Personality and Website Choice

Michal Kosinski **David Stillwell**
University of Cambridge
{mk583,ds617}@cam.ac.uk

Pushmeet Kohli **Yoram Bachrach** **Thore Graepel**
Microsoft Research
{pkohli,yobach,thoreg}@microsoft.com

ABSTRACT
We find that preference for websites, like preference for objects in the offline world, is influenced by personality. We combine personality profiles and website choices of more than 160.000 users and investigate whether different websites attract audience of different personality. Using two independent sources of website choices, we show that website audiences often have distinct personality profiles, that there is a psychologically meaningful relationship between personality and preferences related to website and website categories, and that results are stable across independent data sources. Our findings are useful for researchers interested in website content personalization, text search, search result optimization and online marketing.

Author Keywords
Personality, Browsing, Searching, Social Bookmarks,Website, Preference, Facebook, Online Marketing and Advertising

ACM Classification Keywords
H.3.5 Information Storage and Retrieval: On-line Information Services

General Terms
Measurement, Human Factors

INTRODUCTION
Decades of research in psychology suggests that behavior and preferences of individuals can be explained to a great extent by underlying psychological constructs or so called personality traits [1]. This observation is of great practical value, as it implies that the knowledge of an individual's personality enables prediction of behaviour and preferences across contexts and environments. Moreover, studies in personality assessment have revealed that responses to a relatively short personality questionnaire can allow prediction of human behavior in many aspects of life – including arriving on time and job performance [4], drug use [19], and infidelity [18]. It also is possible to assess personality by

inspecting traces of one's actions in the environment (or behavioral residues) such as records of keyboard and mouse use [14], individual's living spaces [11], personal website [15, 20], and Facebook profiles [9]. We hypothesize that website choices, like preferences for objects in the offline world, are influenced by personality. This implies that personality can be used to understand, describe, and potentially predict website choices of users as well as groups of users.

Related Work Several studies have analyzed the relationship between online preferences, browsing behavior and demographic characteristics of websites' audiences, including age, gender, occupation and education levels, mean income, and race (e.g. [3, 7, 12, 16, 21, 22]). To our knowledge, no attempts were made to relate psychological profiles to website choice, although the psychological literature provides some examples of personality inference based on other aspects of users' behavior in an online setting. For example, [15] and [20] assessed personality using the contents of personal websites, [8] studied the accuracy of personality judgements based on emails, while [2] showed that there is valid personality related information in users' email addresses.

DATA COLLECTION
We used two sources of users' website preferences: self-reports and social bookmarks. Personality information was measured using a standard personality questionnaire.

Participants and Personality scores were obtained using myPersonality[1], a Facebook application that offers its users personality assessments and feedback on their scores. Personality profiles were established using the standard 100 item International Personality Item Pool (IPIP) questionnaire [10] representing Five Factor Model of personality (FFM; [6]). The five personality dimensions are Openness to experience (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N). These dimensions have been shown to efficiently explain a substantial amount of variability in human preferences and behavior across different domains. They have also been observed to be genetically heritable, stable over time and consistent across genders, cultures, and races [13].

Self Reported Website Preferences were collected using a Website Preference Questionnaire (WPQ) designed for this study, that asked users (n=10,897) the frequency with which

[1]http://apps.facebook.com/mypersonality/

	Liked websites				Individuals	
	Min	**Max**	**Mean**	**SD**	**Min**	**Max**
Gender	.00	1.00	.71	.16	61% females	
Age	15.71	51.10	21.24	4.14	13.00	65.00
O	-1.00	1.22	.12	.26	-4.05	2.19
C	-1.21	.92	-.26	.23	-3.26	1.83
E	-1.29	1.02	-.03	.21	-3.13	2.11
A	-1.25	.91	-.10	.20	-3.61	2.24
N	-1.18	1.14	.03	.22	-2.85	2.46
#Users	20	22643	214.69	671.2	–	
#Likes		–			1	2,877

they visit 23 websites on a five point scale (from never to regularly). Websites were selected to be particularly informative about personality and be neither too popular nor too obscure. Extremely popular websites attract visitors of all personality types and thus are not informative, whereas obscure websites do not attract a reasonable fraction of users and thus lead to uncertain predictions.

The WPQ was offered in May 2010 to myPersonality users who had previously taken the IPIP questionnaire; 10,897 individual users completed it. On average, respondents reported that they had visited three of the websites in the questionnaire at least rarely (sd = 1.9). The maximum number of websites endorsed by a respondent was 13, while around 4% of the participants did not visit any of the websites.

Liked Websites Dataset relies on users' preferences expressed by Facebook *Likes*. Users *Like* websites to endorse them to their friends or bookmark them for future reference, and can do so by clicking the Like button directly on the website (an increasing number of websites offer such functionality) or by joining a website's fan page directly on Facebook. Facebook allows third party applications to access Likes stored on users' profiles with their consent. myPersonality respondents could opt-in to provide access to this data. Our sample consisted of more than 153,000 respondents and nearly 75,000 websites each of which was liked by at least 20 distinct users.

ANALYSIS AND RESULTS

Aggregated Website Audience Profiles

First we investigate personality profiles of the websites, established by computing the mean age, gender, and personality profiles of all users who reported to visit (WPQ dataset) or Liked (Liked Websites Dataset) each of the websites. Descriptive statistics of individual users and audience profiles based on Liked URLs are presented in Table 1.

The relationship between the number of Liked websites and individual traits leads to differences in the individual and aggregated profiles. For instance, females constitute 61% of the sample, but as they tend to Like more websites in general, the average website has 71% of females in its audience. To preserve the clarity of the results' presentation and allow for meaningful comparisons between aggregated

profiles, aggregated values were rescaled within each of the samples to zero mean. For instance, the aggregated values of O in the Liked dataset were decreased by its mean value (0.12) presented in Table 1.

An example of a website audience personality profile, deviantART.com, is presented in Figure 1. According to both sources of data, this website attracts an audience that tends to be liberal and artistic rather than conservative and traditional (i.e. with high Openness), spontaneous and flexible rather than well organized (i.e. with low Conscientiousness), shy and reserved rather than outgoing and active (i.e. with low Extraversion), and emotional rather than calm and relaxed (i.e. with high Neuroticism). Both personality theory and common intuition suggest that those results accurately represent the character of deviantART.com users in general – alternative art enthusiasts and artists. Importantly, the findings were consistent between the samples. As shown in Table 3) the average correlation between personality profiles estimated using both datasets was r = 0.83.
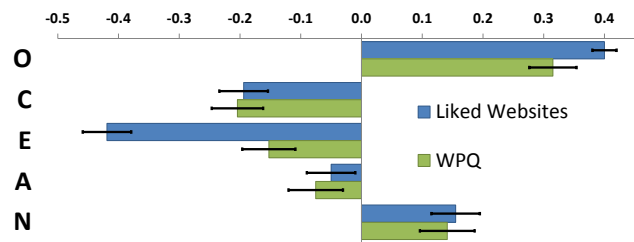
Figure 1. Audience personality profiles for deviantart.com estimated using two different data sources. The error bars show 95% confidence intervals.

Table 3. Correlation between personality profiles estimated using our datasets WPQ and Likes. Correlation coefficients were averaged using Fisher's z transformation.

cafepress.com	deviantart.com	tumblr.com	etsy.com	nba.com	snagajob.com	job.com	gamefaqs.com	digg.com	ancestry.com	myxer.com	pandora.com	barnesandnoble.com	foodnetwork.com	Average correlation
-0.06	0.98	0.78	0.98	0.94	0.85	0.77	0.59	0.86	0.64	0.92	0.84	0.34	0.83	0.83

Table 2 provides further evidence of the psychological validity of our results by presenting ten websites and eight categories characterized by the extreme mean scores for each of the personality traits. Websites were classified into one of the top two-levels of the Open Directory Project (ODP) document hierarchy [17] consisting 219 topical categories such as Arts/Movies, Business/Investing and Sports/Soccer using a methodology described by [5]. A logistic regression classifier with L2 regularization was trained using documents tagged with each category in a 2008 crawl of the ODP index. Using these classifiers, we tagged each liked URL in the dataset with the most likely ODP category. We then com-

Table 2. Websites and website categories with highest and lowest mean personality levels, estimated on the Likes dataset.

Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism
Liberal & Artistic	**Well Organized**	**Outgoing & Active**	**Cooperative**	**Emotional**
Arts.Animation	Reference.Education	Computers.Internet	Reference.Education	Recreation.Pets
Business.Marketing	Shopping.Electronics	Reference.Education	Computers.Internet	Recreation.Scouting
Business. Services	Shopping.Children	Science.Environment	Business.Logistics	Science.Physics
Arts.Photography	Reference.Dictionaries	Arts.Music	Health.Diseases	Sports.Hockey
modcloth.com	lww.com	clubzone.com	abebooks.com	cineplex.com
senate.gov	ecollege.com	ideeli.com	socialsecurity.gov	comparedby.us
boingboing.net	ecnext.com	thanksmucho.com	myrecipes.com	myprofilepimp.com
astrology-online.com	exct.net	discoveryeducation.com	bluemountain.com	barbie.com
gutenberg.org	education.com	list-manage.com	serialssolutions.com	yellowpages.ca
cafeastrology.com	kodak.com	trails.com	ecollege.com	biglots.com
...	...	...	...	...
gateway.com	candystand.com	lyricsty.com	localtribune.org	ncsu.edu
newegg.com	crunchyroll.com	fanfiction.net	funnyjunk.com	sheetmusicplus.com
fitnessmagazine.com	allthetests.com	behindthename.com	sciencebuddies.org	pitt.edu
ourtoolbar.com	bestuff.com	newworldencyclopedia.org	allthetests.com	highschoolsports.net
nhl.com	lyricsdepot.com	personalitypage.com	marvel.com	myrecipes.com
pier1.com	letmewatchthis.com	gaiaonline.com	supercheats.com	lww.com
Reference.Education	Health.Mental Health	Arts.Movies	Kids&Teens.Society	Arts.Photography
Arts.Television	Arts.Music	Shopping.Children	Health.Mental Health	Science.Maths
Sports.Soccer	Arts.Animation	Arts.Literature	Science.Physics	Business.Marketing
Shopping.Children	Arts.Literature	Arts.Comics	Recreation.Pets	Business.Logistics
Conservative	**Spontaneous**	**Shy & Reserved**	**Competitive**	**Calm & Relaxed**

puted the mean of each of the five personality traits for each ODP category.

Users of different personalities prefer different website categories and the differences are consistent with personality. For instance, Extroverted users frequent websites related to Music and Internet (the category that contains Facebook and Twitter), while Introverts prefer websites related to Comics, Literature, and Movies. Similarly, the most liberal, creative, and open to new experience audiences (with high Openness) are especially attracted to (1) modcloth.com, a mod-retro-indie clothing website, (2) boingboing.net, a blog on media, technology and popular culture, (3) astrology-online.com and cafeastrology.com, astrology websites, (4) gutenberg.org, a free e-book repository, (5) failblog.com, containing humorous media content, (6) fineartamerica.com, a fine art website, (7) 911tabs.com, a website specializing in guitar tabs, and (8) senate.gov, the website of the United States senate.

On the other end of the Openness scale, websites for which the user population is estimated to be most conservative and "conventional" include (1) dealspl.us and newegg.com, shopping deal websites, (2) a variety of health, fitness, recipe and style websites such as fda.gov, mydailymoment.com and fitnessmagazine.com, (3) doctorslounge.com, a website specializing in health and medical jobs, (4) gateway.com, which sells information technology products, (4) nhl.com, the website of the National Ice Hockey League in the United States, and (5) pier1.com which sells furniture and accessories.

Website Categories
The relationship between personality and website preferences can be also analysed on the level of website categories. Using classifiers described by [5], we classified each website in the Liked Websites Dataset into one of the top two-levels of the Open Directory Project (ODP) document hierarchy

[17]. A logistic regression classifier with L2 regularization was trained using documents tagged with each category in a 2008 crawl of the ODP index. Using these classifiers, we tagged each liked URL in the dataset with the most likely ODP category. We then computed the mean of each of the five personality traits for each ODP category.

Table 2 presents the categories with highest and lowest mean personality score for each personality trait. Different personalities prefer different categories and these differences are consistent with theory. For instance, Extroverted users frequent websites related to Music and Internet (the category that contains Facebook and Twitter), while Introverts prefer websites related to Comics, Literature, and Movies.

Audience similarity
One practical application of website audience personality profiles is in personalizing search results and suggesting websites of interest to users. Table 4 shows several websites that appear dissimilar on the surface and do not have much overlap in the audience[2] but have similar mean psychological profiles. This avenue for personalization would allow identifying other websites to promote to users based on the similarities in personality profile. We observe that Tumblr.com (a micro blogging platform), etsy.com (a marketplace of handmade craft), gaiaonline.com (advertised as a forum of young open minded people), fanboy.com (marketed as a website for intellectuals with imagination), and rainymood.com (providing sounds of rain to visitors) are frequented by audiences with similar mean personality: liberal, introverted, and rather emotional. Notably, the only website in this group that attracts relatively non-spontaneous and well organized users is etsy.com – a market place of hand-made crafts. Appar-

---
[2]The overlap in the audience between any two of the websites in Table 4 is lower than 2%

Table 4. Similarity of mean personality profiles of various art-related websites estimated using the Liked URL dataset. The columns labeled O through N represent the five personality traits, *freq* indicates the number of distinct users who liked each website. The column labeled *SEM* is the standard error of the mean, which was of similar magnitude for all of the five personality traits and is hence presented in a single column.

Domain	O	C	E	A	N	freq	SEM ($\pm$)	Pearson's correlation					
								1	2	3	4	5	6
(1) deviantART.com	.40	-.19	-.42	-.05	.16	3,154	.01-.02	1					
(2) Tumblr.com	.23	-.23	-.16	-.10	.22	639	.03	.89	1				
(3) Etsy.com	.41	.14	-.26	.07	.10	612	.03	.88	.59	1			
(4) GaiaOnline.com	.33	-.23	-.40	.00	.19	2,076	.02	.99	.91	.82	1		
(5) Fanboy.com	.36	-.27	-.44	-.04	.22	128	.07-.09	.99	.93	.81	1	1	
(6) RainyMood.com	.36	-.22	-.35	.03	.12	236	.06-.07	.99	.88	.84	.99	.98	1

ently, one needs a degree of Conscientiousness in addition to a general arty profile, to trade art.

DISCUSSION AND CONCLUSIONS

In this work we studied the relationship between website choice and personality.

There are several important implications of our work. First, it is valuable for website operators to realize that personality plays role in the website choice. Audience personality profile can be used to personalize website content or services, optimize product recommendations, and adjust marketing communication to fit what is known about the preferences characteristic to people of different personalities. Second, the personality of individuals can be *predicted* based on records of their browsing behaviour. This provides an alternative avenue for psychological research. Until today, most measurement in psychology has relied on self-reported questionnaires completed by relatively small numbers of participants. Like-logs based predictions could enlarge the scope of psychological assessment to unprecedented scale, and may improve the quality of results as it considers actual behavior in an increasingly natural environment rather than self-reported test answers. However, it also highlights the important issue of *user privacy*. While individual personality can often be quickly and accurately assessed by a skilled individual in an off-line setting, similar assessment in an online environment may breach users' expectations of privacy.

ACKNOWLEDGEMENTS

REFERENCES

1. Allport, G. W. The general and the unique in psychological science. *Journal of personality 30* (1962), 405–422.

2. Back, M., Schmukle, S., and Egloff, B. How extraverted is honey.bunny77@hotmail.de? inferring personality from e-mail addresses. *Journal of Research in Personality 42*, 4 (2008), 1116–1122.

3. Baglioni, M., Ferrara, U., Romei, A., Ruggieri, S., and Turini, F. Preprocessing and mining web log data for web personalization. In *Proceedings of the Congress of the Italian Association for Artificial Intelligence* (2003).

4. Barrick, M. R., and Mount, M. K. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology 44*, 1 (1991), 1–26.

5. Bennett, P. N., Svore, K., and Dumais, S. T. Classification-enhanced ranking. In *WWW* (2010).

6. Costa, P. T., and McCrae, R. R. *Revised NEO Personality Inventory (NEO PI-R): manual*. Hogrefe, 2006.

7. De Bock, K., and Van Den Poel, D. Predicting website audience demographics forweb advertising targeting using multi-website clickstream data. *Fundamenta Informaticae 98*, 1 (2010), 49–70.

8. Gill, A., Oberlander, J., and Austin, E. The perception of e-mail personality at zero-acquaintance. *Personality and Individual Differences 40* (2006), 497–507.

9. Golbeck, J., Cristina, R., and Karen, T. Predicting personality with social media. In *CHI* (2011).

10. Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., and Gough, H. G. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality 40*, 1 (2006), 84–96.

11. Gosling, S. D., Ko, S., Mannarelli, T., and Morris, M. E. A room with a cue: Personality judgments based on offices and bedrooms. *Journal of personality and social psychology 82*, 3 (2002), 379–398.

12. Hu, J., Zeng, H.-J., Li, H., Niu, C., and Chen, Z. Demographic Prediction Based on User's Browsing Behavior. In *WWW* (2007), 151–160.

13. John, O., Robins, R., and Pervin, L. *Handbook of Personality: Theory and Research (3rd Edition)*. Guilford Press, NY, 2008.

14. Khan, I. A., Brinkman, W., Fine, N., and Hierons, R. M. Measuring personality from keyboard and mouse use. In *ECCE* (2008).

15. Marcus, B., Machilek, F., and Schütz, A. Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology 90*, 6 (2006), 1014–1031.

16. Murray, D., and Durrell, K. Inferring demographics attributes of anonymous internet users. In *KDD* (1999).

17. Netscape Communication Corporation. Open directory project. http://dmoz.org/.

18. Orzeck, T., and Lung, E. Big-five personality differences of cheaters and non-cheaters. *Current Psychology 24* (2005), 274–287.

19. Roberts, B. W., Chernyshenko, O. S., Stark, S., and Goldberg, L. R. The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology 58*, 1 (2005), 103–139.

20. Vazire, S., and Gosling, S. D. e-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology 87* (2004), 123–132.

21. Weber, I., and Castillo, C. The demographics of web search. In *SIGIR* (2010).

22. Weber, I., and Jaimes, A. Who uses web search for what: and how. In *WSDM* (2011), 15–24.