); Chittaranjan, Blom, and Gatica-Perez (2013); de Montjoye, Quoidbach, Robic, and Pentland (2013) assessed the predictability of personal traits via calls and SMS data from smartphone usage while Xu, Frey, Fleisch, and Ilic (2016) was based on behavioural patterns of smartphone application adoption. Lepri et al. (2012) and Kalimeri (2013) explored the interplay between behavioural traits and

the influence of the social context while Kalimeri, Lepri, and Pianesi (2013); Rauthmann, Sherman, Nave, and Funder (2015) assessed a dynamic view of the personality traits, focusing on the ever-changing nature of actual behaviour/personality states and the elusive role of situational factors from mobile sensor data.

More recently, a plethora of studies employed social media data for the prediction of personality traits from the massive popularity of these sites. The Twitter platform has been one of the first platforms employed for this task (indicatively Golbeck, Robles, Edmondson, & Turner, 2011; Quercia, Kosinski, Stillwell, & Crowcroft, 2011) and remained popular ever since. Kosinski, Stillwell, Kohli, Bachrach, and Graepel (2012) studied the role of personality in website preferences, combining personality profiles and website choices from more than 160,000 users and investigating whether different websites attract audiences with different personalities. Facebook Likes have also been employed to predict the personality traits of the users by means of regression models (Kosinski et al., 2013; Youyou et al., 2015). In both of these works, Openness and Extraversion turn out to be the most predictable traits. Farnadi et al. (2016) presented a comparative analysis of state-of-the-art computational personality recognition methods on a varied set of social media ground-truth data from Facebook, Twitter and YouTube, while fusion of different social media data sources is also proven to be informative (Gou, Zhou, & Yang, 2014; Skowron, Tkalčič, Ferwerda, & Schedl, 2016).

Personality trait prediction is out of the scope of this study, however, we briefly presented the most critical studies on this topic for a fruitful discussion of our findings later on, since the obtained results prove the feasibility of the approaches but are still low regardless of the digital data employed and the scientific maturity on the topic.

2.3. Inferring moral traits

Here, we place the focal point on the prediction of moral traits which is an open issue for the research community. With respect to the dispositional traits of personality, moral traits are considered to be a higher level construct (McAdams, 1995; McAdams & Pals, 2006) clarifying how and when dispositions and attitudes towards interpersonal and intergroup processes relate with persuasion and communication narratives. Studies have pointed out the connection between decision making and moral judgements (Weber, Eden, Huskey, Mangus, & Falk, 2015), essential to understand the perception process of media communication.

In this study, we employ the Moral Foundation Theory (MFT) (Haidt, 2003), which proposes a set of moral principles that could be considered at a higher level with respect to the dispositional traits of personality expressed in the Five-Factor model (Costa & McCrae, 1992). It provides insights on the characteristic adaptations of the individuals (Haidt, Graham, & Joseph, 2009) as described by Dan McAdams's three-level account of personality (dispositional traits, characteristic adaptations, and life stories) (McAdams, 1995; McAdams & Pals, 2006). Focusing on the psychological basis of morality, it identifies the following five moral foundations (see Haidt & Graham, 2007; Haidt & Joseph, 2004); *care/harm*, *fairness/cheating*, *loyalty/betrayal*, *authority/subversion*, and *purity/degradation*, These foundations collapse into two superior foundations, namely the (1) *individualising* and (2) *binding* foundations.

Even if in its infancy, MFT has already impacted not only psychology but various scientific domains (see Graham et al., 2012 for a review). In the computational social science field, Moral Foundation Theory (MFT) (Haidt, 2003) has been studied only recently. Kim, Iyer, Graham, Chang, and Maheswaran (2013) investigated whether digital traces gathered from game play data can reveal fundamental aspects of a person's sacred values or moral identity pointing out several associations. Lin, Hoover, Dehghani, Mooijman, and Ji (2017) assessed the predictability of moral traits from linguistic features using a set of manually annotated texts from Twitter. Their task did not consist on predicting a high or low value related to the orientation of each trait (e.g., care vs. harm), but only its presence in the tweet. They reported satisfying results for care, fairness, and authority (average F-score of 74%), while loyalty and purity were less performant (average F-score of 44%). By comparing the performance of their models against that of a human annotator during the testing phase, they demonstrate the difficulty of the task also for humans.

It is expected that moral traits are expressed more clearly in verbal rather than non-verbal manner, for instance, direct answers to blog discussions, or personal opinions on the Twitter platform (Hoover et al., 2018; Mooijman et al., 2018), rather than the fact that the individual visited a certain blog. The value of our approach is that we try to set a baseline of prediction based on readily available data, which any ISP, mobile carrier, or even mobile application has access to, without being tied to a specific platform such as Facebook or Twitter, while addressing potential implications emerging when making inference on such data. To the best of our knowledge, the present study is one of the first attempts to predict the moral traits from web and smartphone data.

2.4. Inferring human values

The moral domain is also influenced by virtue ethics, education, religion and the ethical codes of specific cultures (Haidt & Joseph, 2004). The Schwartz Theory of Human Values (SHV) (Schwartz, 1992) provides us with a rigorous framework which identifies the following ten basic human values: *self-direction*, *stimulation*, *hedonism*, *achievement*, *power*, *security*, *conformity*, *tradition*, *benevolence*, *universalism*. The above ten values can be clustered into four higher order value types, so-called quadrant values and into two dimensions: *Openness to change* (self-direction, stimulation) vs. *Conservation* (security, conformity, tradition) and *Self-enhancement* (universalism, benevolence) vs. *Self-transcendence* (power, achievement). Therefore, the first dimension captures the conflict between values that emphasise the independence of thought, action, and feelings and readiness for change and the values that emphasise order, self-restriction, preservation of the past, and resistance to change. The second dimension captures the conflict between values that emphasise concern for the welfare and interests of others and values that emphasise the pursuit of one's own interests and relative success and dominance over others. Hedonism shares elements of both openness to change and self-enhancement. According to Schwartz (1992), values are “desirable, trans-situational goals, varying in importance that serve as guiding principles in people's lives”. These values are shown to have cross-cultural validity (Schwartz, 2005) and have been found to correlate to several behaviours (Bardi & Schwartz, 2003; Schwartz, 2003), including among others consumer decisions (Grunert & Juhl, 1995), pro-environmental behaviour (Soyez, 2012) as well as judgements (Torelli & Kaikati, 2009) while there is a demonstrated association with moral traits of the MFT (Feldman, 2018; Parks-Leduc, Feldman, & Bardi, 2015). In the computational social science domain, there have been a few attempts to predict human values from digital data. Chen, Hsieh, Mahmud, and Nichols (2014) presented an analysis of associations between human values and word use in online social media. Youyou et al. (2015) assessed –among other attributes– the prediction of Schwartz human values from Facebook Likes and only reported these results in the Supplementary Information of their study. Both studies reported low prediction scores.

3. Methods

3.1. Data collection

7633 subjects participated in this study after being informed about its content, purpose, and privacy policies, during a probability based recruitment campaign carried out in the United States of America, conducted by a subcontracted marketing company. Informed consent was obtained from all participants enabling the collection, storage, and

Table 1
Complete list of the demographic attributes collected and their respective ranges for the entire sample (7633 participants) that took part in our study. Each option corresponds to a distinct "label" for the classification algorithm. For instance, for the classification of Political Party our classifier had to choose between four labels, "democrats", "republicans", "libertarians" or "independents".

Attribute	Demographic Variables Range	Sample size (N = 7633)	Attribute	Demographic Variables Range	Sample size (N = 7633)
Age	18–24	262 (3.4%)	**Political Party**	Democrat	2973 (38.9%)
	25–34	1308 (17.1%)		Republican	2556 (33.5%)
	35–49	2009 (26.3%)		Libertarian	215 (2.8%)
	50–54	879 (11.5%)		Independent	1889 (24.7%)
	55–64	1759 (23%)			
	65 Plus	1416 (18.6%)			
Education	College Graduate	2624 (34.4%)	**Wealth**	50KLess	2246 (29.4%)
	Post Graduate	2211 (29%)		50K-100K	915 (12%)
	Some College	1680 (22%)		100K-250K	1213 (15.9%)
	High-school	619 (8.1%)		250K-500K	1303 (17.1%)
	Trade School	479 (6.3%)		500K-1000K	1067 (14%)
				1000KPlus	889 (11.6%)
Ethnicity	Asian	338 (4.4%)	**Weight Issues**	No	4239 (55.5%)
	African American	485 (6.3%)		Yes	3394 (44.5%)
	White	6359 (83.3%)			
	Hispanic	347 (4.5%)			
Exercise	No	3328 (43.6%)	**Parent**	No	2533 (33.2%)
	Yes	4305 (56.4%)		Yes	5100 (66.8%)
Gender	Female	4367 (57.2%)	**Smoker**	No	7027 (92.1%)
	Male	3266 (42.8%)		Yes	606 (7.9%)
Income	20KLess	342 (4.5%)	**Marital Status**	Divorced	733 (9.6%)
	20K-30K	492 (6.4%)		Single	1281 (16.8%)
	30K-50K	1222 (16%)		Married	4690 (61.4%)
	50K-75K	1613 (21.1%)		Living Together	617 (8.1%)
	75K-100K	1520 (19.9%)			
	100K-150K	1581 (20.7%)			
	150K-200K	515 (6.7%)			
	200KPlus	348 (4.6%)			

treatment of data. Upon acceptance of the privacy policy[2], the participants were asked to fill in a series of demographic information and validated psychometric questionnaires. Moreover, they were requested to allow access to either their basic mobile or desktop traffic data, for a period of one month. 5008 people (2823 women) consented the access to their desktop data while 2625 people (1544 women) consented access to their mobile data, hereafter referred to as "Desktop" and "Mobile" datasets respectively. All were incentivised to participate in the study.

3.1.1. Demographic data

The intake survey covered basic demographic factors (age, gender, ethnicity[3]), geographic factors (home location, expressed at zip code level), socioeconomic factors (educational level, marital status, parenthood, wealth, income), health-related factors (exercise, smoke, and weight issues) and political orientation. Table 1 presents the complete list of the demographic information gathered, along with the respective range of values for all the 7633 participants. Our predictive models were trained to predict each of the distinct values for each demographic attribute collected and reported in Table 1; for instance, for the classification of Political Party our classifier had to choose between four labels, "democrats", "republicans", "libertarians" or "independent".

[2] The entire research policy can be found here: https://www.researchnow.com/privacy-policy/.

[3] We use the term "African American" as a shorthand for the term "black or African American" as used in the official US census. See http://factfinder.census.gov/help/en/.

3.1.2. Representativeness of the recruited sample

The corpus of this study closely follows the American census data not only in terms of gender distribution but also with respect to age groups, education, income and geographical distribution. We compared our sample distribution (7633 participants) to the official American census data (Bureau, 2015) provided by the US Census Bureau for the respective year of the study (2015) and containing information regarding age groups distribution by gender (Table 2), education, ethnicity, and income (Table 3), and geographical distribution (Fig. 1). We also report breakdown comparisons between the entire sample, the "Mobile" (2625 subjects) and the "Desktop" dataset (5008 subjects): Fig. 1(a) depicts the geographical distribution of the recruited sample (observed values) in the entire corpus as compared to the American Census (expected values). Fig. 1(b) and 1(c) present insights on the geographical distributions for the two subsets, "Mobile" and "Desktop". Finally, Table 4 compares the age, gender, education, ethnicity and income distribution between both subsets. Importantly, all the above attributes of the entire sample employed in this study, as well as the two subsets "Mobile" and "Desktop", are approximating the expected distributions according to the US census.

3.2. Psychometric measures

3.2.1. Moral foundations

The five moral foundations from the Moral Foundation Theory (Haidt & Graham, 2007; Haidt & Joseph, 2004) were assessed through the Moral Foundations Questionnaire (Graham et al., 2011), a validated measure of the degree to which individuals endorse each of five

Table 2
Age and Gender Representativeness. The distribution of the participants according to their age group and gender shows that the sample represents closely the demographics of the US.

Age distribution by gender	US census	Recruited sample (N = 7633)
Entire sample		
18 to 24	13.0%	3.4%
25 to 34	17.6%	17.1%
35 to 49	24.4%	26.3%
50 to 54	9.2%	11.6%
55 to 64	16.7%	23.0%
≥ 65	19.2%	18.6%
Male subset		
18 to 24	13.6%	2.1%
25 to 34	18.2%	12.0%
35 to 49	24.9%	25.4%
50 to 54	9.2%	11.2%
55 to 64	16.6%	25.1%
≥ 65	17.5%	24.1%
Female subset		
18 to 24	12.3%	4.4%
25 to 34	17.0%	21.0%
35 to 49	23.9%	27.0%
50 to 54	9.1%	11.7%
55 to 64	16.8%	21.5%
≥ 65	20.8%	14.4%

Table 3
Comparative statistics for the demographic attributes, education, ethnicity, and income of the entire dataset employed for this study and the US census. US census data are provided by US Census Bureau (2011–2015 American Community Survey 5-Year Estimates (⋆)). Education refers to the percent of persons age 25 years + 2011–2015 (⋆⋆) and income is expressed in 2015 US dollars (⋆⋆⋆).

Demographic Variables	US census (⋆)	Recruited sample (N = 7633)
Education		
High-school graduate or higher (⋆⋆)	86.7%	93%
Bachelor's degree or higher (⋆⋆)	29.8%	28%
Ethnicity		
Asian	5.7%	4.4%
African American	13.3%	6.3%
White	76.9%	83.3%
Hispanic	17.8%	4.5%
Income		
Median income (⋆⋆⋆)	$53K	$50K - $75K

dimensions:

- *care/harm*, basic concerns for the suffering of others, including virtues of caring and compassion (Cronbach's alpha = 0.76);
- *fairness/cheating*, concerns about unfair treatment, inequality, and more abstract notions of justice (Cronbach's alpha = 0.74);
- *loyalty/betrayal*, concerns related to obligations of group membership, such as loyalty, self-sacrifice, and vigilance against betrayal (Cronbach's alpha = 0.73);
- *authority/subversion*, concerns related to social order and the obligations of hierarchical relationships like obedience, respect, and proper role fulfilment (Cronbach's alpha = 0.73);
- *purity/degradation*, concerns about physical and spiritual contagion, including virtues of chastity, wholesomeness, and control of desires (Cronbach's alpha = 0.82).

The questionnaire is based on self-assessment evaluations and consists of 30 items, resulting in a unique numerical value from 0 to 30 per person. According to the MFQ, six items (on a 6-point Likert scale) per foundation were averaged to produce the individuals' scores on each of the five foundations. All the obtained scores were quantised into two classes, “Low” and “High”, with the threshold placed at the respective median value per foundation. We opted for a binary strategy for all the numeric variables, which is a common practice in the computational social science field to limit phenomena of extremely unbalanced classes.

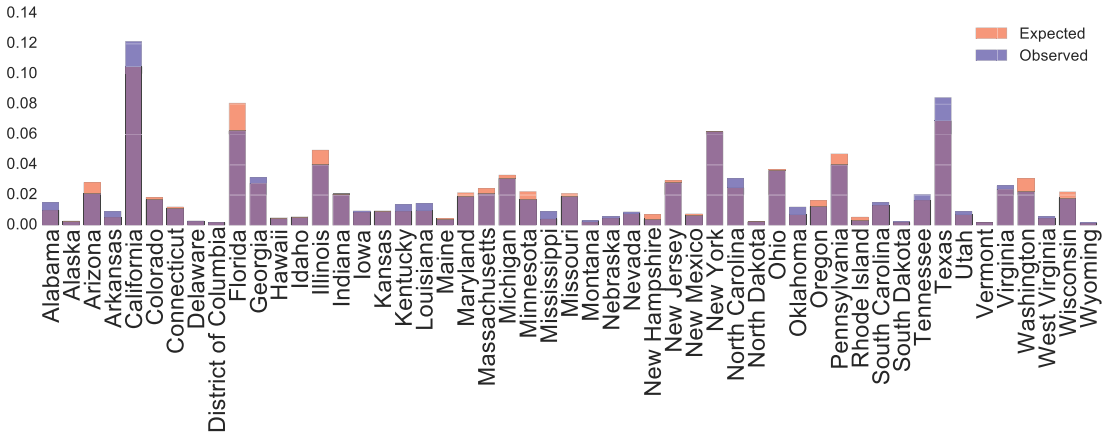
Additionally, we assessed the two superior foundations into which those five collapse according to Haidt and Graham (2007): *individualising* (care and fairness) and *binding* (loyalty, authority and purity). The individualising foundation asserts that the basic constructs of society are the individuals and hence focuses on their protection and fair treatment, defending their right to pursue their own goals. In contrast, the binding foundation focuses on group-binding, based on the respect of leadership and traditions, and the defence of the family as the nucleus of society (Haidt & Graham, 2007).

3.2.2. Schwartz basic human values

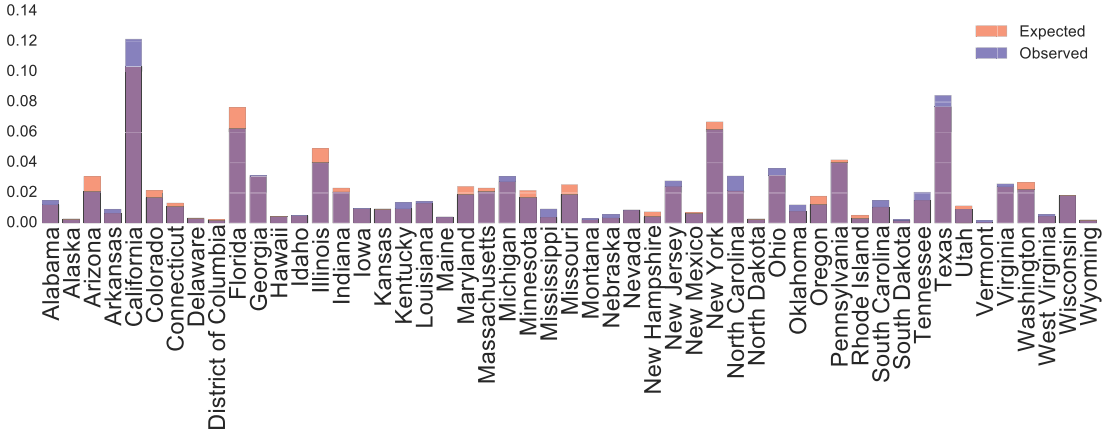
The *Schwartz human values* were assessed by the Portrait Values Questionnaire (Schwartz, 2012), whose validity across cultures is validated in studies performed on 82 countries and samples belonging to highly diverse geographic, cultural, linguistic, religious, age, gender, and occupational groups. The questionnaire is based on self-assessments resulting in a numerical value per person for each of the ten basic values:

- *self-direction*, independent thought, action-choosing, creating, exploring (Cronbach's alpha = 0.67);
- *stimulation*, need for variety and stimulation to maintain an optimal level of activation (Cronbach's alpha = 0.85);
- *hedonism*, related to organismic needs and the pleasure associated with satisfying them (Cronbach's alpha = 0.83);
- *achievement*, personal success through demonstrating competence according to social standards (Cronbach's alpha = 0.86);
- *power*, the attainment or preservation of a dominant position within the more general social system (Cronbach's alpha = 0.62);
- *security*, safety, harmony, and stability of society, of relationships, and of self (Cronbach's alpha = 0.73);
- *conformity*, restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms (Cronbach's alpha = 0.83);
- *tradition*, symbols and practices or groups that represent their shared experience and fate (Cronbach's alpha = 0.60);
- *benevolence*, concern for the welfare of close others in everyday interaction (Cronbach's alpha = 0.81);
- *universalism*, this value type includes the former maturity value type; including understanding, appreciation, tolerance, and protection for the welfare of all people and for nature (Cronbach's alpha = 0.80).

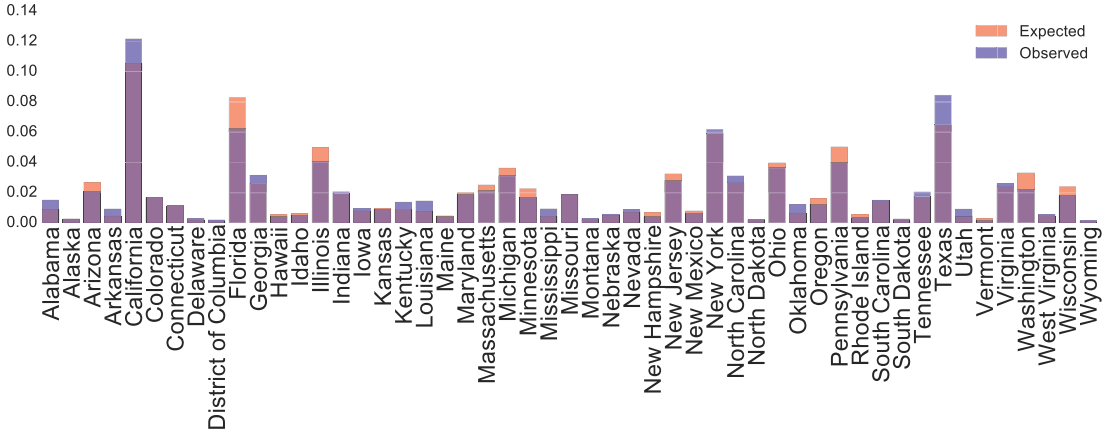
The questionnaire is based on self-assessment evaluations on a 7-point Likert scale. Following Schwartz (2012), we averaged the respective items per value and we accounted for individual differences. The above ten values can be clustered into four higher order values, so-called quadrant values and into two dimensions, as the sum of the individual items of which they consist: *Openness to change* (self-direction, stimulation) vs. *Conservation* (security, conformity, tradition) and *Self-enhancement* (universalism, benevolence) vs. *Self-transcendence* (power, achievement). Therefore, the first dimension captures the conflict between values that emphasise the independence of thought, action, and feelings and readiness for change and the values that highlight order, self-restriction, preservation of the past, and resistance to change. The second dimension captures the conflict between values that stress concern for the welfare and interests of others and values that emphasise the pursuit of one's own interests and relative success and



(a) Participants Geography



(b) Mobile Participants Geography



(c) Desktop Participants Geography

Fig. 1. Distribution of the geographical distribution of the recruited sample (observed values) in the entire dataset (expected values) as compared to the American census (Fig. 1a). The geographical distribution of the mobile (Fig. 1b) and the Desktop (Fig. 1c) datasets follow closely the geographical distribution of participants.

dominance over others. Hedonism shares elements of both openness to change and self-enhancement. The obtained scores were quantised into two classes, “Low” and “High”, with the threshold placed at their respective median value as for the MFQ's values.

3.3. Digital behaviours

Additionally to the demographic and psychometric questionnaires, our participants were asked to provide us with their digital data for one month; either desktop or mobile and in both cases they were financially incentivised to do so. As mentioned before, out of the 7633 participants, 5008 permitted us to assess their desktop web browsing data and 2625 to their mobile data. In the following paragraphs, we describe in detail the information contained in these data.

3.3.1. Desktop data

For the participants who permitted the logging of their desktops' web browsing data, 5008 in total, we captured: (i) the domain names, and (ii) the average time spent online and (iii) the number of visits per day on each domain. All this information was aggregated by day, and

Table 4
Comparison of age, gender, education, ethnicity and income distribution between the Mobile and Desktop subsets.

Demographic Variables	Mobile subset	Desktop subset
Age		
18 to 24	5.7%	2.3%
25 to 34	27.7%	11.6%
35 to 49	35.2%	21.7%
50 to 54	10.6%	12.0%
55 to 64	14.5%	27.5%
≥ 65	6.4%	24.9%
Gender		
Female	58.8%	56.4%
Male	41.4%	43.6%
Education		
College graduate	36.2%	33.4%
Post graduate	26.8%	30.1%
Some college	23.4%	21.3%
High-school	6.7%	8.8%
Trade school	6.5%	6.2%
Less than high-school	0.3%	0.2%
Ethnicity		
Asian	4.6%	4.4%
African american	8.5%	5.2%
White	78.5%	85.8%
Hispanic	6.6%	3.5%
Other	1.8%	1.1%
Income		
200K or more	3.9%	4.9%
150K–200K	6.9%	6.7%
100K–150K	20.8%	20.6%
75K–100K	18.7%	20.6%
50K–75K	21.5%	20.9%
30K–50K	16.5%	15.7%
20K–30K	7.0%	6.1%
Less than 20K	4.6%	4.4%

only the domain names (and not the page or section of the websites) were stored, to ensure the privacy of the participants. Having performed an initial exploratory analysis (see *Effects of Quantity and Quality of Available Information* in Section 4), we set the minimum of amount of web domain visits required per subject to $N = 30$; users with a lower number of visits than this, were discarded from the analysis leaving us with a total of 4781 participants.

3.3.2. Mobile data

Participants were asked to download an application which, upon agreement of the privacy policy, logged their web browsing activity and application usage.

- *Applications' Data.* Application usage was captured whenever the application was running in the "foreground". Foreground usage means an application is open on someone's device, regardless of whether the application is currently being engaged with or not. Application usage data included for each participant, records of the date and timestamp, the local time zone, and time spent on the application (in seconds).
- *Mobile browsing Data.* URL data was captured from the native browser on the subject's device (not any 3rd party browsers). URLs for both secure and non-secure traffic were captured, though only the URLs' domain was stored for privacy issues. Similar to the desktop browsing data, users with a number of visits less than $N = 30$ unique domains were discarded from the analysis leaving us with a total of 2406 participants. The storage of mobile web browsing data followed the same criterion as for desktop web browsing: only the domain names of the websites were stored to preserve the privacy of the subjects.

Noteworthy is the fact that "Mobile" and "Desktop" browsing data provide the same information they only express different modes of web navigation i.e. mobile vs desktop.

3.4. Classification method: random forest

To automatically identify the morality and human values of the participants having access only to their digital data, we conceived this study as a supervised classification problem. The digital behavioural traces obtained for each participant were used to predict their individual measures while the performance of the prediction was assessed based on the ground-truth provided by their responses in the questionnaires, and their demographic attributes. Ground-truth levels for the moral traits and human values were obtained by considering the self-assessed questionnaires filled by the subjects without any additional labelling by external judges. All psychometric measures were thresholded at the median point and treated as binary variables, namely *Low, High*, in a single-class classification scenario. Instead, for the demographic attributes, we followed a multi-class classification scheme with a one-against-all strategy (Bishop, 2006) whenever the range of possible values was more than two, for instance, the "age" attribute, and a single-class approach (similar to the psychometric one) when only two values were possible, for instance the "gender" attribute; Table 1 reports all the possible classes per demographic attribute.

We designed a generic experimentation scheme, aiming to assess and compare the predictive power, i.e. the ability to correctly anticipate unseen data, of (i) the distinct behavioural sequences alone (desktop web browsing, mobile web browsing, mobile application usage) and (ii) the fusion of the separate mobile modalities (web browsing and apps for the "Mobile" dataset). Our data fusion policy consisted of "early" fusion at a feature level; for each mobile user we aggregated the two separate data vectors, the one expressing the web browsing activity and the one expressing the mobile application usage. We opted for a Random Forest (RF) classifier (Breiman, 2001) due to its ability to deal with sparse data (web browsing activity) and unbalanced labels (see Table 1). The input to the RF classifier is a matrix whose rows (*feature vectors*, hereafter) correspond to the activity of each participant, its columns to the visited domains or apps by all participants, while the values represent the frequency of visits to the specific domain, or app by the specific participant, if any. Then, the classifier trained on the activity of a subset of subjects, predicts for the new unseen participants, based on their activity, the probability that they are closer to one or the other class label or else the target value (i.e., a specific attribute like "gender") as a continuous value between 0.0 (Low) and 1.0 (High).

In this study, we fitted a model for each target employing a five-fold cross-validation procedure in which 4-folds are used for model training, and the remaining one is used to evaluate the predictions, in a round-robin scheme. The aim of the cross-validation step is to use mutually exclusive subjects for the training (i.e., fitting) of the model and avoiding the common issue of overfitting during its consequent evaluation. The predictions were evaluated in terms of the weighted AUROC statistic (see Subsection 3.5 for a definition of AUROC). Inside each training stage, the hyper-parameter optimisation was performed using a grid-search procedure; the hyper-parameter combinations explored are reported in Table 5. For each hyper-parameter combination, a random forest model was trained by using an inner five-fold cross-validation inside the training set, and its performance was evaluated against the test set. Finally, we report the average AUROC performance over all folds. In this way, a data sample is used in the validation phase only once and was never seen during the training phase of the model.

As a final step, we estimated the most predictive behaviours (i.e., features) for each target variable by computing the relative rank (i.e., depth) of the features in the different trees that conform the random

Table 5
Random Forest hyper-parameter space and respective set of attributes exploited for each model in our experimental scenarios.

Hyperparameter	Value Range
Number of trees	150, 300, 600
Maximum number of features (N)	$\{\frac{1}{2}, 1, 2\} * \sqrt{N}$
Maximum tree depth	5, 7, 15
Minimum number of samples in leaf	5
Class weight function	'entropy', 'gini'

forest classifiers, according to the *Gini impurity function* (Breiman, 2001).[4] The higher the rank of a feature, the higher is its importance for the prediction of the specific target variable; hence the relation of the specific behaviour to the specific target variable.

3.5. Accuracy measurement: the weighted AUROC

Model performance was computed in terms of the *weighted Area Under the Receiver Operating Characteristic (AUROC)* statistic (Fawcett, 2006; Hanley & McNeil, 1982; Li & Fine, 2010). The AUROC is a performance measure for binary classifiers that uses a discrimination threshold to distinguish between a High and a Low class. Using a discrimination threshold implies that the classifier produces a continuous rank value for each sample, and the samples whose rank is above the threshold are classified as High, while the remaining ones are classified as Low. The AUROC was preferred over the commonly-used *accuracy* metric (i.e., the proportion of true positives and true negatives among the total number of samples) as it takes into account the effect of unbalanced labels, which holds true for most of our demographic attributes (Mason & Graham, 2002). In the case of highly unbalanced classes, even if a classifier is unable to discriminate between High and Low, obtaining accuracies close to 1.0 is easier since they are easily achieved simply by predicting always the most common class label, which can lead to misleading results and interpretations. Instead, in a similar scenario, the AUROC metric would give a result of 0.5 since taking into account the true positive (TPR) and false positive (FPR) rates, each class would be inversely weighted with its relative frequency.

The AUROC resumes into a single quantity the trade-off between the false positive rate (FPR) or *Type I* error and the true positive rate (TPR) or sensitivity, by measuring the variation of the TPR as a function of the FPR as the classifier's discrimination threshold moves from its minimum to its maximum. The curve followed by this function is called ROC curve, and the AUROC is the area under this curve, bounded between 0 and 1. The diagonal line between points (0,0) and (1,1) is called *no-discrimination line*, as a random classification algorithm (i.e., one that chooses between the High and Low-classes with equal probability) would move over it, and its expected AUROC value would be 0.5.[5] The AUROC score for a perfect classifier is 1, and any AUROC value above 0.5 is considered *better than random*.

We remark that the AUROC's definition is equivalent to the probability that a randomly-chosen sample among the High-class ones is given a higher rank than a randomly-chosen sample among the Low-class ones (Hanley & McNeil, 1982). For the multi-class classifications (for instance, income, education, wealth, etc.), we report the weighted AUROC statistic, which is the average of the AUROC's of single-class classifiers (one-against-all approach), weighted by the prevalence of each class (Fawcett, 2006; Li & Fine, 2010).

[4] The implementation of the model was performed using code from the scikit-learn project (Pedregosa et al., 2011).

[5] We conventionally refer to the AUROC values as percentages throughout this paper, using the notation 50% instead of 0.5.

4. Effects of quantity and quality of Available Information

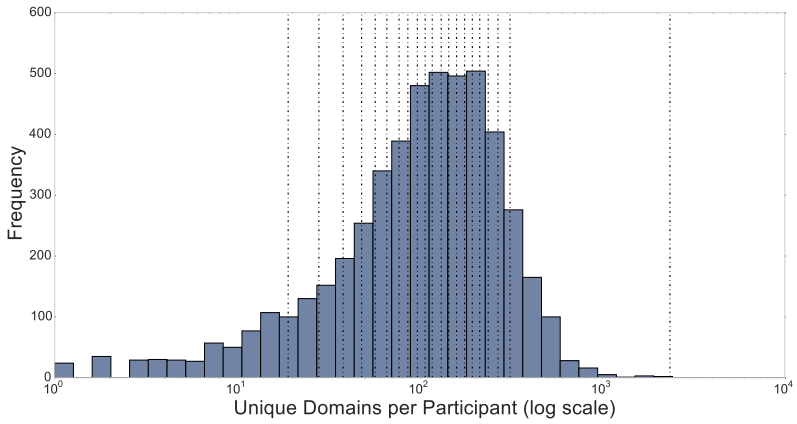
A common strategy followed in classification tasks with sparse data is to rank the feature vectors according to their total frequency (i.e., the columns containing the most frequently used web domains and/or apps are placed first), improving the learning of the model and hence its prediction accuracy. In this way, the model learns a "forced" reality, requiring at the same time access to large datasets of long-term observations. In cases where access to large datasets is not possible or when the training needs to be performed on a thin-slice of observations, the prediction of the model often drops. We present here an exploratory analysis, focusing only on the "gender" prediction inferring from the desktop web browsing data. In the following paragraphs, we aim to compare the predictive power of the Random Forest classification models evidencing the biases due to the quantity or the quality of information available each time.

4.1. How much digital information is needed for a successful prediction?

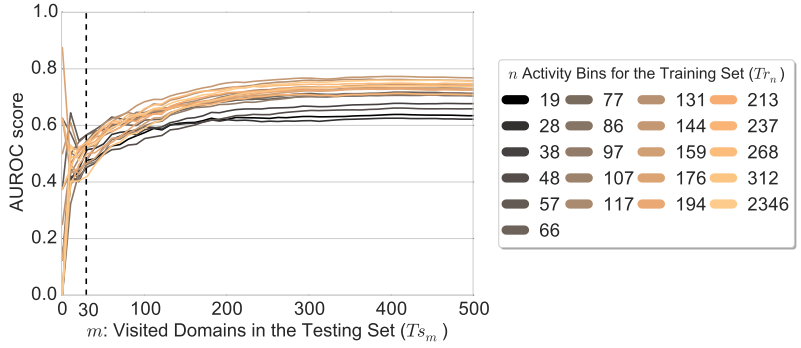
To address this question, the focus of attention is placed on understanding the relationship between the prediction score and the amount of information required to achieve that score. Our expectation is that the more information we have about an individual the better our models will perform until they reach a saturation level. To exploit this trade-off we performed an extensive analysis for the prediction of the "gender" inferred from the desktop browsing activity, while increasing steadily the amount of information (number of domains) contained in the training set. In the desktop dataset, there are users that have visited only one domain (referred to as minimum user activity) while others have visited 2346 unique domains (referred to as maximum user activity). We split the entire dataset into a training set Tr, consisting of the 80% of the samples and a testing, Ts, containing the 20% of the samples. Leaving the testing set aside, we further split the training set creating 20 subsets one per activity bin n, namely Tr_n. We quantised the total activity of users considering $n = 20$ activity bins that ranged from the minimum to the maximum user activity (with $n \in [1,2346]$). Fig. 2a shows in dotted lines the ranges of the activity bins. In this way, the first activity bin $n = 1$ contains users that have visited up to 19 websites and so forth (see legend of Fig. 2b for the exact number of maximum domains per activity bin n). This approach ensures that in each step there is a steady increase of the amount of information in terms of unique domains. Therefore, in each subset Tr_n, are included only the users that have visited, in total, an amount of unique websites that are in the range defined by the activity bin n; for instance, Tr_3 includes all users that have visited less than 38 unique websites. We trained one Random Forest (RF) model for each training set Tr_n, employing a five-fold validation approach.

We validated each RF model against the testing set Ts which was initially kept apart. The amount of information included in the testing set also influences the performance of a classifier, therefore, following a similar logic, we created a series of testing subsets, namely Ts_m. In these subsets where only users that visited at most m domains were included. Overall, we considered 100 activity bins, m, with $m \in [1,1002]$ since the maximum user activity in the testing set Ts was 1002 unique domains.

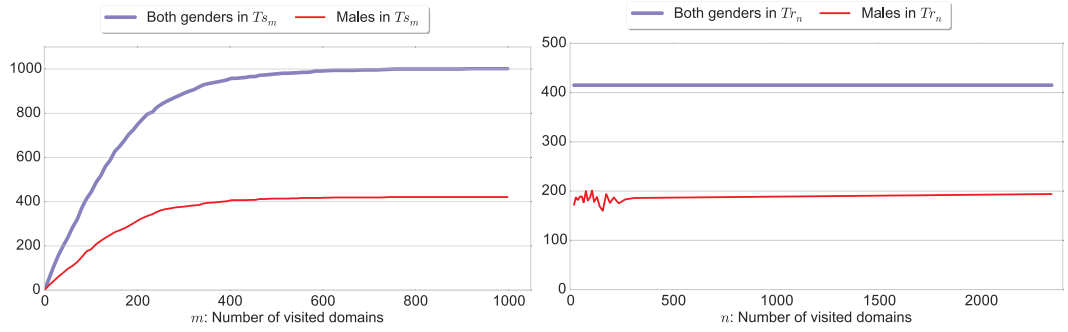
Fig. 2b depicts the performance of each RF model, as a continuous line whose colour mapping ranges from darker to brighter tones according to the increasing activity bin n. Each RF model is trained on a specific Tr_n and validated against all the testing subsets Ts_m. The more information is contained in the RF model in the training phase the brighter the colour mapping of its performance over the different testing sets. From this comparison, we can set a minimum activity threshold of web browsing activity that the users should meet for the learning models to be robust. This threshold may be set approximately at $m = 30$ domains, after which the fluctuations in AUROC metric scores seem to stabilise for all models (see dotted line in Fig. 2b). As expected, the more information is present in the training set the higher



(a) Distribution of unique domains seen by user (in log scale).



(b) AUROC values for increasing number of domains.



(c) Proportion of labels in the Ts_m.

(d) Proportion of labels in the Tr_n.

Fig. 2. Fig. 2a presents the activity bins in which we segmented the user activity in log scale. Fig. 2b reports the AUROC performances for each trained model Tr_n, with n ranging from 20 to 2346, compared against all the testing sets Ts_m. Each model is represented by a coloured line ranging from darker to brighter colours for increasing values of n. Note in Fig. 2b, for Ts_m, we plotted only the activity range between 0 and 500 to zoom in the initial interval of interest. For n greater than 500 the behaviour of the model remains invariant. Fig. 2c and d instead depict the size the testing and training sets respectively. We note that the size of all training sets, Tr_n is kept fixed for all models. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the AUROC scores. The user activity improves the AUROC score from 70% to 80% (see Fig. 2b); models trained on users that have only visited a few domains (darker colours) are systematically performing worse than the ones trained on users who have visited plenty of domains (models with brighter colours). Models trained on less than 50 domains (darker lines) deviate significantly from the others, while they fluctuate more intensively when the testing set contains subjects with only few visited web browsing domains $m < 30$. These observations indicate that a generic activity threshold set at 30 domains benefit the model both in the training and in the testing phase.

Fig. 2c and 2d depict the distribution of labels (males, females) for all the training and testing subsets. Worth-mentioning is the fact that the size of the training set is forced to be fixed for all the Tr_n to avoid biases related to the size of the training/testing set instead of its informative content. We defined a maximum sample size for men and women, C_m and C_w respectively, to be included in the respective training sets. So for example if in Ts_1 we have approximately 200 women and 400 men, then for every Ts_n we would randomly pick 200 women and 400 men. Fig. 2c and 2d depict the distribution of the labels for each training Tr_n and testing Ts_m subset.

4.2. Sensitivity of the model to the quality of available information

This section presents an extensive analysis performed only on the most "active" users; as "active" we define a user that has visited at least 200 unique domains, given that the average number of unique domains

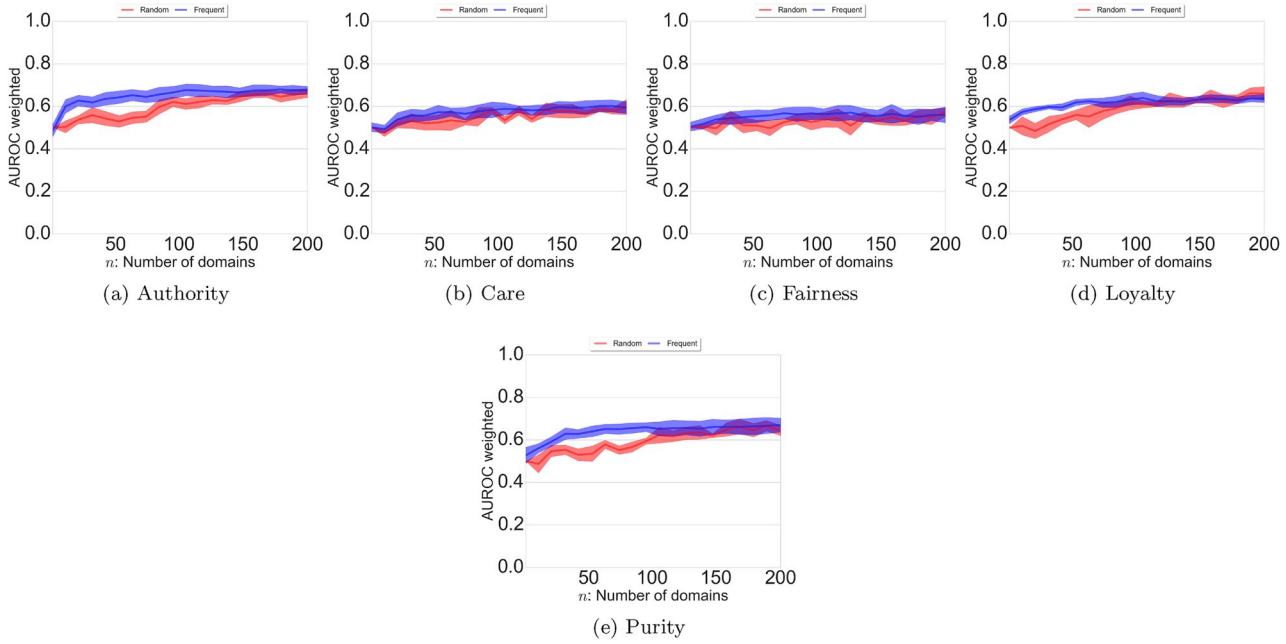


Fig. 3. AUROC analysis for the MFT attributes. Increasing the number of domains visited according to their frequency of visit (blue line, *Case 1*) and increasing the number of domains visited regardless of their frequency of appearance (red line, *Case 2*), in the training sets, while always validated on the same testing set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

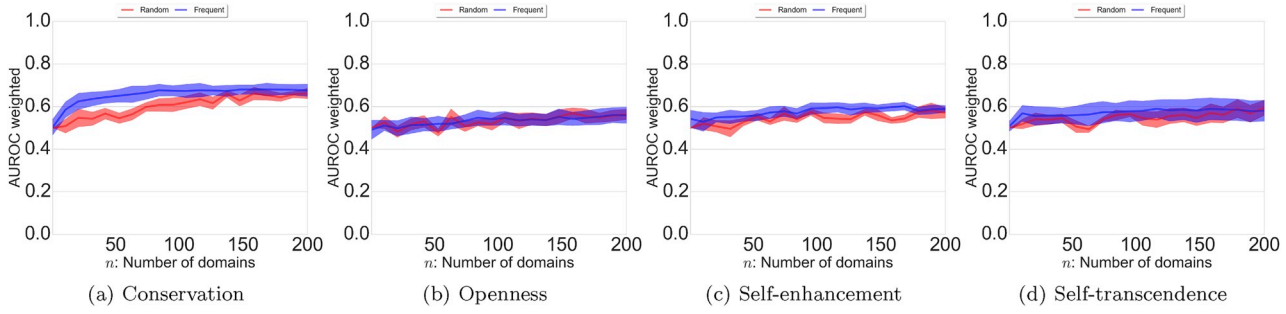


Fig. 4. AUROC analysis for the Schwartz Quadrants. Increasing the number of domains visited according to their frequency of visit (blue line, *Case 1*) and increasing the number of domains visited regardless of their frequency of appearance (red line, *Case 2*), in the training sets, while always validated on the same testing set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

visited by a user equals to 178 in the entire desktop dataset. The scope of this analysis is to assess the sensitivity of the models to the amount of informative content contained in the training data and its effect on the final prediction accuracy. We performed this exploration for all the demographic and psychometric attributes.

Initially, we filtered our entire dataset with a minimum activity threshold set at *thresh* = 200 domains, maintaining only the highly active users for the rest of our analysis (1210 participants in total). Sequentially, we split this into a training set, *Tr active*, and a testing set, *Ts active*, consisting of the 80% and 20% of the samples, respectively. The *Ts active* is only employed later in the validation phase. At this point we want to explore what is the effect of sample size in training in the following two cases; first (i) picking first the most active users (users with a plethora of visits to different websites) and second (ii) randomly picking users regardless of their activity, always from our *Tr active* set. For the first case, we ranked the users by their number of visits in unique domains, while for the second case the selection is random, hence, their probability of taking part in the training set is the same.

Case 1. *Websites Ranked by Amount of Visits.* Having ranked our users according to their web browsing activity, we further split our training set *Tr active* in 20 subsets, namely Tr_n *freq*, with n uniformly distributed in the interval [1,200]. I.e., for Tr_n *freq*, the feature vector for each user would contain his/her n most visited websites. For each attribute, we trained an RF model on each Tr_n *freq* employing a five-fold validation approach. Successively, we validated all Tr_n *freq* models against the *Ts active* set.

Case 2. *Websites Selected Randomly.* Here, we split our initial training set *Tr active* in 20 subsets, namely Tr_n *random*, with n uniformly distributed in the interval [1,200]. I.e., for Tr_n *active*, the feature vector for each user would consist on n web domains randomly picked from his/her pool of domains. For each attribute, we trained an RF model on each Tr_n *random* employing a five-fold validation approach. Successively, we validated all Tr_n *random* models against the *Ts active*.

Blue curves in Figs. 3–6 depict the weighted AUROC metric for each considered attribute of the *Case 1*, the most frequent domain selection. Each point of the blue line indicates the average weighted AUROC metric, for the five-folds, for the model trained on the Tr_n *freq* set and tested on the *Ts active* set; while the shaded interval indicates its standard deviation. From this exploratory analysis, we note that increasing the information contained in the training sets has a limited effect on the weighted AUROC metric after a brief stabilisation step noted approximately for $n < 40$ domains. Red curves in Figs. 3–6 depict the weighted

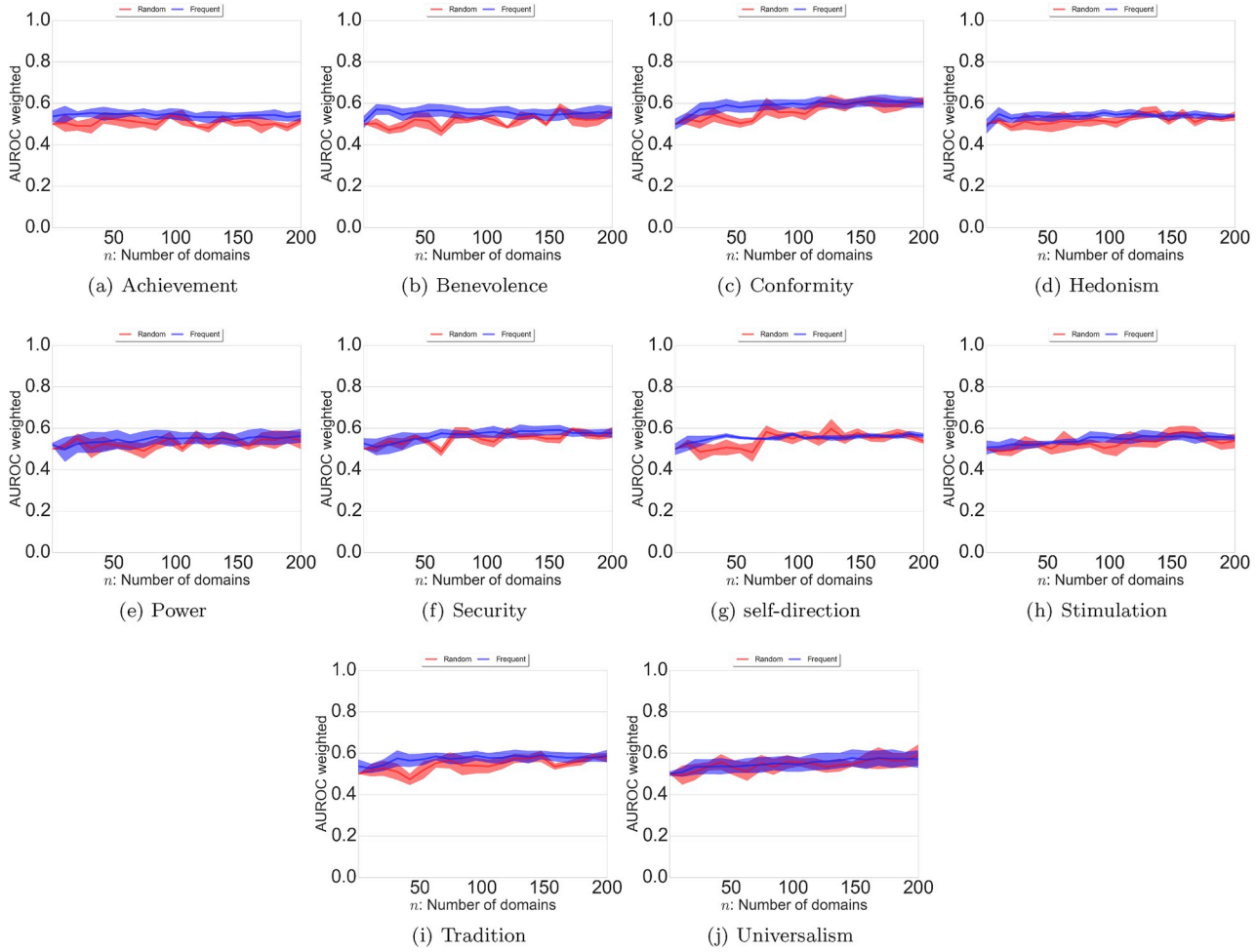


Fig. 5. AUROC analysis for the Schwartz Values. Increasing the number of domains visited according to their frequency of visit (blue line, *Case 1*) and increasing the number of domains visited regardless of their frequency of appearance (red line, *Case 2*), in the training sets, while always validated on the same testing set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

AUROC metric for each attribute of the *Case 2*, the random domain selection. Each point of the red line indicates the average weighted AUROC metric, for the five-folds, for the model trained on the Tr_n *random* set and tested on the *Ts active* set; while the shaded interval indicates its standard deviation.

Interestingly, even with the experimentation with random selection of active users in the training set we noticed only a limited effect on the average weighted AUROC metric. There are, however, more fluctuations due exactly to the random amount of information included in each training set with respect to the first approach where the increase of the training set size, only improved the performances. This finding suggests, as expected, that models trained on the most frequent domains have higher prediction ability, even if trained on a small number of participants.

5. Results

5.1. Moral foundations

For the moral foundations and the individualistic/binding hypercluster, we obtained the following prediction scores: 66% for authority, 62% for care, 58% for fairness, 63% for loyalty, 67% for purity, and 66% for individualists, respectively.[6] The web domains from Desktop browsing were more informative than mobile data for the prediction. The first part of Table 6 reports the average weighted AUROC metric and the respective standard deviations over five-folds, for models trained on the web domains of the Desktop dataset (first column), on the web domains of the Mobile dataset (second column), on solely Application usage (third column), and finally on the fusion of web browsing and application usage of the Mobile dataset (fourth column).

In Table 7 we indicatively report the ten most predictive domains (in typewriter style) and applications (in uppercase style) for each model. Finally, Table 8 presents the most important predictors focusing on the individualistic versus binding foundations.

5.2. Schwartz human values

The prediction of Schwartz's basic human values and quadrants proved to be a more difficult classification problem with respect to the moral foundations one. The average weighted AUROC metric for all the attributes was around 60%, with the most accurate predictions being the ones for the conservation 63% and universalism 62% values. As for the moral foundation prediction, we employed as predictors (i) the desktop browsing domains, (ii) the mobile browsing, (iii) the application usage, and (iv) the fusion of mobile browsing and applications. The second part of Table 6 reports the weighted AUROC metric for each of the Schwartz's basic human values and quadrants, while Table 9 reports the ten most predictive domains and applications for each model.

Comparing these findings with the state-of-the-art literature we see

[6] All prediction scores refer to the weighted AUROC metric.

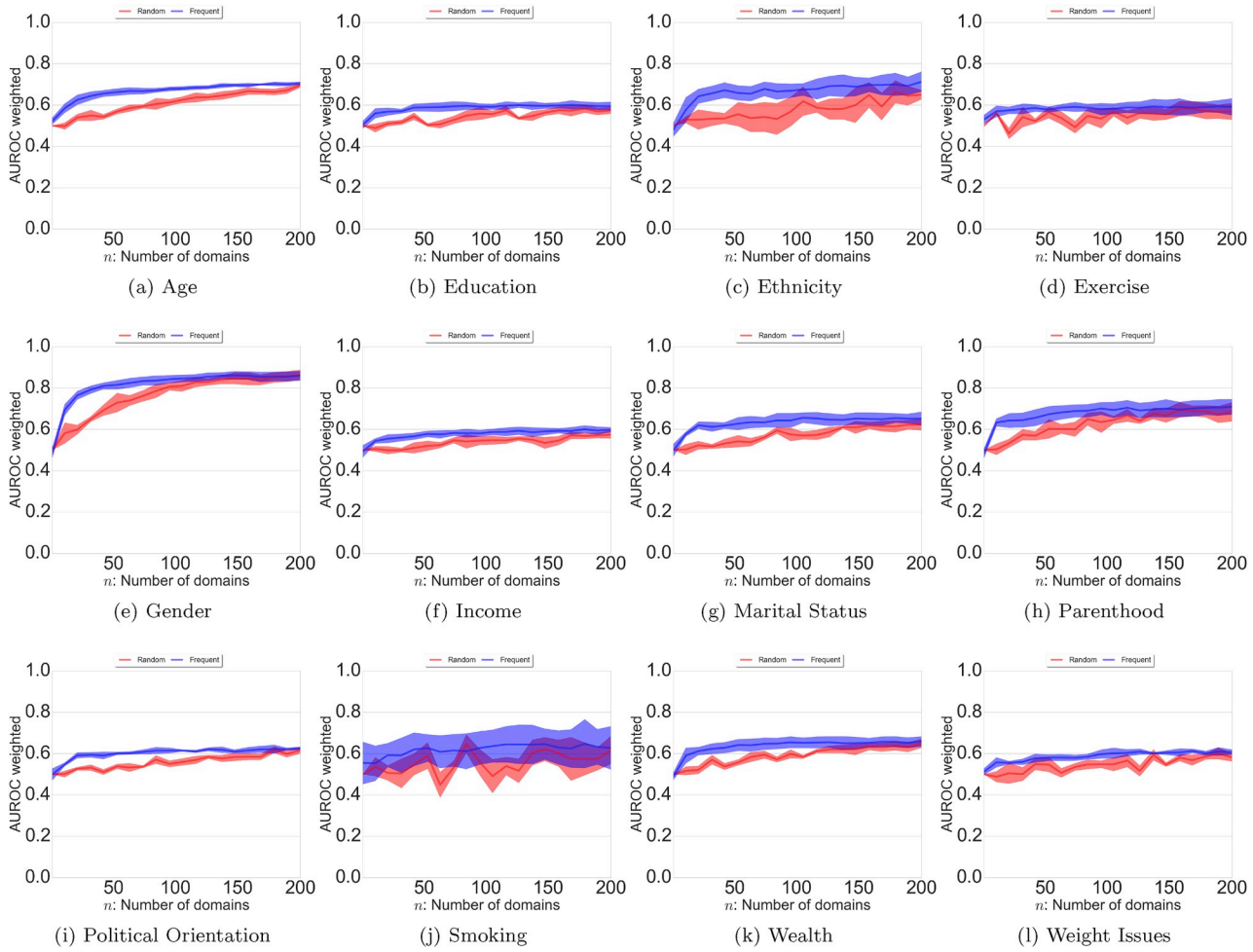


Fig. 6. AUROC analysis for the demographic attributes. Increasing the number of domains visited according to their frequency of visit (blue line, *Case 1*) and increasing the number of domains visited regardless of their frequency of appearance (red line, *Case 2*), in the training sets, while always validated on the same testing set. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the performances obtained despite being low, are in line. A study we can directly compare with, was performed by Chen et al., 2014, who employed supervised binary machine learning algorithms to classify human values in a binary scheme with above-median levels of each value dimension and reported their results using AUROC metric. From their best-performing classifiers, they obtained 60% for Self-Transcendence, 56% for Self- Enhancement, 59% for Conservation, 61% for Openness to change, and 61% for Hedonism. Except from the fact that Chen et al., 2014 inferred on linguistic features from social media data, the performances they obtained are close to ours demonstrating the difficulty of predicting these values.

5.3. Demographic data

The third part of Table 6 reports in detail the weighted AUROC metric for all the demographic attributes, while Table 10 indicatively reports the ten most predictive domains and applications for each model. We note that the demographic data were more accurately predicted with the “gender” attribute outperforming all others with a 90% AUROC score. Moving to less commonly studied demographic attributes as age 71%, marital status 67%, parenting 72%, and ethnicity 74% our models have a satisfactory performance. Income and wealth were inferred with 60% and 66%, respectively. Table 11 reports the findings of the state-of-the-art studies with which we can directly compare our findings since they reported also the AUROC metric. For the “age” attribute the prediction scores vary according to the number of classes considered; in our study we considered 6 classes and achieved 71% which is a good performance. Moreover, we addressed health-related attributes like weight issues and the practice of physical exercise, achieving 63% and 64% weighted AUROC scores respectively. We were also able to capture the smoking behaviour of individuals with an accuracy of 64%. These AUROC scores show that digital footprints can contribute to the health communication campaigns targeting individuals in need.

For the prediction of demographic variables, we considered all the categories (labels) reported in Table 1, which of course hardens the classification tasks. Complex attributes like education level 59% and political orientation 60% were difficult to infer, but if we consider only the two major political parties, namely Democrats and Republicans, the weighted AUROC prediction score reaches 70%. Table 1 provides a complete list of all the categories considered for each predicted attribute.

6. Discussion

The obtained results for moral foundations and human values suggest that online behaviours are potentially informative of the individuals' worldviews and ideals. The poor to medium prediction scores are probably related to the complexity of the attributes that we are trying to infer; in fact, morals are often expressed in subtle ways in everyday life, and only occasionally in a more intense way under specific circumstances, making them more difficult to assess from digital

Table 6
The prediction scores of moral attributes, human values and demographics expressed in average weighted AUROC scores. We trained a model for each attribute and for each category of predictors (i) features of desktop web browsing, (ii) mobile browsing, (iii) mobile application usage and (iv) fusion of mobile browsing and application usage. In parenthesis, the weighted AUROC's standard deviation (STD) of the five-fold cross-validation process is reported. The best-performing features are indicated in **bold**.

Morals	Desktop Web Domains (STD)	Mobile Web Domains (STD)	Mobile Apps (STD)	Mobile Apps Domains (STD)
Authority	**0.66 (0.02)**	0.63 (0.03)	0.60 (0.02)	0.64 (0.05)
Care	**0.62 (0.02)**	0.55 (0.05)	0.57 (0.04)	0.55 (0.03)
Fairness	**0.58 (0.01)**	0.57 (0.04)	0.55 (0.01)	0.55 (0.02)
Loyalty	**0.63 (0.02)**	0.60 (0.03)	0.55 (0.04)	0.58 (0.05)
Purity	0.64 (0.01)	0.62 (0.02)	**0.67 (0.04)**	0.66 (0.04)
Individualist/Binding	**0.66 (< 0.01)**	0.65 (0.02)	0.64 (0.02)	**0.66 (0.02)**
Human Values	**Desktop Web Domains (STD)**	**Mobile Web Domains (STD)**	**Mobile Apps (STD)**	**Mobile Apps Domains (STD)**
Conservation	**0.63 (0.02)**	0.61 (0.05)	0.59 (0.06)	0.62 (0.05)
Openness	0.56 (0.02)	0.56 (0.03)	0.57 (0.03)	**0.58 (0.04)**
Self-enhancement	**0.59 (0.01)**	0.51 (0.01)	0.57 (0.04)	0.56 (0.04)
Self-transcendence	0.58 (0.01)	0.57 (0.02)	0.58 (0.05)	**0.59 (0.02)**
Achievement	0.56 (0.03)	0.54 (0.03)	0.56 (0.04)	**0.59 (0.04)**
Benevolence	0.54 (0.01)	0.53 (0.04)	0.50 (0.02)	**0.56 (0.01)**
Conformity	**0.59 (0.01)**	0.57 (0.04)	0.58 (0.03)	0.58 (0.04)
Hedonism	0.56 (0.02)	0.57 (0.03)	0.59 (0.05)	**0.60 (0.03)**
Power	0.56 (0.02)	0.55 (0.03)	0.55 (0.01)	**0.58 (0.05)**
Security	**0.58 (0.02)**	0.56 (0.03)	0.53 (0.04)	0.55 (0.03)
Self-direction	0.56 (0.02)	0.53 (0.03)	**0.57 (0.03)**	0.53 (0.02)
Stimulation	0.56 (0.01)	0.55 (0.03)	0.56 (0.03)	**0.58 (0.02)**
Tradition	0.58 (0.02)	**0.60 (0.03)**	0.58 (0.06)	**0.60 (0.03)**
Universalism	0.59 (0.01)	0.57 (0.02)	**0.62 (0.06)**	0.61 (0.04)
Demographics	**Desktop Web Domains (STD)**	**Mobile Web Domains (STD)**	**Mobile Apps (STD)**	**Mobile Apps Domains (STD)**
Age	**0.71 (0.01)**	0.68 (0.03)	**0.71 (0.03)**	**0.71 (0.02)**
Education	**0.59 (0.01)**	0.57 (0.01)	**0.59 (0.02)**	**0.59 (0.01)**
Ethnicity	0.73 (0.02)	0.69 (0.02)	0.72 (0.05)	**0.74 (0.02)**
Exercise	0.61 (0.02)	0.59 (0.02)	**0.63 (0.04)**	0.60 (0.03)
Gender	0.86 (0.01)	0.88 (0.02)	**0.90 (0.02)**	0.89 (0.02)
Income	**0.60 (0.01)**	0.55 (0.02)	0.60 (0.02)	0.58 (0.01)
Marital Status	**0.67 (0.02)**	0.61 (0.02)	0.64 (0.01)	0.63 (0.03)
Parent	0.71 (0.01)	0.66 (0.02)	**0.72 (0.04)**	0.69 (0.04)
Political Orientation	0.58 (0.01)	0.58 (0.02)	0.59 (0.03)	**0.60 (0.02)**
Smoker	0.63 (0.06)	0.59 (0.06)	0.62 (0.03)	**0.64 (0.04)**
Wealth	**0.66 (0.01)**	0.60 (0.02)	0.61 (0.02)	0.62 (0.02)
Weight Issues	**0.62 (0.02)**	0.58 (0.02)	0.58 (0.03)	0.60 (0.02)

behaviours with respect to the well-studied topic of personality traits recognition. Taking as an example the Big-Five model (Goldberg, 1990) for personality assessment, Kosinski et al. (2013), one of the most influential studies on the topic, obtained Pearson correlations in the range [0.29–0.43] on their regression models for the prediction of Big-Five traits. A major issue for automatic personality recognition is insufficient benchmark data and a consensus on the reported metrics and labelling strategies, leading to the lack of a direct comparison of the obtained results across the state of the art studies. Despite that, a common finding that emerge is that the best performing attribute in the state-of-

Table 7
Top ten websites and applications emerging as the best predictors for each moral foundation from mobile records (fusion of Web browsing and Application usage). Uppercase style indicates an APPLICATION while typewriter style indicates a Web Domain. (+) and (−) signs refer to whether the specific behaviour was employed by the model to classify an individual as “higher” or “lower” in the specific moral.

Authority	Care	Fairness	Loyalty	Purity
iflscience.com (−)	MOBILE HOTSPOT (+)	SHOPKICK (−)	americanexpress.com (+)	BIBLE (+)
npr.org (−)	SNAPCHAT (−)	ijreview.com (−)	GMAIL (−)	YELP (−)
HANGOUTS (−)	theanimalrescuesite.com (+)	mookie1.com (−)	INSTAGRAM (−)	google.com (−)
thinkprogress.org (−)	TWITTER (−)	MOBILE HOTSPOT (+)	PANDORA (−)	MAPS (−)
huffingtonpost.com (−)	PAPA JOHN’S (−)	accuweather.com (−)	wikipedia.org (−)	LINKEDIN (−)
thoughtcatalog.com (−)	AMAZON KINDLE (−)	facebook.com (+)	CLOUD (+)	iflscience.com (−)
REWARDS (−)	BING (−)	marriott.com (+)	YOUTUBE (+)	PANDORA (−)
petco.com (−)	change.org (+)	change.org (+)	craigslist.org (−)	facebook.com (−)
SOLITAIRE (+)	about.com (+)	imgur.com (−)	MAPS (−)	yp.com (+)
huffpost.com (−)	ihg.com (+)	care2.com (+)	yahoo.com (−)	wikipedia.org (−)

Table 8
Top ten websites and applications emerging as the best predictors of individualistic and binding foundations for two distinct models; the first one trained on Desktop browsing data and the second one on the fusion of mobile browsing and applications usage. The weighted AUROC metric is 66% for both cases. We note that among the predictors we find many news websites which are in line with the expected profiles of the two foundations. Moreover, there is consistency in the correlation of certain predictors (e.g. wikipedia.org) with a specific foundation indicating the stability of the approach. (*B*) and (*I*) signs refer to whether the specific behaviour was employed by the model to classify an individual as "Binders" or "Individualists".

Desktop Web Domains	Mobile Web and Apps
google.com (I)	huffingtonpost.com (I)
foxnews.com (B)	FOX NEWS (B)
dailykos.com (I)	BIBLE (B)
yelp.com (I)	google.com (I)
imdb.com (I)	accuweather.com (B)
cnsnews.com (B)	HANGOUTS (I)
wikipedia.org (I)	EMERGENCY ALERTS (B)
mrctv.org (B)	facebook.com (I)
theblaze.com (B)	GOSPEL LIBRARY (B)
thepetitionsite.com (I)	wikipedia.org (I)

the-art is the extroversion trait commonly associated with out-going people who engage more in social interactions. These behaviours are in general quantifiable; the assessment of intense face to face interaction and communication patterns is straightforward by a variety of digital data (Finnerty et al., 2016; Vinciarelli & Mohammadi, 2014). Conversely, fairness as defined in the Moral Foundation Theory, may have many facets and ways to be expressed depending on the event that triggered the behaviour, resulting in complex dependencies between the observed data and the respective attribute value.

Taking a deeper look into the most predictive features of the moral foundations, ranked according to their Gini importances (Table 7), interesting insights emerge. *The Huffington Post* and *ThinkProgress* which are usually considered progressive sources of news and information, that perceive the notion of authoritarian displays of power and arbitrary order as negative (Mitchell, Gottfried, Kiley, & Matsa, 2014), arise as one of the main indicators of subversion. Visiting change.org acts as an important indicator of concern for equity and others' suffering since it organises around topics meant to invoke social change, especially for those experiencing victim status. TheAnimalRescue.com is evocative of nurturing and caring sentiments, and emerges as one of the main indicators for care. Individuals who consider themselves as pure can be identified by the frequent use of applications like the *Bible* or the *Gospels*, and they tend to follow *Fox News* rather than *The New York Times*. For the individualist and binding foundations (Table 8), the websites and applications that emerge as differentiators are strongly reflecting the choice of information sources, for instance huffingtonpost.com for the individualists and foxnews.com for the binders. These findings are in line with existing literature studies on news selectivity and mechanisms of political opinion formation (Yeo, Cacciatore, & Scheufele, 2015). Moreover, according to a year-long Pew Report regarding political polarisation and the sources people usually get informed from, people "with ideological views on the left and right have information streams that are distinct from those of individuals with more mixed political views and very distinct from each other" (Mitchell et al.,

Table 9
Top ten websites and applications emerging as the best predictors for each of the Schwartz values and quadrants from mobile records (fusion of Web browsing and Application usage). Uppercase style indicates an APPLICATION while typewriter style indicates a Web Domain. (+) and (−) signs refer to whether the specific behaviour was employed by the model to classify an individual as "higher" or "lower" in the specific value.

Conservation	Openness	Self-enhancement	Self-transcendence	Achievement	Benevolence
SNAPCHAT (−)	SNAPCHAT (+)	WHATSAPP (+)	EBAY (−)	GMAIL (+)	EBAY (−)
iflscience.com (−)	INSTAGRAM (+)	SOUTHWEST (+)	GAMESTOP (−)	INSTAGRAM (+)	WEATHER (+)
MAPS (−)	facebook.com (+)	linkedin.com (+)	bitecharge.com (+)	FACEBOOK (+)	mpstat.us (+)
indiegogo.com (−)	FOX NEWS (−)	SNAPCHAT (+)	MINT (−)	kinja.com (+)	imdb.com (−)
UBER (−)	POLARIS OFFICE (+)	ACCUWEATHER (−)	mlnwap.com (−)	NETFLIX (−)	battle.net (−)
lgbtqnation.com (−)	MAPS (+)	SCOUT (−)	kmart.com (−)	BIBLE (−)	emgn.com (+)
npr.org (−)	EMERGENCY ALERTS (−)	xojane.com (+)	nytimes.com (+)	target.com (+)	LINE (−)
amazon.com (−)	SPRINT HOTSPOT (+)	GMAIL (+)	FANTASY SPORTS (−)	GOOGLE MOBILE (+)	facebook.com (+)
huffpost.com (−)	CAPITAL ONE (+)	TWITTER (+)	huffpost.com (+)	lds.org (−)	sparkpeople.com (−)
BIBLE (+)	PINTEREST (−)	MINT (+)	FACEBOOK (−)	TWITTER (+)	FIREFOX (−)
Conformity	**Hedonism**	**Power**	**Security**	**Self-direction**	**Stimulation**
zimbio.com (−)	INSTAGRAM (+)	aarp.org (−)	MAPS (−)	FACEBOOK (−)	MAPS (+)
iflscience.com (−)	YELP (+)	TAPPED OUT (+)	REGAL (+)	YOUTUBE (−)	google.com (+)
FOX NEWS (+)	YOUTUBE (+)	simplyhired.com (−)	papajohns.com (−)	PHOTO GRID (−)	facebook.com (+)
BIBLE (+)	PANDORA (+)	att.com (−)	YAHOO! MAIL (+)	PINTEREST (−)	uproxx.com (+)
YOUTUBE (+)	FOURSQUARE (+)	nerdist.com (+)	myfitnesspal.com (−)	PLAY STORE (−)	YOUTUBE (+)
KROGER (+)	FACEBOOK (+)	PLAY STORE (+)	SPRINT HOTSPOT (−)	bitecharge.com (+)	FACEBOOK (+)
wikipedia.org (−)	IBOTTA (−)	INKPAD (−)	instagram.com (−)	amazon.com (+)	SHAZAM (+)
yelp.com (−)	WHATSAPP (+)	SPOTIFY (+)	SPOTIFY (−)	ttaa.com (+)	ebay.com (+)
GALAXY APPS (+)	MAPS (+)	imdb.com (+)	WALMART (+)	CANDY CRUSH SAGA (−)	buzzfeed.com (+)
amazon.com (−)	SHAZAM (+)	southwest.com (−)	aarp.org (+)	groupon.com (−)	YELP (+)
Tradition	**Universalism**				
amazon.com (−)	FOX NEWS (−)				
YELP (−)	TWITTER (−)				
YOUTUBE (−)	FACEBOOK (−)				
guff.com (−)	PEOPLE (−)				
SNAPCHAT (−)	INSTAGRAM (−)				
iflscience.com (−)	EBAY (−)				
facebook.com (−)	twitter.com (+)				
FACEBOOK (−)	EVERNOTE (+)				
wikipedia.org (−)	GROUPON (−)				
google.com (−)	couponsherpa.com (−)				

Table 10
Top ten websites and applications emerging as the best predictors for each of the demographic attributes from mobile records (fusion of Web browsing and Application usage). Uppercase style indicates an APPLICATION while typewriter style indicates a Web Domain. The signs next to each behaviour refer to whether the specific behaviour was employed by the model to classify an individual to a specific demographic group.

Education	Income	Political Party	Wealth	Ethnicity
linkedin.com (*postGrad*)	linkedin.com (200*KPlus*)	MAPS (*L*)	YOUTUBE (50*KLess*)	YELP (*A*)
PLAY STORE (*someCollege*)	GAMESTOP (20*KLess*)	reddit.com (*D*)	buzzfeed.com (50*KLess*)	SAFEWAY (*A*)
YOUTUBE (*someCollege*)	SNAPCHAT (20*KLess*)	complex.com (*D*)	INSTAGRAM (50*KLess*)	LINE (*A*)
LINKEDIN (*postGrad*)	YOUTUBE (20*KLess*)	facebook.com (*D*)	AMERITRADE (1000*KPlus*)	slickdeals.net (*A*)
jw.org (*high − school*)	TUMBLR (20*KLess*)	FLIPBOARD (*D*)	FACEBOOK (50*KLess*)	YOUTUBE (*H*)
google.com (*postGrad*)	FACEBOOK (30*Kto*50*K*)	imgur.com (*R*)	PINTEREST (50*KLess*)	PINTEREST (*W*)
MI MOBILE (*high − school*)	GMAIL (150*Kto*200*K*)	COLOR NOTE (*L*)	SNAPCHAT (50*KLess*)	wetpaint.com (*AA*)
att.com (*tradeSchool*)	adadvisor.net (200*KPlus*)	iflscience.com (*D*)	AMAZON (100*Kto*250*K*)	WECHAT (*A*)
gamestop.com (*high − school*)	google.com (200*KPlus*)	diply.com (*L*)	facebook.com (50*KLess*)	WHATSAPP (*H*)
FACEBOOK (*someCollege*)	PLAY STORE (30*Kto*50*K*)	NFL FANTASY FOOTBALL (*L*)	NYTIMES (1000*KPlus*)	BIBLE (*AA*)

Gender	Exercise	Age	Marital Status	Parent
PINTEREST (*F*)	FITBIT (*Y*)	SNAPCHAT (18*to*24)	GMAIL (*S*)	SNAPCHAT (*N*))
CARTWHEEL BY TARGET (*F*)	MAPS (*Y*)	INSTAGRAM (18*to*24)	NCPMOBILE(*W*)	GMAIL (*N*)
FACEBOOK (*F*)	MAPMYRUN (*Y*)	buzzfeed.com (25*to*34)	FLIPBOARD(*W*)	elitedaily.com (*N*)
playbuzz.com (*F*)	active.com (*Y*)	YOUTUBE (25*to*34)	NCP(*W*)	YELP (*N*)
INSTAGRAM (*F*)	ANGRY BIRDS (*N*)	FACEBOOK (25*to*34)	SNAPCHAT (*S*)	buzzfeed.com (*N*)
buzzfeed.com (*F*)	MYFITNESSPAL (*Y*)	PINTEREST (25*to*34)	google.com (*S*)	TUMBLR (*N*)
SPORTS CENTER (*M*)	yelp.com (*Y*)	google.com (18*to*24)	jetblue.com (*M*)	INSTAGRAM (*N*)
PLAY STORE (*M*)	YELP (*Y*)	TUMBLR (18*to*24)	CHASE (*D*)	SWARM (*N*)
gap.com (*F*)	yahoo.com (*N*)	facebook.com (25*to*34)	YOUTUBE (*S*)	UBER (*N*)
target.com (*F*)	ESPN (*Y*)	aarp.org (55*to*64)	FACEBOOK (*LT*)	MAPS (*N*)

Weight	Smoker
FACEBOOK (*Y*)	SNAPCHAT (*N*)
express.com (*N*)	sprintpcs.com (*Y*)
MAPS (*N*)	mailchimp.com (*N*)
WORDSWITHFRIENDS (*Y*)	google-analytics.com (*N*)
AMAZON KINDLE (*Y*)	psychcentral.com (*Y*)
LOSE IT! (*Y*)	TRIP ADVISOR (*N*)
YAHOO! MAIL (*Y*)	about.com (*N*)
bodybuilding.com (*N*)	18andabused.com (*Y*)
UBER (*N*)	SPORTS CENTER (*N*)
GMAIL (*N*)	google.com (*N*)

2014).

Regarding the demographic attributes' prediction, the “gender” is predicted with the highest AUROC score 90%. The study by Kosinski et al. (2013) reported an AUROC of 93%, when employing the users' “Likes” on the Facebook platform as predictors and is the only one that outperforms our score. The prediction score for gender is of great importance not because of the high accuracy achieved but because it is the most broadly studied attribute in the literature and one of the few that can be employed assess the validity and potentials of our study with respect to the state-of-the-art literature. Table 11 reports the findings of the state-of-the-art studies against which we can directly compare our findings since they reported also the AUROC metric on gender, age, parenthood and marital status from various sources of digital data. For the “age” attribute the prediction scores vary according to the number of classes considered; in our study we considered six classes and achieved 71% which is a good performance. To compare our methodology against the study of Malmi and Weber (2016), we re-trained the predictive model for the “age” attribute as single-class classification; dividing our sample in approximately two balanced subsets, under and over 49 years old. The AUROC in this case was 86% which is slightly higher than the performance of Malmi and Weber (2016). The “smoker” attribute was predicted with a higher score using Facebook likes 73% in Kosinski et al. (2013) against 64% in our case. It is not straightforward to perform a direct comparison of the remaining demographic attributes with studies close to ours since they often opted for binary classifiers for all labels, while in our scenario we employed multi-class classification. In general, multi-class classification is a much harder task; taking the political party prediction as an example, we remark our reported AUROC score of 60% increases to 70% if we focus only on the two most populated classes (“Democrats” and “Republicans”). The latter score is higher than the AUROC of 63% reported in Youyou et al. (2015) for the binary prediction of the political orientation of Facebook users, but lower than the 85% value in Kosinski et al. (2013). Indicatively, considering a binarisation strategy for our multi-class demographic attributes we have the following scores: education 63%, ethnicity 77%, income 70%, marital status 65%, and wealth 71% (see Supplementary Information 2, Table 2 for details regarding the binarisation strategy).

Shifting our attention to the most important indications of demographic attributes; among the patterns for education levels, Linkedin.com emerges as the most significant indicator of higher education, and FACEBOOK as an indicator of college education. Interestingly, the website of the Jehovah's Witnesses Church is an indicator of a lower educational level (high-school); these findings are consistent with studies on educational trends and social media usage, as reported in the Gallup Poll results in the US (Newport, 2012) and on educational trends and religion (Stark & Iannaccone, 1997). The top gender indicators *Pinterest* and *Cartwheel by Target* were also pointed out as representative of gender in another study (Malmi & Weber, 2016). As for the prediction of age, the social network preferences are the most predictive cues, for instance SNAPCHAT which is known to be very popular in the age group of 18*to*24. Indicatively, *LinkedIn* and *adadvisor.net* emerged as the most prevalent predictors of income for the higher income class, while the most predictive applications and websites for the “exercise” were related to fitness applications like FITBIT and MAPMYRUN. These indicators are cross-referenced by other

Table 11
Related studies on demographic attribute prediction which reported their results in terms of AUC metric which allows for a direct comparison with our results.

Related Study	Gender (AUC)
Facebook Likes Kosinski et al., 2013	93%
Search Queries Bi et al., 2013	80%
Client web browsing history Goel, Hofman, & Sirer, 2012	85%
Apps - Category and Content Seneviratne et al., 2015	74%
Location check-ins Zhong et al., 2015	85–86%
User Applications Malmi & Weber, 2016	90%
Smartphone Call Logs Ying et al., 2012	85%
Smartphone Call Logs Felbo et al., 2015	79.7%
Social Networks Dong et al., 2014	80%
Web browsing Hu et al., 2007	50%
Related Study	**Marital Status (AUC)**
User Applications Malmi & Weber, 2016	79%
Smartphone Call Logs Ying et al., 2012	79%
Related Study	**Parenthood (AUC)**
User Applications Malmi & Weber, 2016	68%
Related Study	**Age (AUC)**
Web browsing Hu et al., 2007	50% (5 classes)
Smartphone Call Logs Ying et al., 2012	77% (2 classes)
Social Networks Dong et al., 2014	73% (3 classes)
Smartphone Call Logs Felbo et al., 2015	63% (3 classes)
User Applications Malmi & Weber, 2016	85% (2 classes)

studies contributing to the validity of our approach; however, their exhaustive interpretation is out of our scope.

Shedding light on the limitations of this study, as a first point we must acknowledge the nature of our data. Our web browsing data (of both desktop and mobile) consist of only the higher level domains which are for sure not as informative as the entire URL. For example, from our data, we know that two individuals visited *wikipedia.com* and spend the same amount of time reading an article, however, the topics of the articles are unknown to us. From the quantitative results obtained, we note the performance of the moral foundations' and human values' prediction (AUROC scores between 7% and 17% above the AUROC of a random classifier) is much lower than the average performance for demographic attributes (AUROC scores between 9% and 40% above the AUROC of a random classifier). This observation might be due to the fact that moral and human values are often expressed in specific occasions, perhaps not evident in the everyday digital activity, which instead can provide straightforward information clearly depicting a specific demographic attribute such as “gender”. Moreover, these attributes are often expressed in verbal rather than non-verbal manner, for instance, direct answers to blog discussions or personal opinions on the Twitter platform (Mooijman et al., 2018), rather than the fact that the individual visited a certain blog. All these elements, together with larger volumes of information in terms of a period of digital observation would benefit the performance of Random Forest classifiers. This is evident also from the exploitative study presented in Section 4 where we showed that the AUROC score increases as we include more active users (users that visited plenty of domains). Another point that must be acknowledged is that even if our cohort is closely representative to the US census, we could not account for potential and unavoidable self-selection bias that might have occurred during the recruitment phase. Moreover, even though the data collection period was quite long, we can not account for potential changes in the way our participants navigate the web or use the applications due to the fact that they knew their activity was scrutinised.

Overall, the proposed approach gives insights into the population's behaviour, culture and preferences, providing in-depth insights in the internal decision-making steps of the inference procedure embedded in the predictive models. Exploiting the effect of the quality, nature and/or quantity of the initial data, we aimed at pointing out the importance of understanding how algorithmic and/or data collection choices may impact the findings; issues that are often overlooked. Understandably, there are many possible sources of biases to account for, however, it is important to raise awareness and ensure the transparency of the processes, especially when the outcome of these models is to serve as evidence for decision-making in another field or influence the policy-making.

7. Conclusions

Exploiting digital trails of human behaviour has found numerous applications in the contexts of learning analytics (Gray, 2014; Nistor, Dascalu, & Trausan-Matu, 2016), measurement-based care (Scott & Lewis, 2015), the promotion of well-being (Luhmann, 2017) and crime prevention (Almagor, 2014). We assessed the predictive power of low level digital behavioural sequences on complex psychological attributes like moral traits, human values and a series of advanced demographic attributes through a cross-validated machine learning classification framework. Previous work showed the possibility of inferring demographic and psychometric attributes but most studies are based on platform-specific digital information and without a possibility to obtain ground-truth. In the present study, the cohort engaged is a sample of the US population, closely representative to the US Census (see subsection 3.1.2) with respect to major demographic variables, and not affiliated to a specific web platform or application. For this reason, our design avoids many cultural and demographic biases inherent to the users of specific platforms (Golder & Macy, 2014b). Since our recruited sample is not tied to a specific social network, it is not directly subject to algorithmic manipulation and exposure to content (Kramer, Guillory, & Hancock, 2014). Moreover, its validation was based on self-reported information provided through a concrete survey designed for the scopes of this study.

Our findings suggest that digital traces can be informative of the demographics and can be used to sketch a portrait of emerging cybercultures. Shifting to morals and human values the task becomes much more complex and the prediction accuracy drops significantly, calling for further investigation. These findings are result of an exploitation of different digital data sources and modalities (web browsing, mobile browsing, application usage) which allowed us to perform a comparative study on the predictive power of each modality as well as their combination. Overall, deviations for performance between models inferencing from web browsing and/or smartphone usage were found to be minimal for all attributes.

Undoubtedly machine learning is an essential tool for understanding patterns in human culture and behaviour, exposing stereotypes inherent in our everyday lives, sometimes uncomfortable to acknowledge without such hard evidence. This study highlights the possibility of learning complex demographic and psychological attributes by assessing web browsing data and/or mobile usage while evidencing the extent to which certain attributes can be predicted. Providing a balanced perspective of the risks associated to readily available data like top level domains of the web browsing history, our quantitative results contribute to pointing out the ease in predicting demographic attributes including gender, and ethnicity from all available data sources. Further research is required in this direction for a deeper understanding of bias embedded in the predictive models and decision-making algorithms which may lead to involuntary discrimination (Dwork, Hardt, Pitassi, Reingold, & Zemel, 2012; Edelman, Luca, & Svirsky, 2017; van Miltenburg, 2016; Zliobaite, 2015).

In spite of this, the psychometric attribute prediction case is slightly different; accurate prediction of complex psychometric attributes was still far from fully disclosing psychological profiles calling for further

investigation. With prediction performances ranging from 60% to 70%, we highlight both the potentials and the limitations of such approaches, providing with a realistic dimension of the possibilities for personalised web services as well as the privacy and surveillance concerns that keep raising. Considering that smartphones are among the most widely deployed technologies of human history, the capability of automatically predicting demographic attributes and human/moral values from multi-modal passively collected data is potentially a key enabler for delivering better targeted and more effective interventions at the population scale or nowcasting important social issues like poverty.

Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgement

The authors acknowledge support from the "Lagrange Project" of the ISI Foundation funded by the Fondazione CRT. The authors would like to thank all the participants of the study. The authors also acknowledge Dr. Johan Bollen for his comments and feedback on earlier version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2018.11.024.

References

Almagor, R. (2014). People do not just snap: Watching the electronic trails of potential murderers. *Journal of Civil & Legal Sciences, 3* 2169–0170.

Amin, A. B., Bednarczyk, R. A., Ray, C. E., Melchiori, K. J., Graham, J., Huntsinger, J. R., et al. (2017). Association of moral values with vaccine hesitancy. *Nature Human Behaviour, 1*, 873–880. https://doi.org/10.1038/s41562-017-0256-5.

Bardi, A., & Schwartz, S. H. (2003). Values and behavior: Strength and structure of relations. *Personality and Social Psychology Bulletin, 29*, 1207–1220.

Bi, B., Shokouhi, M., Kosinski, M., & Graepel, T. (2013). Inferring the demographics of search users: Social data meets search queries. *Proceedings of the 22Nd international conference on world wide web WWW '13* (pp. 131–140). New York, NY, USA: ACM. https://doi.org/10.1145/2488388.2488401. URL: http://doi.acm.org/10.1145/2488388.2488401.

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Bonanomi, A., Rosina, A., Cattuto, C., & Kalimeri, K. (2017). Understanding youth unemployment in Italy via social media data. *28th IUSSP international population conference*.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Bureau, U. S. C. (2015). *Pepagesex - annual estimates of the resident population for selected age groups by sex for the United States, states, counties, and Puerto Rico commonwealth and municipios: April 1, 2010 to july 1, 2014, generated by kyriaki kalimeri, using american factfinder*. URL: http://factfinder2.census.gov.

Burke, M., & Kraut, R. (2013). Using facebook after losing a job: Differential benefits of strong and weak ties. *Proceedings of the 2013 Conference on computer supported cooperative work CSCW '13* (pp. 1419–1430). New York, NY, USA: ACM. https://doi.org/10.1145/2441776.2441936. URL: http://doi.acm.org/10.1145/2441776.2441936.

Butt, S., & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior, 24*, 346–360.

Cesare, N., Grant, C., & Nsoesie, E. O. (2017). *Detection of user demographics on social media: A review of methods and recommendations for best practices.* arXiv preprint arXiv:1702.01807.

Chen, D., Fraiberger, S. P., Moakler, R., & Provost, F. (2017). Enhancing transparency and control when drawing data-driven inferences about individuals. *Big Data, 5*, 197–212.

Chen, J., Hsieh, G., Mahmud, J. U., & Nichols, J. (2014). Understanding individuals' personal values from social media word use. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 405–414). ACM.

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing, 17*, 433–450.

Costa, P. T., Jr., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6*, 343.

Dong, Y., Yang, Y., Tang, J., Yang, Y., & Chawla, N. V. (2014). Inferring user demographics and social strategies in mobile social networks. *Proceedings of the 20th ACM SIGKDD international Conference on knowledge Discovery and data mining KDD '14* (pp. 15–24). New York, NY, USA: ACM. https://doi.org/10.1145/2623330.2623703. URL: http://doi.acm.org/10.1145/2623330.2623703.

Dubois, A., Zagheni, E., Garimella, K., & Weber, I. (2018). *Studying migrant assimilation through facebook interests.* arXiv preprint arXiv:1801.09430.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *Proceedings of the 3rd innovations in theoretical computer science conference* (pp. 214–226). ACM.

Edelman, B., Luca, M., & Svirsky, D. (2017). Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics, 9*, 1–22.

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., et al. (2016). Computational personality recognition in social media. *User Modeling and User-adapted Interaction, 26*, 109–142. URL: https://doi.org/10.1007/s11257-016-9171-0. doi:10.1007/s11257-016-9171-0.

Fatehkia, M., Kashyap, R., & Weber, I. (2018). Using facebook ad data to track the global digital gender gap. *World Development, 107*, 189–209.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters, 27*, 861–874.

Felbo, B., Sundsoy, P. R., Pentland, A., Lehmann, S., & de Montjoye, Y. A. (2015). *Using deep learning to predict demographics from mobile phone metadata. CoRR, abs/1511.06660*. URL: http://arxiv.org/abs/1511.06660.

Feldman, G. (2018). *Personal values and moral foundations: Towards an integrated perspective by examining meaning, structure, and relations*.

Finnerty, A. N., Lepri, B., & Pianesi, F. (2016). Acquisition of personality. *Emotions and personality in personalized services* (pp. 81–99). Springer.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*, 1012.

Goel, S., Hofman, J., & Sirer, M. I. (2012). Who does what on the web: Studying web browsing behavior at scale. *International conference on weblogs and social media* (pp. 130–137). .

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from twitter. *Privacy, security, risk and trust (PASSAT) and 2011 IEEE third inernational conference on social computing (SocialCom), 2011 IEEE third international conference on* (pp. 149–156). IEEE.

Goldberg, L. R. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology, 59*, 1216.

Golder, S. A., & Macy, M. W. (2014a). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology, 40*, 129–152. URL: https://doi.org/10.1146/annurev-soc-071913-043145. doi:10.1146/annurev-soc-071913-043145. arXiv:https://doi.org/10.1146/annurev-soc-071913-043145.

Golder, S. A., & Macy, M. W. (2014b). Digital footprints: Opportunities and challenges for online social research. *Sociology, 40*, 129.

Gou, L., Zhou, M. X., & Yang, H. (2014). Knowme and shareme: Understanding automatically discovered personality traits from social media and user sharing preferences. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 955–964). ACM.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., et al. (2012). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology* (in press).

Graham, J., Nosek, B. a., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*, 366–385. https://doi.org/10.1037/a0021847.

Gray, G. (2014). A review of psychometric data analysis and applications in modelling of academic achievement in tertiary education. *Journal of Learning Analytics, 1*, 75–106.

Grunert, S. C., & Juhl, H. J. (1995). Values, environmental attitudes, and buying of organic foods. *Journal of Economic Psychology, 16*, 39–62.

Haidt, J. (2003). The moral emotions. *Handbook of Affective Sciences, 11*, 852–870.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*, 98–116.

Haidt, J., Graham, J., & Joseph, C. (2009). Above and below leftright: Ideological narratives and moral foundations. *Psychological Inquiry, 20*, 110–119. https://doi.org/10.1080/10478400903028573.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Dædalus, 133*, 55–66.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology, 143*, 29–36.

Hoover, J., Johnson, K., Boghrati, R., Graham, J., & Dehghani, M. (2018). Moral framing and charitable donation: Integrating exploratory social media analyses and confirmatory experimentation. *Collabra: Psychology, 4*.

Hu, J., Zeng, H.-J., Li, H., Niu, C., & Chen, Z. (2007). Demographic prediction based on user's browsing behavior. *Proceedings of the 16th international Conference on world wide web WWW '07* (pp. 151–160). New York, NY, USA: ACM. https://doi.org/10.1145/1242572.1242594. URL: http://doi.acm.org/10.1145/1242572.1242594.

Jackson, L. A., Ervin, K. S., Gardner, P. D., & Schmitt, N. (2001). Gender and the internet: Women communicating and men searching. *Sex Roles, 44*, 363–379.

Jansen, B. J., & Solomon, L. (2010). Gender demographic targeting in sponsored search. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 831–840). ACM.

Kalimeri, K. (2013). *Traits, states and situations: Automatic prediction of personality and situations from actual behavior*Ph.D. thesis. University of Trentohttps://doi.org/10.13140/2.1.2847.9043.

Kalimeri, K., Lepri, B., & Pianesi, F. (2013). Going beyond traits: Multimodal classification of personality states in the wild. *Proceedings of the 15th ACM on international Conference on multimodal interaction ICMI '13* (pp. 27–34). New York, NY, USA: ACM. https://doi.org/10.1145/2522848.2522878. URL: http://doi.acm.org/10.1145/2522848.2522878.

Kim, E., Iyer, R., Graham, J., Chang, Y.-H., & Maheswaran, R. (2013). Moral values from

simple game play. *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 56–64). Springer.
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*, 5802–5805.
Kosinski, M., Stillwell, D., Kohli, P., Bachrach, Y., & Graepel, T. (2012). Personality and website choice. *ACM web sciences 2012. ACM conference on web sciences*. URL: https://www.microsoft.com/en-us/research/publication/personality-and-website-choice/.
Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*, 8788–8790. URL: http://www.pnas.org/content/111/24/8788. doi:10.1073/pnas.1320040111. arXivhttp://www.pnas.org/content/111/24/8788.full.pdf.
Lepri, B., Subramanian, R., Kalimeri, K., Staiano, J., Pianesi, F., & Sebe, N. (2012). Connecting meeting behavior with extraversiona systematic study. *IEEE Transactions on Affective Computing, 3*, 443–455.
Li, J., & Fine, J. P. (2010). Weighted area under the receiver operating characteristic curve and its application to gene selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 59*, 673–692.
Lin, Y., Hoover, J., Dehghani, M., Mooijman, M., & Ji, H. (2017). *Acquiring background knowledge to improve moral value prediction.* arXiv preprint arXiv:1709.05467.
Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social media fingerprints of unemployment. *PloS One, 10*, 1–13. https://doi.org/10.1371/journal.pone.0128692. URL: https://doi.org/10.1371/journal.pone.0128692.
Low, M., & Wui, M. G. L. (2015). Moral foundations and attitudes towards the poor. *Current Psychology*, 1–11. URL: https://doi.org/10.1007/s12144-015-9333-y. doi:10.1007/s12144-015-9333-y.
Luhmann, M. (2017). Using big data to study subjective well-being. *Current Opinion in Behavioral Sciences, 18*, 28–33.
Malmi, E., & Weber, I. (2016). You are what apps you use: Demographic prediction based on user's apps. *ICWSM,* 635–638.
Mason, S. J., & Graham, N. E. (2002). Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation. *Quarterly Journal of the Royal Meteorological Society, 128*, 2145–2166. https://doi.org/10.1256/003590002320603584. URL: https://doi.org/10.1256/003590002320603584.
McAdams, D. P. (1995). What do we know when we know a person? *Journal of Personality, 63*, 365–396. https://doi.org/10.1111/j.1467-6494.1995.tb00500.x. URL: https://doi.org/10.1111/j.1467-6494.1995.tb00500.x.
McAdams, D., & Pals, J. (2006). A new big five: Fundamental principles for an integrative science of personality. *American Psychologist, 61*, 204–217. https://doi.org/10.1037/0003-066X.61.3.204.
Miles, A., & Vaisey, S. (2015). Morality and politics: Comparing alternate theories. *Social Science Research, 53*, 252–269. URL: http://www.sciencedirect.com/science/article/pii/S0049089X15001143https://doi.org/10.1016/j.ssresearch.2015.06.002.
van Miltenburg, E. (2016). *Stereotyping and bias in the flickr30k dataset.* arXiv preprint arXiv:1605.06083.
Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., & Rosenquist, J. N. (2011). Understanding the demographics of twitter users. *ICWSM, 11*, 25.
Mitchell, A., Gottfried, J., Kiley, J., & Matsa, K. E. (2014). *Political polarization and media habits.* Pew Research. URL: http://www.journalism.Ωorg/2014/10/21/political-polarization-media-habits.
de Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 48–55). Springer.
Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour,* 1.
Newport, F. (2012). *The Gallup Poll: Public opinion 2011.* Rowman & Littlefield Publishers.
Nistor, N., Dascalu, M., & Trausan-Matu, S. (2016). Newcomer integration in online knowledge communities: Exploring the role of dialogic textual complexity. *12th international conference on learning sciences* (pp. 914–917). .
Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401–421.
Parks-Leduc, L., Feldman, G., & Bardi, A. (2015). Personality traits and personal values: A meta-analysis. *Personality and Social Psychology Review, 19*, 3–29.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.
Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). Our twitter profiles, our selves: Predicting personality with twitter. *Privacy, security, risk and trust (PASSAT) and 2011 IEEE third inernational conference on social computing (SocialCom), 2011 IEEE third international conference on* (pp. 180–185). IEEE.
Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How peoples personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality, 55*, 98–111.
Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1–65. https://doi.org/10.1016/S0065-2601(08)60281-6.
Schwartz, S. H. (2003). A proposal for measuring value orientations across nations. *Questionnaire Package of the European Social Survey,* 259–290.
Schwartz, S. H. (2005). Basic human values: Their content and structure across countries. *Valores e comportamento nas organizações,* 21–55.
Schwartz, S. H. (2012). An overview of the schwartz theory of basic values. *Online Readings in Psychology and Culture, 2*, 11.
Scott, K., & Lewis, C. C. (2015). Using measurement-based care to enhance any treatment. *Cognitive and Behavioral Practice, 22*, 49–59.
Seneviratne, S., Seneviratne, A., Mohapatra, P., & Mahanti, A. (2015). Your installed apps reveal your gender and more!. *ACM SIGMOBILE - Mobile Computing and Communications Review, 18*, 55–61.
Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016). Fusing social media cues: Personality prediction from twitter and instagram. *Proceedings of the 25th international conference companion on world wide web* (pp. 107–108). International World Wide Web Conferences Steering Committee.
Soyez, K. (2012). How national cultural values affect pro-environmental consumer behavior. *International Marketing Review, 29*, 623–646.
Stark, R., & Iannaccone, L. R. (1997). Why the jehovah's witnesses grow so rapidly: A theoretical application. *Journal of Contemporary Religion, 12*, 133–157.
Torelli, C. J., & Kaikati, A. M. (2009). Values as predictors of judgments and behaviors: The role of abstract and concrete mindsets. *Journal of Personality and Social Psychology, 96*, 231.
Vinciarelli, A., & Mohammadi, G. (2014). A survey of personality computing. *IEEE Transactions Affective Computing, 5*, 273–291. https://doi.org/10.1109/TAFFC.2014.2330816. URL: https://doi.org/10.1109/TAFFC.2014.2330816.
Wang, P., Guo, J., Lan, Y., Xu, J., & Cheng, X. (2016). Your cart tells you: Inferring demographic attributes from purchase data. *Proceedings of the ninth ACM international Conference on web Search and data mining WSDM '16* (pp. 173–182). New York, NY, USA: ACM. https://doi.org/10.1145/2835776.2835783. URL: http://doi.acm.org/10.1145/2835776.2835783.
Weber, I., & Castillo, C. (2010). The demographics of web search. *Proceedings of the 33rd international ACM SIGIR Conference on Research and Development in information retrieval SIGIR '10* (pp. 523–530). New York, NY, USA: ACM. https://doi.org/10.1145/1835449.1835537. URL: http://doi.acm.org/10.1145/1835449.1835537.
Weber, René, Eden, Allison, Huskey, Richard, Mangus, J. Michael, & Falk, Emily (2015). Bridging media psychology and cognitive neuroscience. *Journal of Media Psychology, 27*(3), 146–156.
Weber, I., & Jaimes, A. (2011). Who uses web search for what: and how. *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 15–24). ACM.
Winterich, K. P., Zhang, Y., & Mittal, V. (2012). How political identity and charity positioning increase donations: Insights from moral foundations theory. *International Journal of Research in Marketing, 29*, 346–354. URL: http://www.sciencedirect.com/science/article/pii/S0167811612000638https://doi.org/10.1016/j.ijresmar.2012.05.002 Special Issue on Consumer Identities .
Wolsko, C., Ariceaga, H., & Seiden, J. (2016). Red, white, and blue enough to be green: Effects of moral framing on climate change attitudes and conservation behaviors. *Journal of Experimental Social Psychology, 65*, 7–19. URL: http://www.sciencedirect.com/science/article/pii/S0022103116301056https://doi.org/10.1016/j.jesp.2016.02.005.
Xu, R., Frey, R. M., Fleisch, E., & Ilic, A. (2016). Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study. *Computers in Human Behavior, 62*, 244–256.
Yeo, P. D. S. K., Cacciatore, M. A., & Scheufele, D. A. (2015). News selectivity and beyond: Motivated reasoning in a changing media environment. *Publizistik und gesellschaftliche Verantwortung* (pp. 83–104). Springer.
Ying, J. J. C., Chang, Y. J., Huang, C. M., & Tseng, V. S. (2012). Demographic prediction based on users mobile behaviors. *Mobile Data Challenge,* 1–6.
Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, 112*, 1036–1040. https://doi.org/10.1073/pnas.1418680112. URL: http://www.pnas.org/content/112/4/1036.abstracthttp://www.pnas.org/content/112/4/1036.full.pdf.
Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review, 43*, 721–734.
Zhao, X. W., Guo, Y., He, Y., Jiang, H., Wu, Y., & Li, X. (2014). We know what you want to buy: A demographic-based system for product recommendation on microblogs. *Proceedings of the 20th ACM SIGKDD international Conference on knowledge Discovery and data mining KDD '14* (pp. 1935–1944). New York, NY, USA: ACM. https://doi.org/10.1145/2623330.2623351. URL: http://doi.acm.org/10.1145/2623330.2623351.
Zhong, Y., Yuan, N. J., Zhong, W., Zhang, F., & Xie, X. (2015). You are where you go: Inferring demographic attributes from location check-ins. *Proceedings of the eighth ACM international Conference on web Search and data mining WSDM '15* (pp. 295–304). New York, NY, USA: ACM. https://doi.org/10.1145/2684822.2685287. URL: http://doi.acm.org/10.1145/2684822.2685287.
Zliobaite, I. (2015). *A survey on measuring indirect discrimination in machine learning.* arXiv preprint arXiv: 1511.00148.