

Erratum

to *Manifestations of user personality in website choice and
behaviour on online social networks*
originally published in Machine Learning Journal (MLJ), 2013

by Michal Kosinski, Yoram Bachrach, Pushmeet Kohli,
David Stillwell, Thore Graepel

May 7, 2014

An error was identified in Sections 3.4 and 3.5 of our paper. While the results presented there reliably represent the relationship between websites, website categories and the personality of an audience, we have mistakenly used an older dataset for this section. As we have used a newer dataset for remainder of the manuscript, we have re-run the calculations to prepare a new version of this section which uses the same dataset as the rest of the paper.

Below we include a corrected fragment of Section 3.4, corrected version of Section 3.5 and updated contents of Tables 2 and 3. We have also expanded Table 3 to include additional 2 traits: Satisfaction with Life and Intelligence.

We would also like to thank Filip Radlinski for his generous help and advice on the original paper and spotting the error corrected here. Also, we would like to thank Paul Bennett, who provided the web page classifier used to produce Table 3 in the original that is replaced here with the updated version.

3.5 Aggregated Website Audience Profiles

(...)

Table 2 provides further evidence of the psychological validity of our results by presenting the six websites with highest and lowest mean scores for each of the personality traits. For example we see that the most liberal, creative, and open to new experience audiences (with high Openness) are especially attracted to (1) takepart.com, a website with articles, commentary, video, and campaigns for social change, (2) phys.org, a portal with science-related news, (3) holytaco.com, containing humorous media content, (4) incaseofzombies.com, a portal that helps to prepare oneself for a zombie invasion and (5) dailyhoroscope.com (note that high Openness can lead to both, fascination with science, and extraordinary spirituality).

On the other end of the Openness scale, we see that websites for which the user population is estimated to be most conservative and “conventional” include (1) berecruited.com, a college sport teams recruitment website, (2) abc.com, a website of a predominantly conservative TV station (3) pinchingyourpennies.com offering shopping deals, (4) 4x4offroads.com, and (5) simplyshredded.com - a website about working out and fitness.

3.6 Website Categories

To better understand the relationship between personality and website preference, personality profiles are aggregated across website categories.

Websites in the Liked dataset were categorized based on the Open Directory Project¹ (ODP), a people-reviewed web directory categorized by language, subject and location. ODP categorizes websites in a hierarchical structure, with categories such as: "Arts.Television.Programs.Horror", where "Arts" represent the top level category, "Television" a subcategory of "Arts", and "Programs" a subcategory of "Television". Top-level categories "World" and "Regional" containing great numbers of non-US websites were removed from this analysis, as the sample used was predominantly composed of the US users.

The names of objects in the Liked dataset were matched with ODP website titles resulting in $n = 16,169$ matching websites. We used the second level of the ODP hierarchy, consisting of 396 categories such as "Education", "Social Sciences", and "Religion and Spirituality". Average audience personality profiles for each of the categories were computed in the following way. First, the number of websites Liked by each of the users across categories was computed. Number of websites Liked in a given category was used to estimate weighted means for each category and for each personality trait.

Table 3 presents the categories with highest and lowest mean personality score for each of the five personality traits. Results show that users of different personalities prefer different website categories. For instance, Extroverted users frequent websites related to Sports and Health, while Introverts prefer online dictionaries, Encyclopaedias, and websites related to models. Competitive individuals visit websites related to drugs, guns, and tile games, while agreeable users are interested in support groups, human resources and recreational camps.

¹<http://dmoz.org/>

Openness	
Liberal, Artistic	Conservative
TakePart.com	beRecruited.com
PhysOrg.com	ABC.com
HolyTaco.com	PinchingYourPennies.com
InCaseOfZombies.com	4x4OffRoads.com
DailyHoroscope.com	Simplyshredded.com

Conscientiousness	
Organized	Spontaneous
Guns.com	HolyTaco.com
PoliceOne.com	399animeshop.com
Idealist.org	AZLyrics.com
Scrubshopper.com	AlbinoBlacksheep.com
KMOV.com	JrockRevolution.com
FreedomProject.org	nerf-this.com

Extroversion	
Outgoing, Active	Shy, Reserved
FragranceX.com	PsychCentral.com
TotalFratMove.com	FanFiction.Net
clubflyers.com	mamacreepy.com
elnutrition.com	ShanaLogic.com
MaxPreps.com	Congress.org
LilWayneHQ.com	Fark.com

Agreeableness	
Cooperative	Competitive
SelfGrowth.com	Tattoofunny.com
NaturalWellbeing.com	AngryYoungandPoor.com
Gracehope.com	WEBN.com
GenealogyBuff.com	FrightCatalog.com
Crosswalk.com	AZLyrics.com
Proliferocks.com	TheYBF.com

Neuroticism	
Emotional	Calm, Relaxed
Tattoofunny.com	LiveMixtapes.com
DogTipper.com	BlackFolkzThink.com
RandomCreepyGuy.com	Guns.com
VampireAwareness.com	NUCLacrosse.com
PrankDialer.com	CBSSports.com
iDyeMyHair.com	Superpages.com

Table 2: Websites with highest and lowest mean personality for each of the five personality traits, estimated on the Likes dataset.

Table 3: Categories of websites characterized by the highest and lowest levels of aggregated personality traits, satisfaction with life and intelligence.

Openness	
Liberal, Artistic	Conservative
Computers.Home Automation	Sports.Bowling
News.Analysis and Opinion	Sports.Fantasy
Reference.Thesauri	Business.Accounting
Arts.Digital	Business.Human Resources
Computers.Intranet	Shopping.Tools
Science.News and Media	Home.Rural Living
Computers.Speech Technology	Health.Search Engines
Arts.Costumes	Sports.Rodeo
News.Alternative	Shopping.Recreation
Recreation.Models	Home.Home Improvement

Conscientiousness	
Organized	Spontaneous
Sports.Bowling	Sports.Airsoft
Business.Accounting	Computers.Consultants
Business.Information Technology	Recreation.Kites
Computers.Data Communications	Games.Trading Card Games
Sports.People	Reference.Ask an Expert
Business.Human Resources	Computers.Usenet
Shopping.Tools	Games.Conventions
Sports.Fencing	Computers.Speech Technology
Reference.Biography	Sports.Laser Games
Computers.Algorithms	Sports.Skateboarding

Extroversion	
Outgoing, Active	Shy, Reserved
Sports.Youth and High School	Reference.Encyclopaedias
Games.Party Games	Recreation.Models
Sports.Rodeo	Health.Search Engines
Shopping.Weddings	Business.Information Technology
Science.Software	Reference.Dictionaries
Sports.Golf	Games.Conventions
Sports.People	Reference.Bibliography
Business.Textiles and Nonwovens	Computers.Consultants
Sports.Fantasy	Reference.Ask an Expert
Health.Beauty	Business.Human Resources

Agreeableness	
Cooperative	Competitive
Recreation.Camps	Business.International Business and Trade
News.Museums and Archives	Computers.Home Automation
Business.Human Resources	Sports.College and University
Science.Software	Computers.Consultants
Health.Support Groups	Society.Sexuality
Sports.Laser Games	Recreation.Drugs
Shopping.Recreation	Recreation.Guns
News.Weather	Reference.Encyclopaedias
Health.Women's Health	Computers.Artificial Life
Health.Weight Loss	Games.Tile Games
Neuroticism	
Emotional	Calm, Relaxed
Reference.Biography	Sports.Youth and High School
Computers.Virtual Reality	Sports.People
Computers.Intranet	Science.Software
Society.Advice	Business.Small Business
Society.Transgendered	Sports.Cycling
Health.Child Health	Sports.Strength Sports
Games.Conventions	Business.Human Resources
Health.Search Engines	Sports.Boxing
Arts.Digital	Sports.Basketball
Recreation.Kites	Sports.Bowling
Satisfaction with Life	
High	Low
Computers.Bulletin Board Systems	Shopping.Office Products
Sports.Cycling	Reference.Quotations
Shopping.Photography	Science.Education Resources
Sports.Gymnastics	Computers.Usernet
Arts.Education	Computers.Robotics
Science.Anomalies and Alternative Science	Computers.Virtual Reality
Health.Support Groups	Recreation.Drugs
Shopping.Flowers	Sports.Fantasy
Business.Textiles and Nonwovens	Society.Subcultures
Shopping.Niche	Society.Crime
Intelligence	
High	Low
Health.Child Health	Computers.Open Source
Reference.Libraries	Society.Organizations
News.Media Industry	Recreation.Antiques
Arts.Architecture	Society.Advice
Recreation.Climbing	Shopping.Ethnic and Regional
Sports.Flying Discs	Arts.Education
Games.Coin-Op	Business.Opportunities
Society.History	Society.Ethnicity
Games.Puzzles	Shopping.Auctions
Society.Politics	Home.Home-making