Bad News for Democracy? Browsing Behavior Can Signal Political Attitudes

Nora Kirkizh
Technical University of Munich
GESIS – Leibniz Institute for the Social Sciences
eleonora.kirkiza@gesis.org

Roberto Ulloa
GESIS – Leibniz Institute for the Social Sciences
roberto.ulloa@gesis.org

Sebastian Stier
GESIS – Leibniz Institute for the Social Sciences
sebastian.stier@gesis.org

Jürgen Pfeffer
Technical University of Munich
juergen.pfeffer@tum.de

ABSTRACT

Browsing behavior and visits to website such as donation platforms, social media, streaming service providers, and even online gambling can reflect individuals' life-style, while, as research shows, life-style itself is a predictor of individuals' political issue preferences and attitudes. In this paper, we linked 19,000,000 website visits generated from web tracking of 1,000 users in Germany to self-reported political attitudes to investigate whether website choices can predict individuals' political orientations? Our machine learning model was best at signaling individuals' interest in politics, Euroscepticism, and democratic attitudes. Browsing behavior of individuals with populist inclinations as well as attitudes towards immigrants were moderately informative for the model. The web browsing histories could not predict individuals perceptions of climate change and level of trust to public institutions. By showing the potential of web browsing data to reveal individuals' political orientations such as authoritarian or anti-immigrant attitudes, our cross-validated evidence has normative implications for political campaigns, online privacy, and democracy in general. This study also makes methodological contribution to the literature on classifying political orientations with online behavioral data.

KEYWORDS

browsing behavior, political attitudes, democracy, machine learning, computational social science

ACM Reference Format:
Nora Kirkizh, Roberto Ulloa, Sebastian Stier, and Jürgen Pfeffer. 2020. Bad News for Democracy? Browsing Behavior Can Signal Political Attitudes. In *Working paper 2021 (XXX '21), Month 00–00, 2021, City, Country*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

For political parties, individuals who have particular attitudes towards specific policies but did not vote are valuable to target with political messages. Since people often prefer to read and hear information that confirms their existing believes [34] and if political actors know what people believe in by monitoring their web-traffic, they can much better create targeted adds that comply with voters' believe system, and therefore, increase the manipulative power of advertising. Research has already shown that right-leaning websites are tracking their audience more extensively than left-leaning websites [1], while individuals prefer to keep their political views or attitudes private.[1,2] Nevertheless, budgets for political advertising are growing every electoral cycle. Availability of digital trace data and modeling to infer people's stands on issues has inflated political campaigns' spending for the US mid term elections in 2020 to $5 billion, the largest in history even compared to the presidential election in 2016. Political campaigns use the service of commercial data brokers who track users and match their purchasing histories with voter registration information.[3] As a response to increasing interest of political parties to digital marketing, which often followed by misinformation and polarization, prior the US presidential election in 2020, Google and Facebook significantly altered their political ads policies to avoid displaying political ads with 'demonstrably' false information and adjusted their microtargeting policies[4,5] while Twitter entirely banned political ads from its platform.[6]

However, political parties or other interest groups may use alternative strategies to reach their electorate by means of targeting individuals based on their subscriptions to YouTube channels or Facebook public pages. Research has shown that based on Facebook likes models can predict if a person is Democrat or Republican [29], or even vote choice [7]. Visits to untrustworthy news websites are related to people's populist attitudes [3] and party affiliation [21]. In this study, we argue that in the time of rising political polarization, voters of specific political views might be identifiable based on their web browsing behavior. Building on previous work that found a strong relation between lifestyles and political orientations [11], we test this argument based on three-month web browsing histories from 1,000 individuals living in Germany. We examine if individuals' political attitudes are identifiable from *general* website choices,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
XXX '21, April 19–23, 2021, City, Country
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$00.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

[1] https://policies.google.com/privacy/key-terms#toc-terms-sensitive-info
[2] https://www.propublica.org/article/how-companies-have-assembled-political-profiles-for-millions-of-internet-us
[3] https://www.cnet.com/news/how-your-personal-data-is-used-to-create-a-perfect-midterm-election-ad/
[4] https://about.fb.com/news/2020/01/political-ads/
[5] https://blog.google/technology/ads/update-our-political-ads-policy
[6] https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html

not just their news-related behavior that is the focus of most previous research. Overall, we analyze 19,026,887 website visits and the associated URLs. Prior to the tracking, we asked our panelists to answer survey questions measuring their attitudes towards *(1) immigration, (2) democracy, (3) climate change policies, (4) trust in public institutions, and (5) populist attitudes.* — policy dimensions that reflect manifestos of major political parties in Germany, and parties' Facebook pages [4]. We also measured panelists' political ideology, interest in politics, and attitudes towards the European Union (EU) integration. We chose Germany because the country represents a state, whose political landscape has been experiencing a rise of anti-system right-wing populist parties, which are challenging the norms of democracy. In the challenging for democracy political context, the studies of potential weaknesses in voters' privacy protection on the Internet become even more important.

We found that from web browsing histories, simple machine learning models are able to detect signals about individuals' attitudes towards democracy, the EU, and interest in politics. Our model was also able to signal perceptions of immigration and crime, and partially populist attitudes. Attitudes towards climate change and trust in public institutions, however, are not distinguishable from zero. Consistent with our theory on the connection between life-style related websites and political orientation, our predictions improve significantly when we apply visit duration thresholds to keep only websites, where people spend considerable amount of time.

The contribution of this paper lies in several dimensions. First, we find that sensitive information about' political attitudes can be revealed from individuals' browsing histories. Hence, further development of digital privacy policies should consider that this knowledge could be potentially used by ads distributors like Google [12] and their customers such as political actors to target voters without asking them about their attitudes towards specific policies directly. Despite the generally small persuasive effects of political advertising [9], more targeted messages might exacerbate attitude polarization with regard to specific issues. Coupled with the fact that these processes are hidden to citizens and the general public, the findings are troubling for democracy. Second, predictability of attitudes towards democracy from visits to life-style related websites suggests that political polarization goes beyond political news consumption. Hence, tracking differences in website preferences between democratic and authoritarian individuals over time can potentially be used as a measure of dynamic polarization at scale. Third, from a methodological perspective, in this paper we introduce an approach on how to draw classification from domains that are relevant for social science research questions, and that measurement of peoples' attitudes based on web tracking data is compatible to self-reported attitudes. Finally, the paper contributes to a growing body of research showing that digital traces can be used for inferences about people's political and social attitudes [7, 25, 40, 41] by showing that web browsing histories can be stronger predictors of individuals' political orientations than social media because website visits reflect peoples' life-style.

2 RELATED WORK

A number of studies have used digital footprints, — Facebook likes, smartphone or website logs — to identify peoples' personality traits and other attributes [35, 39]. Here we provide an overview of the recent literature on using digital trace data to learn about peoples' political tastes, and major studies on association of political attitudes with individuals' lifestyle.

2.1 Predicting Individuals' Attributes from Online Behavior Patterns

2.1.1 Social media. Social media data can help to find individuals' perfect job match [25], to measure emotions [32] or personality traits. Patterns of Facebook usage such as number of friends, followed groups or published photos can predict big five personality traits like openness, extraversion, and agreeableness among others [6, 13, 19]. Facebook likes can predict individuals' attributes and life-style characteristics including alcohol drinking, religion, and even relationship types [29].

2.1.2 Web browsing logs. However, social media data might display socially desirable picture of individuals because people do not want the public to know about their true interests. Self-reported website choices, however, might be more revealing and consistent with peoples' personalities. Websites choices can predict personality traits [30]. For instance, high level of emotionality correlates with visits to websites related to sports, while calm personality is associated with photography. Observational data with websites extracted from Facebook likes showed that self-reported website choices are robust in predicting peoples' personalities [28]. Collected from smartphones data can reveal personalities as well. For instance, phone activity correlates with extraversion, and music apps with openness [41]. Individuals' browsing behavior is also capable to predict demographics features. Age and gender are the most predictable from browsing behavior [24, 29] setting a benchmark for other attributes.

2.1.3 Digital trace data and political attributes. The review of related literature shows that researchers has been focusing on personality traits and demographic attributes of individuals while political orientations remain understudied. Some studies include politics-related attributes into their prediction models as a secondary variables of interest. For example, using Facebook likes, machine learning models can predict individuals' vote choice [7], whether a user is a Democrat or Republican [29], while visits to untrustworthy websites are associated with populist attitudes [3] and party affiliation [21]. In this paper, we offer a direct prediction of broader set of individuals' political attitudes based on their browsing histories.

2.2 Political Attitudes, Political Behavior and Lifestyle

Political attitudes are based on a set of beliefs with which individuals approach political issues. Attitudes are much more stable than opinions, and they are hard to shift [33], which makes them a good predictor of individuals' political ideology or even vote choice. In fact, political attitudes often become a source of decision-making

and political behavior like voting [10]. For instance, in Europe, anti-immigration attitudes are a predictor of voting for radical-right parties [38], attitudes towards climate change policies are associated with voting for green parties [23]. Research also demonstrates that people who have never voted before often rely on their political attitudes when voting [2]. Hence, political parties or candidates, when drafting their policy platforms, often appeal to voters' political attitudes. This makes identification of political attitudes especially desirable for political parties or candidates while individuals would prefer their attitudes towards, for instance immigration or foreign workers, remain private.

In this paper we attempt to predict political attitudes from individuals' web browsing behavior. However, can website choices signify about political attitudes and what are the mechanisms? We rely on existing literature, which demonstrates that people with distinct lifestyle preferences also have specific political attitudes and personality [11, 14, 18, 40].[7] For example, immigration attitudes can be linked to lack of compassion or traveling [26], negative attitudes towards climate change policies and support of authoritarian policies to deal with crime are associated with low level of generalized trust [17, 36], populist inclinations can be the result of an individual being in debt or loosing job [42], or low level of agreeableness . Assuming that offline behavior is generally reflected in online behavior, each of the suggested associations might be derived from visits to specific websites. For instance, hotel and flights booking platforms can be a proxy of cosmopolitan vs. nationalist orientation, whereas gambling and charity websites can be a proxy of individuals with high level of empathy. However, people choose websites according to their predispositions. We build on the theory of selective exposure [16, 37], implying that individuals choose political information based on their existing world view. We apply this theory to browsing behavior assuming that individuals choose websites based on their preferences and attitudes.

3 DATA AND MEASUREMENT

In this paper, we use two types of data: web browsing logs and online survey responses. The data was collected with approval of the Oxford Internet Institute's Departmental Research Ethics Committee at the University of Oxford (Reference Number SSH IREC 18 004).

3.1 Web Tracking

We acquire web browsing histories of respondents from an online access panel maintained by Netquest, a market research company. Respondents enter the panel only via invitation, which they receive via e-mail under user consent. The panel consists of respondents who agreed to install plugins tracking their web browsing behavior on desktop computers. If they consent to participate, panelists receive additional incentives in case they do not stop the tracking for longer than seven days. Participants have the possibility to pause the tracking tools at any time. The tracking tools would then be interrupted for 15 minutes. Personally identifiable information is algorithmically anonymized by Netquest. We utilize web browsing histories from 1,003 panelists living Germany. The tracking period is between mid-March and mid-June, 2019. The dataset includes anonymized IDs, visited URLs, domains, and time spent on the web page. Overall, the dataset comprises 19,026,887 URLs, and 96,093 unique domains. Figure 1 and Table 1 shows the distribution for mean duration and number of visits per panelist and other main summary statistics.

We also tested if our data represent behavior of the general population. Since our panelists were aware of the tracking, they might have altered their behavior. For instance, they might have started to visit more news websites to learn more about political issues or the other way around; in short, respondents who are being tracked might be more careful in revealing their political interests and inclinations. We correlated visits made by our panel to media websites with ground truth data on the visits of the German general population recorded by the "Informationsgemeinschaft zur Feststellung der Verbreitung von Werbeträgern e.V." (IVW). The

[7] https://www.pewresearch.org/politics/2014/06/12/section-3-political-polarization-and-personal-life/

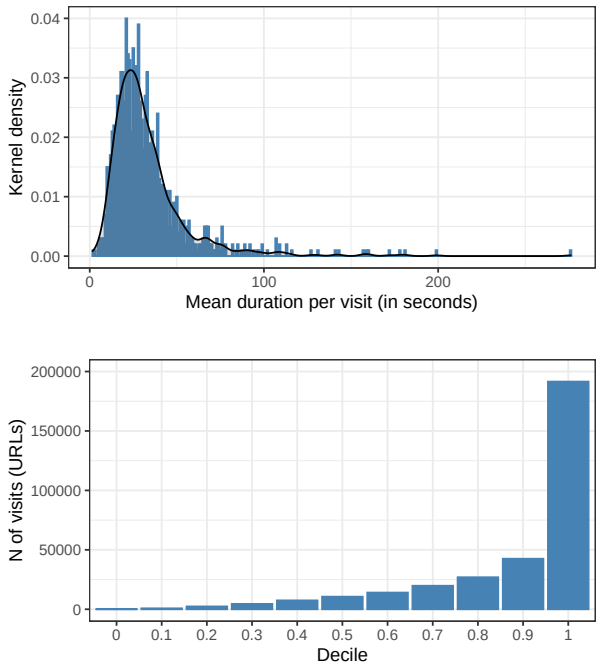

Figure 1: Kernel distribution for mean duration and number of visits made by panelist in each decile.

Table 1: Descriptive statistics of web tracking variables. There were 1,000 panelists, 19,026,887 URLs, and 96,093 unique domains.

Statistic	N	Mean	St. Dev.	Min	Max
N visited URLs	1,000	18,080.07	23,864.05	53	191,526
Duration (sec.)	1,000	469,971.90	539,768.10	569	4,660,753
N unique domains	1,000	362.04	328.97	9	2,279
μ visits per unique domain	1,000	43.36	37.30	3.28	376.76
μ duration per unique domain (sec.)	1,000	1,373.64	1,955.56	56.90	44,116.23
μ duration per URL (sec.)	1,000	33.34	23.58	1.62	276.12

Table 2: Descriptive statistics of predicted survey-based variables that are measuring individuals' political attitudes.

Statistic	N	Mean	St. Dev.	Min	Max
Interest in politics	1,019	2.86	0.86	1.00	4.00
Trust in parliament	1,019	3.24	1.17	1.00	5.00
Trust in the police	1,020	2.54	1.10	1.00	5.00
Left-light ideology	1,003	5.65	2.01	1	11
EU integration	871	6.79	2.92	1.00	11.00
Income redistribution	871	3.30	1.14	1.00	5.00
Big business and the people	869	3.72	1.03	1.00	5.00
Social benefits and laziness	870	2.79	1.15	1.00	5.00
Islam	940	3.46	1.31	1.00	5.00
Immigrants and jobs	1,020	2.83	0.91	1.00	4.00
Immigrants and crime	1,020	2.04	0.89	1.00	4.00
Climate change and humans	869	3.49	0.89	1.00	5.00
Free elections	866	9.60	2.19	1.00	11.00
People obey their rulers	866	3.96	2.92	1.00	11.00
Strong leader	870	2.04	0.96	1.00	4.00
Democratic political system	868	3.39	0.69	1.00	4.00
Satisfaction with democracy	1,019	2.63	0.80	1.00	4.00
Gender	1,018	1.51	0.50	1.00	2.00
Age	1,020	46.92	14.09	16.00	84.00
Education	942	1.97	0.70	1.00	3.00
Income	1,015	4.62	2.68	1.00	10.00

Political attitudes scales: interest in politics (1 - not at all, 4 - quite interest); trust in parliament and police (1 - not at all, 5 - a great deal); political left-right ideology (1 - left, 11 - right); EU integration (1 - gone too far, 11 - should be pushed further); big business takes advantage of ordinary people and social benefits make people lazy, and Islam promotes violence more than other religions, immigrants take jobs away from German people, immigrants make crime problems worse (1 - strongly disagree, 5 - strongly agree); climate change is caused by natural processes, human activity, or both (1- natural processes, 5 - human activity); the following things are essential characteristic of democracy (1 - not essential for democracy, 11 - essential for democracy); having a strong leader who does not have to bother with Parliament and elections (1 - very good, 5 - very bad); satisfaction with democracy (1 - not at all, 4 - very satisfied). *Demographics:* gender (1 - female, 2 - male), education (1 - low, 2 - middle, 3 - higher); income in EUR (1 - less than 100, 5 - between 2,200 and 2,600, 10 - more than 5,390).

correlation between the ranking of news sites visited by our German tracking panelists and the IVW data is strong (ρ = 0.73). These evaluations give us confidence that our tracking data provides a fairly accurate representation of visited websites by internet users in Germany.

In addition, following [21] we evaluate to what extent privacy attitudes of tracking panelists diverge from panelists who participate in surveys but do not have tracking tools installed. To identify a potential "opt in bias", we implemented the same privacy attitude battery in a sample of German participants drawn from the regular online access panel of Netquest without web tracking. In total, 1,00o participants were sampled according to German population margins for gender, age and education. Respondents were presented the following statements and asked about their (dis)agreement on a five-point scale: Personalized advertising makes me afraid; I am concerned about how much data there is about me on the Internet; and My privacy on the Internet does not matter to me. Figure 8 in the appendix shows that there are minor differences in the privacy attitudes of online panelists with and without web tracking technology installed, which brings the results of this paper closer to generalizability.

3.2 Survey

To combine digital trace data with panelists' responses [5], we surveyed the study participants parallel to the web tracking. To

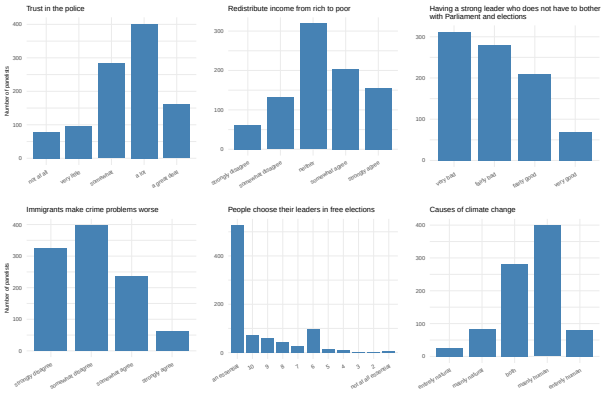

Figure 2: Distribution of selected survey items measuring political attitudes from each topic: trust in institutions (e.g. the police), populist attitudes (redistribute incomes from rich to poor), democracy (demand in strong leader, and choosing the leader in free elections), immigration (immigrants and crime), climate change.

infer their political attitudes, panelists were asked a set of questions related to diverse political issues: *immigration*, *democracy*, *climate change*, *trust to public institutions*, *populist attitudes* as well as *political ideology* on the left-right scale, political interest and attitudes towards the EU integration. The responses are placed on Likert (from strongly disagree to strongly agree) or 1-11 scales among others. In addition, we asked demographic questions such as age, gender, education, and income. We used the question wording established by prominent annual survey panels such as Eurobarometer, European Social Survey, and World Values Survey. Table 2 summarizes scales for each survey item, and questions wording in the note of the table. Figure 2 shows the distribution of answers for selected survey items. Differences in the number of questions per dimension represent the heterogeneous underlying concept structures. For instance, the dimension on immigration requires to ask about attitudes towards immigrants and their association with crime as well as jobs, while climate change attitudes can be measured with only one question, aiming to identify acceptance or rejection of climate change.

4 ANALYSIS AND RESULTS

Before running the regression analysis, we first check if the browsing behavior of panelists from different ideological groups, an important confounder of political activity, are balanced and hence not driven by how active they are online. The attitudes that we are focusing in this study potentially can be explained by a latent variable *political ideology*. For instance, people with a left-wing political ideology are more likely to support climate change policies than people from the right wing political ideology because they believe that climate change is caused by human activity [27]. Attitudes towards immigrants [22] and other policies would be in the same direction. We use the left-right political ideology spectrum to test the balance in browsing behavior between individuals from the left and the right. Figure 3 illustrates the distribution of visits and mean

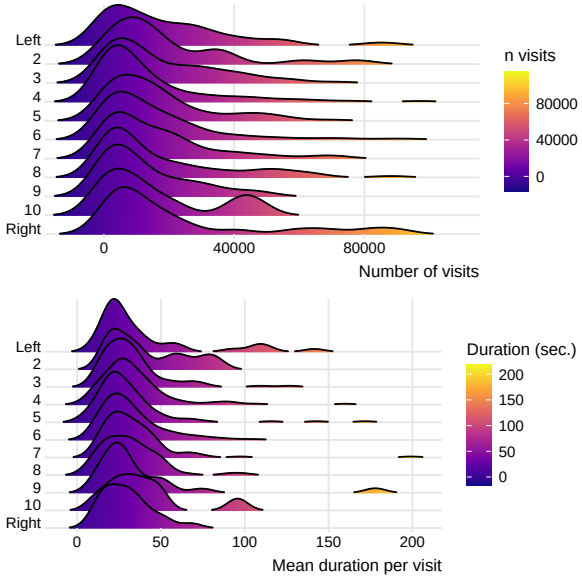

Figure 3: Distribution of number of visits and mean duration per visit across panelists' political ideology.

duration by the political ideology of the panelists. Both groups' browsing behavior, the left and the right, appeared to be balanced hence the degree of activity should not drive the prediction in our regression models.

In the analysis we use domains as a unit of observation that will be predicting political attitudes. We choose domains instead of URLs because similar domains appear in the data set much more often across individuals' browsing histories than URLs, i.e., the use of URL would produce a very sparse representation. We use two different thresholds for further analysis of the web tracking data: domain visits with duration one and five minutes. These thresholds are important because based on theory, we are interested in website visits that can signify the lifestyle of an individual. Domains where individuals spent more than five minutes are more likely deliberately visited. We choose the one minute cutoff as an additional robustness test. After we removed "short" domain visits, where individuals spent less than five minutes, the number of panelists drops to 803 and visits drops to 204,774 URLs and 7,999 unique domains; for one minute, the data generates 1,632,769 visits (36,074 unique domains) with 1,003 respondents.

4.1 Method

To predict political attitudes, we follow the methodology used by Kosinski et al. [31] to predict personality traits with Singular Value Decomposition (SVD, [20]) based on likes to Facebook pages. We assume that analyzing web tracking data poses a similar scenario because (1) both of them can be considered digital traces, and (2) a visit to a website indicates interest in a similar way that a like to a Facebook page does. As opposed to Facebook likes, however, visits to websites are not binary and also contain a duration. Below, we explain the adjustments taken to deal with these two features.

4.1.1 Pre-processing. We represent the browsing data in a matrix, where rows are individuals (respondents) and columns are unique domains. Each cell of the matrix contains the number of visits that a participant made to a particular domain. Additionally to the initial 5min/1min cutoffs explained before, we trim the matrix by excluding domains that were visited by less than 10 respondents and respondents that visited less than 10 domains. This procedure removes rare domains and occasional participants that do not contribute a lot of information to the model considering that our relatively small sample size would not represent these visits properly.

4.1.2 Decreasing the number of dimensions. Since our respondent-domain matrix is sparse, we apply Singular Value Decomposition (SVD), a popular dimensionality reduction approach based on eigen-decomposition [20]. Given the sparsity of our matrix, SVD also helps representing the data in a more compact dimensional space. We preferred SVD over other methods as it is computationally efficient and suitable for our exploration. In contrast, methods like LDA (also suggested in [31]) offered small improvements to predictions at very high computational costs.

SVD "dimensions" (i.e., left and singular values) can be interpreted directly if they are properly rotated (using varimax rotation). Each dimension then will be represented by all domains and a "coefficient" for each them that can be used to sort the domains according to their importance inside each dimension, also taking consideration that such coefficients could be negative values. Thus, SVD allows to identify domains with large explanatory power or domains that are the most relevant for the prediction of political attitudes.

One caveat with any dimensionality reduction approach is the selection of k; using a scree plot of the explained variance, a good k is often selected by visually identifying the "elbow", i.e., the point after which adding more dimensions does not decrease the explained variance. However, given the exploratory nature of this study and the difficulty of the task, we identify the best k by training and testing models for each k in a comprehensive set of values ranging from 5 to 500: $\{5\} \cup \{10i : 1 <= i <= 20\} \cup \{25i : 8 < i <= 12\} \cup \{50i : 6 < i <= 10\}; k \in \mathbb{Z}$.

4.1.3 Regression model. For each k, we fit generalized linear regression models using the k dimensions of SVD as independent variables for each of our dependent variables in Table 2. We can write the regression models formally as follows:

$$Political\_Attitude = \alpha + \beta_1 K_1 + \beta_2 K_2 + ... + \beta_n K_n, \quad (1)$$

where $Political\_Attitude$ represents survey-based attitudes, α is the intercept, and β is a regression coefficient for every SVD component K. Overall, we have 16 questions measuring political attitudes, which we include into the model one by one.

For each model, we measured its ability to predict new cases using 10-fold cross-validation, i.e., we ran 10 repetitions of the cross-validation process while randomizing the selection of the 10 folds each time. Each 10-fold cross-validation repetition splits the data in 10 fixed parts and uses 9 to predict the 10th. Repeating the cross-validation ensures that the prediction was not an artifact of the selection of the 10 fixed parts. We only considered a dependent variable to be "predictable" for a given k, if the prediction was

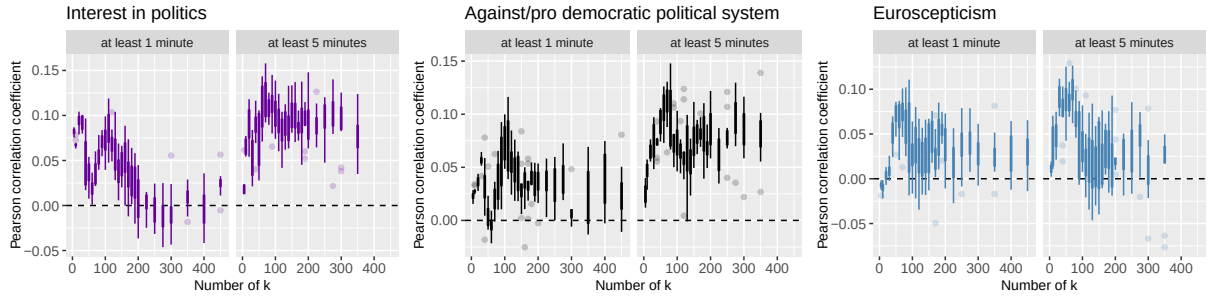

Figure 4: Attitudes with stable significant correlations in cross-validations as a function of the number of SVD components k by at least one and five minutes duration per visit cutoff.

statistically significant ($p < 0.05$) in all the cases in which the cross validation was repeated (i.e., 10 out of 10). The statistical significance was calculated using Pearson correlations between the predictions and the dependent variables on the test splits.

4.2 Regression results

The regression analysis focuses on five dimensions of political attitudes: immigration, democracy, climate change policies, populism, trust as well as interest in politics in general. We first report the results for the most predictable and stable attitudes and continue with the less predictable and unstable attitudes.

4.2.1 Political attitudes with significant predictions. In Figure 4 we report Pearson correlation coefficients by every SVD component k and averaged by the cross-validation folds for political attitudes with best predictions. The estimations show that web browsing behavioral patterns can predict individuals interest in politics, attitudes towards democratic political system, and Euroscepticism. The coefficients become much larger when we apply visit duration thresholds: at least one and five minutes. Interest in politics has median $r = 0.09$, attitude towards democratic political system — 0.07, and $r = 0.04$ for Euroscepticism, all obtained with $k = 125$. Best predictions are $r = 0.15$, 0.13, and 0.15 respectively, with $k = 70$, 80, and 60. The level of accuracy is compatible even with age, which has median $r = 0.08$ also with $k = 125$, and best prediction $r = 0.22$ with $k = 50$ (age usually serves as a benchmark in the literature, because it is a continuous variable, as opposed to a scale, and it is often reported accurately in surveys). Predictability of individuals interested in politics implies that browsing behavior can potentially reveal people who are a particularly relevant target for politicians or political parties. In combination with authoritarian inclinations and Euroscepticism, interest in politics becomes even more valuable target for political campaigns since political parties often aim individuals, who can be potential voters.

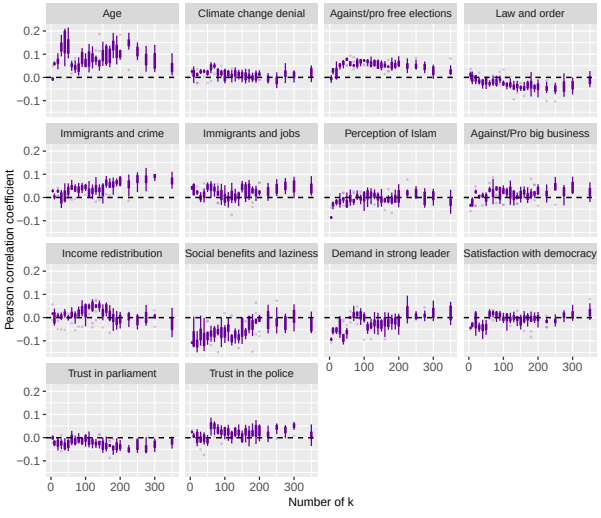

Figure 5: Correlations of political attitudes as a function of the number of SVD components k based on domain visits with duration of at least five minutes.

The number of k components that is most informative for the model is 100 under one-minute cutoff, and between 50 and 100 under five-minute cutoff depending on the political attitude. Overall, correlations drop significantly after 100 components depending on the duration cutoff, indicating that SVD dimensions beyond that do not have more explanatory power. As mentioned before, the cutoffs for the duration of a visit also contribute to the model performance. Figure 4 shows that the results are improving when we apply a five-minute cutoff for duration per visit, which is consistent with our theory that websites where individuals spend more than five minutes should reveal more about individuals' lifestyle and therefore political attitudes. The stronger results with the five-minute cutoff also signify that duration and therefore deliberate visit made by an individual matters in the models that utilize web tracking data compared to Facebook pages, which users often forget to revisit.[8]

4.2.2 Political attitudes with unstable predictions. Figure 5 summarizes the results for the rest of the attitudes based on the visits with at least five minutes duration. Similar to Figure 4, we plot Pearson correlation coefficients as a function of the number of k components obtained from SVD, and aggregated by the cross-validation folds.

[8] Although pages are active in posting various content and even send notifications to their subscribers about updates, in practice users tend to ignore them and therefore eventually stop seeing the updates in the newsfeed because the Facebook algorithm assumes that a user is not interested in that content.

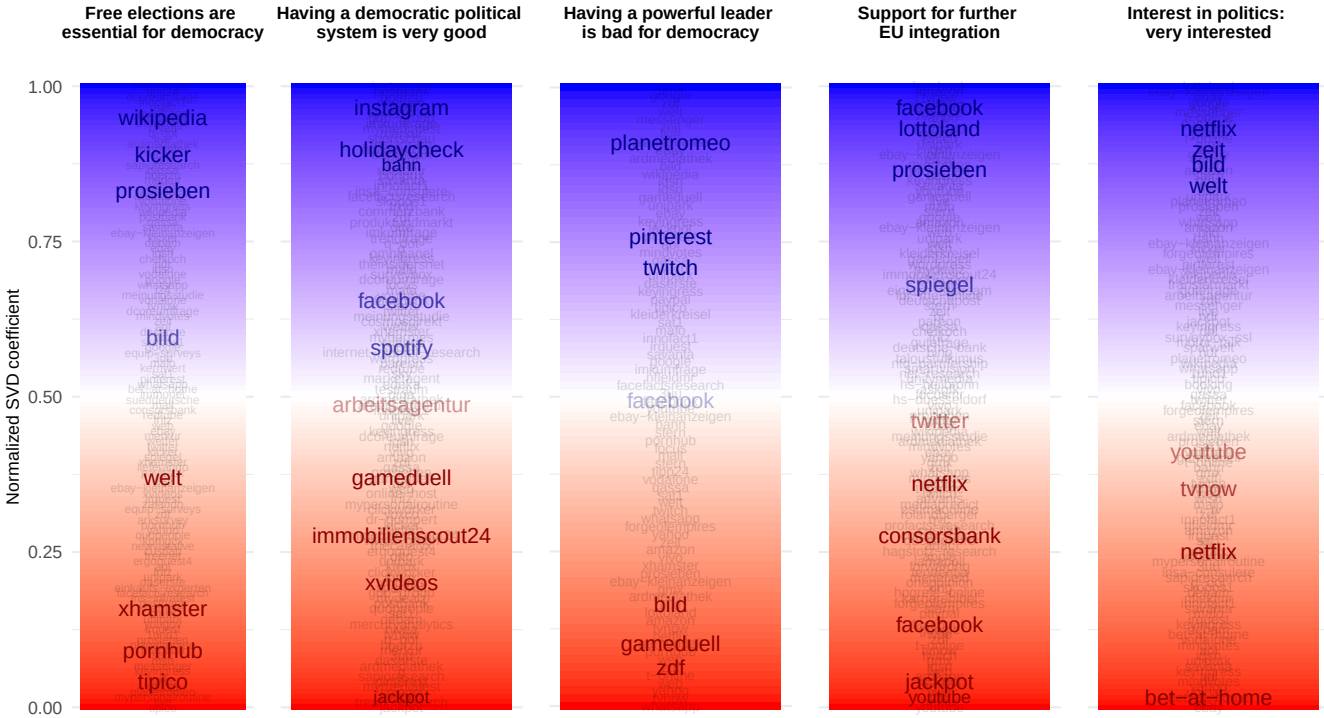

Figure 6: Association of political attitudes with selected domains. Domains on the top are positively and domains on the bottom are negatively associated with a corresponding attitude. Normalized SVD coefficients represent loadings or weights for every domain. Because we place domains from different clusters on mutual scale, some domains like Facebook appear both on top and on the bottom implying that in one cluster Facebook associated with pro-democratic and in the other – with authoritarian attitudes.

Compared to coefficients for age, attitudes towards free elections highest and relatively stable predictions followed by perceptions of crime and immigration (highest r = 0.13). While preferences for income distribution by the government and taxing big businesses are not significant, the attitudes towards social benefits has a significant but unstable correlation (the highest r = -0.15), suggesting that web browsing behavior can reveal people who are against state spending on social welfare.[9] With less stability, the data signals individuals' attitudes towards law and order policies, perception of Islam, and demand for a stronger leader. Correlations for attitudes towards immigrants taking natives' jobs, climate change denial, trust in public institutions, and satisfaction with democracy are not distinguishable from zero.

4.2.3 Association of political attitudes with domains. We explore which websites specifically are associated with political attitudes. For each political attitude with stable predictions and cross-validation, we took only SVD components with which they have significant correlation. We then filtered top 20 domains with positive loadings, and top 20 domains with negative loadings within each component, signifying the direction in which every domain is associated with the attitude of interest. Figure 6 illustrates association of political attitudes with selected domains that are related to our theoretical expectations. We normalized domain ladings between 0 and 1 and placed it on two colour gradient, where blue means positive association and red – negative association of a domain with an attitude. First three political attitudes represent democracy dimension. Websites that are related to recreation or creativity like Instagram, Spotify, Pinterest, as well as soccer news outlet Kicker are associated with pro-democratic attitudes. Hotel and train booking websites HolidayCheck and Deutsche Bahn are most representative of individuals, who value democratic political system. One possible mechanism that could explain this connection is that people, who are traveling and seeing other cultures benefit from democratic system and therefore support it. Visits to a gay dating platform is associated with opposition of having undemocratic leader, which is also confirms the existent evidence in the literature that LGBT community tend to support democratic principles because they protect LGBT rights. Visits to quality press magazine Spiegel signal people with pro-EU integration attitudes, whereas, as expected, visits to media outlets Zeit, Bild, and Welt point at individuals who are interested in politics.

[9] A closer inspection to the data is necessary to explain why such predictions are in the opposite directions.

Websites that are related to low quality entertainment and information (pornhub and xvideos), as well as online gambling sites (tipico, jackpot, and bet-at-home) signal individuals with authoritarian attitudes. Visits to estate marketplace (Immobilienscout24) and micro-credit banks (Consorsbank) are associated with anti-democratic and anti-EU attitudes. This suggests that people who are in search for apartments or in debt might feel insecure financially and therefore does not feel that current political system and the EU policies meets the needs of the people. In contrast, German federal job agency does not correlate with either of sides contradicting to the literature that connect right-wing populist support with unemployment.

Social media platforms such as Facebook, Twitter, video hosting platform YouTube, streaming website Nexflix, and online gaming appear on both poles (with YouTube appearing more often on the right), implying that social media plays significant role for people with any political views. Further in depth text-analysis of URLs is required to learn about differences in social media use between individuals with left and right ideological views.

5 DISCUSSION

The results of this study demonstrate that information about website choices available from individuals' browsing histories can signal their political attitudes. Our limited dataset on 1,003 panelists in Germany, which generated approximately 19,000,000 URLs after three months of tracking, showed that interest in politics, attitudes towards democracy, Euroscepticism, and perception of immigration and its relation to crime, as well as effectiveness of social benefits were especially identifiable from the web tracking data. Our straightforward estimation also shows that some political attitudes have stable zero predictability: attitudes towards climate change and trust in public institutions, among others. Moreover, the predictions of political attitudes are improving when we apply strict visit duration thresholds. In order to filter for relevant websites that meaningfully inform about individuals' lifestyle and values, we applied a five-minute threshold to filter out accidental or not significant websites. Results improved drastically pointing out to the importance of weighting that every website has in the daily life of an individual.

Although we performed several tests of our sample and the data on generalizability (see the appendix A), our results represent a conservative estimation of the predictive power of web browsing data. Our estimation is based on bounded ordinal independent variables that are common in the political science literature to measure political attitudes, but not always informative for machine learning models. With larger samples, better representations of URLs that are not limited to domains, alternative continuous instead of categorical measure of attitudes, and various model specifications, we expect the findings to gain more accuracy and robustness. Despite the limitations of our data and measurement, the results are compatible with previous studies of individuals' personalities with larger samples. Our highest predictions for interest in politics, Euroscepticism, attitudes towards immigration and democracy varies from r = 0.09 to 0.15 compared to 0.17 for satisfaction with life also measured on five-point scale in [29], [0.20, 0.40] average estimation in [41] and in [15]. However, political attitudes related to populism and immigration demonstrated only relative stability in cross-validations, sample dependence, and low although still significant correlation coefficients, which is still consistent with the literature on classifying political orientation with social media data [8].

In summary, this paper introduces an empirical strategy for analyses of web tracking data with application to political behavior and attitudes. The study also makes suggestions for further research that seeks to explain political phenomena through the analysis of browsing histories. Finally, our results have important implications for policy makers in digital privacy and the society in general by emphasising beneficial as well as disadvantages for democracy potential of web browsing behavior to reveal peoples' political attitudes and issue preferences.

ACKNOWLEDGMENTS

The Volkswagen Foundation funded this project, Grant #94758. We thank the faculty of GESIS CSS Department and the TUM School of Government for helpful comments.

REFERENCES

[1] Pushkal Agarwal, Sagar Joglekar, Panagiotis Papadopoulos, Nishanth Sastry, and Nicolas Kourtellis. 2020. Stop Tracking Me Bro! Differential Tracking of User Demographics on Hyper-Partisan Websites. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1479–1490. https://doi.org/10.1145/3366423.3380221

[2] Luciano Arcuri, Luigi Castelli, Silvia Galdi, Cristina Zogmaister, and Alessandro Amadori. 2008. Predicting the Vote: Implicit Attitudes as Predictors of the Future Behavior of Decided and Undecided Voters. *Political Psychology* 29, 3 (May 2008), 369-387. https://doi.org/10.1111/j.1467-9221.2008.00635.x

[3] Anonymous authors. [n.d.]. ([n. d.]).

[4] Anonymous authors. [n.d.]. ([n. d.]).

[5] Anonymous authors. [n.d.]. ([n. d.]).

[6] Yoram Bachrach, Michal Kosinski, Thore Graepel, Pushmeet Kohli, and David Stillwell. 2012. Personality and Patterns of Facebook Usage. In *Proceedings of the 4th Annual ACM Web Science Conference* (Evanston, Illinois) *(WebSci '12)*. Association for Computing Machinery, New York, NY, USA, 24–32. https://doi.org/10.1145/2380718.2380722

[7] Roberto Cerina and Raymond Duch. 2020. Measuring public opinion via digital footprints. *International Journal of Forecasting* (March 2020). https://doi.org/10.1016/j.ijforecast.2019.10.004 To appear.

[8] Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on Twitter: It's not easy!. In *Seventh international AAAI conference on weblogs and social media.*

[9] Alexander Coppock, Seth J. Hill, and Lynn Vavreck. 2020. The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments. *Science Advances* 6, 36 (2020). https://doi.org/10.1126/sciadv.abc4046

[10] Russell J. Dalton. 2000. Citizen Attitudes and Political Behavior. *Comparative Political Studies* 33, 6-7 (2000), 912–940. https://doi.org/10.1177/001041400003300609

[11] Daniel DellaPosta, Yongren Shi, and Michael Macy. 2015. Why Do Liberals Drink Lattes? *Amer. J. Sociology* 120, 5 (2015), 1473–1511. https://doi.org/10.1086/681254

[12] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-Million-Site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna, Austria) *(CCS '16)*. Association for Computing Machinery, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313

[13] D. C. Evans, S. Gosling, and A. Carroll. 2008. What Elements of an Online Social Networking Profile Predict Target-Rater Agreement in Personality Impressions?. In *ICWSM.*

[14] Matthias Fatke. 2017. Personality Traits and Political Ideology: A First Global Assessment. *Political Psychology* 38, 5 (2017), 881–899. https://doi.org/10.1111/pops.12347

[15] David C. Funder and Daniel J. Ozer. 2019. Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science* 2, 2 (2019), 156–168. https://doi.org/10.1177/2515245919847202 arXiv:https://doi.org/10.1177/2515245919847202

[16] R. Kelly Garrett. 2009. Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate. *Journal of Communication* 59, 4 (12 2009), 676–699. https://doi.org/10.1111/j.1460-2466.2009.01452.x

[17] Gordon Gauchat. 2018. Trust in climate scientists. *Nature Climate Change* 8, 6 (2018), 458–459. https://doi.org/10.1038/s41558-018-0147-4
[18] Alan S. Gerber, Gregory A. Huber, David Doherty, and Conor M. Dowling. 2010. Personality and Political Attitudes: Relationships across Issue Domains and Political Contexts. *American Political Science Review* 104, 1 (Feb. 2010), 111-133. https://doi.org/10.1017/S0003055410000031
[19] Jennifer Golbeck, Cristina Robles, and Karen Turner. 2011. Predicting Personality with Social Media. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI EA '11)*. Association for Computing Machinery, New York, NY, USA, 253–262. https://doi.org/10.1145/1979742.1979614
[20] G. H. Golub and C. Reinsch. [n.d.]. Singular value decomposition and least squares solutions. 14, 5 ([n. d.]), 403–420. https://doi.org/10.1007/BF02163027
[21] Andrew M. Guess, Brendan Nyhan, and Jason Reifler. 2020. Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour* 4, 5 (2020), 472–480. https://doi.org/10.1038/s41562-020-0833-x
[22] Jens Hainmueller and Daniel J. Hopkins. 2014. Public Attitudes Toward Immigration. *Annual Review of Political Science* 17, 1 (2014), 225–249. https://doi.org/10.1146/annurev-polisci-102512-194818
[23] Eelco Harteveld, Andrej Kokkonen, and Stefan Dahlberg. 2017. Adapting to party lines: the effect of party affiliation on attitudes to immigration. *West European Politics* 40, 6 (June 2017), 1177-1197. https://doi.org/10.1080/01402382.2017.1328889
[24] Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen. 2007. Demographic Prediction Based on User's Browsing Behavior. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada) *(WWW '07)*. Association for Computing Machinery, New York, NY, USA, 151–160. https://doi.org/10.1145/1242572.1242594
[25] Margaret L. Kerna, Paul X. McCarthyb, Deepanjan Chakrabartyb, and Marian-Andrei Rizoiud. 2019. Social media-predicted personality traits and values can help match people to their ideal jobs. *Proceedings of the National Academy of Sciences of the United States of America* 116, 52 (dec 2019), 26459-26464. https://doi.org/10.1073/pnas.1917942116
[26] Olga M. Klimecki, Matthieu Vétois, and David Sander. 2020. The impact of empathy and perspective-taking instructions on proponents and opponents of immigration. *Humanities and Social Sciences Communications* 7, 1 (2020), 91. https://doi.org/10.1057/s41599-020-00581-0
[27] David M. Konisky, Jeffrey Milyo, and Lilliard E. Richardson. 2008. Environmental Policy Attitudes: Issues, Geographical Scale, and Political Trust*. *Social Science Quarterly* 89, 5 (2008), 1066–1085. https://doi.org/10.1111/j.1540-6237.2008.00574.x
[28] Michal Kosinski, Yoram Bachrach, Pushmeet Kohli, David Stillwell, and Thore Graepel. 2014. Manifestations of user personality in website choice and behaviour on online social networks. *Mach Learn* 95, 3 (jun 2014), 357–380. https://doi.org/10.1007/s10994-013-5415-y
[29] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America* 110, 15 (April 2013), 5802-5805. https://doi.org/10.1073/pnas.1218772110
[30] Michal Kosinski, David Stillwell, Pushmeet Kohli, , Yoram Bachrach, and Thore Graepel. 2012. Personality and Website Choice. In *ACM Web Sciences 2012* (acm web sciences 2012 ed.). ACM Conference on Web Sciences.
[31] Michal Kosinski, Yilun Wang, Himabindu Lakkaraju, and Jure Leskovec. 2016. Mining big data to extract patterns and predict real-life outcomes. 21, 4 (2016), 493–506. https://doi.org/10.1037/met0000105
[32] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America* 111, 24 (jun 2014), 8788-8790. https://doi.org/10.1073/pnas.1320040111
[33] Jon A. Krosnick. 1991. The Stability of Political Preferences: Comparisons of Symbolic and Nonsymbolic Attitudes. *American Journal of Political Science* 35, 3 (1991), 547–576. http://www.jstor.org/stable/2111553
[34] Z Kunda. 1990. The case for motivated reasoning. *Psychological bulletin* 108, 3 (November 1990), 480—498. https://doi.org/10.1037/0033-2909.108.3.480
[35] R. Lambiotte and M. Kosinski. 2014. Tracking the Digital Footprints of Personality. *Proc. IEEE* 102, 12 (2014), 1934–1939.
[36] Sergio Lo Iacono. 2019. Law-breaking, fairness, and generalized trust: The mediating role of trust in institutions. *PLOS ONE* 14, 8 (08 2019), 1–14. https://doi.org/10.1371/journal.pone.0220160
[37] Markus Prior. 2013. Media and Political Polarization. *Annual Review of Political Science* 16, 1 (2013), 101–127. https://doi.org/10.1146/annurev-polisci-100711-135242
[38] Matthijs Rooduijn. 2018. What unites the voter bases of populist parties? Comparing the electorates of 15 populist parties. *European Political Science Review* 10, 3 (2018), 351–368. https://doi.org/10.1017/S1755773917000145
[39] Michele Settanni, Danny Azucar, and Davide Marengo. 2018. Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking* 21, 4 (2018), 217–228. https://doi.org/10.1089/cyber.2017.0384 PMID: 29624439.
[40] Feng Shi, Yongren Shi, Fedor A. Dokshin, James A. Evans, and Michael W. Macy. 2017. Millions of online book co-purchases reveal partisan differences in the consumption of science. *Nature Human Behaviour* 1, 4 (April 2017). https://doi.org/10.1038/s41562-017-0079
[41] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D. Gosling, Gabriella M. Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, Heinrich Hussmann, Bernd Bischl, and Markus Bühner. 2020. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences* 117, 30 (2020), 17680–17687. https://doi.org/10.1073/pnas.1920484117
[42] Andreas Wiedemann. 2020. Austerity, Indebtedness, and Political Behavior. Evidence from the U.K. *Working paper* (2020).

A EVALUATION OF WEB TRACKING DATA AND THE PANEL OF RESPONDENTS

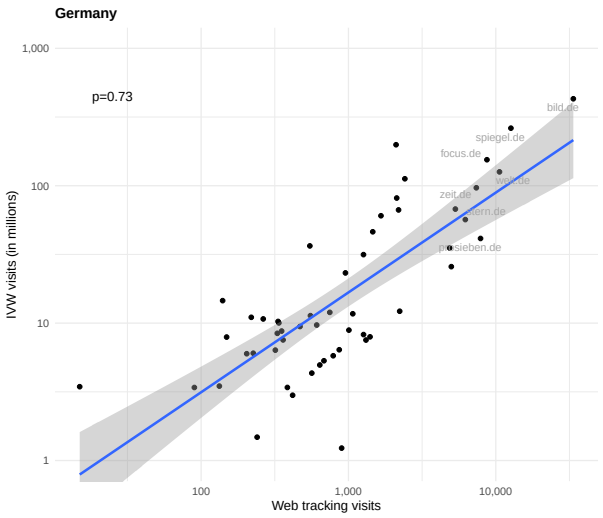

Figure 7: Comparison of IVW domain visit rankings and news domains visited by our web tracking panelists in Germany. Both axes are logged; p is Spearman's rank correlation.

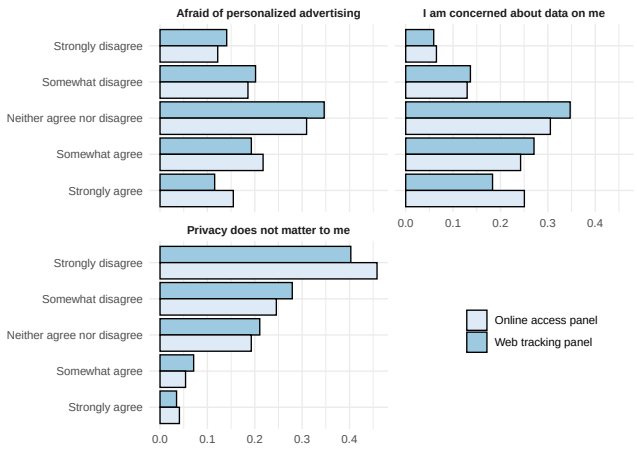

Figure 8: Comparison of privacy attitudes in a survey of German online access panelists and web tracking panelists.