Social Desirability and the Willingness to Provide Social Media Accounts in Surveys. The Case of Environmental Attitudes

Beate Klösch[1], Markus Hadler[1], Markus Reiter-Haas[2], Elisabeth Lex[2]
[1] Department of Sociology, University of Graz, Austria, [2] Institute of Interactive Systems and Data Science, Graz University of Technology, Graz, Austria.

Abstract

This paper contributes to the research on combining public opinion surveys and social media data by a) analyzing the effects of social desirability on the willingness to provide social media account information in surveys, and b) evaluating the congruence of opinions expressed in the survey and on social media. We analyze these questions by considering the willingness to make a sacrifice for the environment, i.e., the willingness to pay higher taxes and higher prices. Our results show that Facebook users who oppose environmental measures are less likely to share their account information in the survey, whereas this effect could not be found among Twitter users. Considering the congruence of opinions expressed in the survey and on Twitter, we find similar tendencies both at the aggregate and the individual level.

Keywords: *Survey; Social Media; Facebook; Twitter; Sentiment Analysis; Environmental policies.*

1. Introduction

In a previous paper (Hadler et al. 2022), we showed that the likelihood of providing one's social media account information in surveys is higher among respondents who are in favor of various COVID-19 policy measures. One explanation is the occurrence of a social desirability bias. Social desirability refers to the effect that respondents tend to report behaviors and opinions that are generally assessed positively and give socially desirable answers rather than share their true thoughts if those deviate from social norms (Grimm 2010). Anonymous interviewing and ensuring confidentiality are known to reduce this bias (Larson 2019). Providing access to their social media accounts removes the respondents' anonymity, as researchers will know their actual identities. In line with the idea of social desirability, respondents who share the mainstream opinion to support COVID-19 measures more often provide their account information than respondents who oppose them.

The COVID-19 pandemic and related policy measures, however, are specific topics as they resulted in a polarized public opinion (Reiter-Haas et al. 2022). The current paper thus tests whether our findings on the effects of social desirability are also applicable to another topic – environmental attitudes, i.e., the willingness to pay higher taxes and prices to protect the environment. Finding similar effects would support the general idea of a bias in providing one's social media account towards respondents who share mainstream views. Therefore, our first research question is: *Do attitudes towards environmental protection measures influence the willingness to provide social media accounts in a public opinion survey?*

Alongside this question, we also compare opinions towards environmental measures expressed on social media to those stated in our survey. First, we compare attitudes reported in a public opinion survey with sentiments expressed on social media at the aggregate level. Second, we also take a closer look at the congruence at the individual level and check whether single individuals express the same opinions in the survey and on social media. Therefore, our second research question is: *Do the attitudes expressed in surveys match the social media sentiments at the aggregate and the individual level?*

2. Methods

Our analyses are based on three data sources, i.e., a public opinion survey, the tweets of survey respondents who shared their social media account names, and Twitter data for the same time period. The survey was conducted online in the DACH region (i.e., Germany, Austria, Switzerland) in the summer of 2020. The sample comprises a total of 2560 respondents and resembles the sociodemographics of each country. The questions included attitudes towards the COVID-19 pandemic, environmental attitudes, the use of various social media platforms, and sociodemographics. Respondents were also asked to share the name of

their Facebook and Twitter accounts. However, we were only able to access the Twitter data due to Facebook's terms and conditions.

Our dependent variable is derived from the questions on sharing one's account information and includes the following groups for Facebook and Twitter users respectively: a) respondents without an account, b) account holders who were not willing to share their account name, and c) account holders who shared their account name. Independent variables include the sociodemographic variables gender, age, and education, as well as attitudes towards environmental measures ('How willing would you be to a) pay higher prices, and b) pay higher taxes in order to protect the environment?'). Responses to this item are measured on a five-point scale with 1 = no acceptance and 5 = high acceptance. We also computed an index, displaying the mean score of the two attitudes. An overview of the survey variables is provided in Table 1.

Table 1. Sample characteristics

Variables	Mean (SD) or %
Acceptance of environmental measures **(1 = no acceptance, 5 = high acceptance)**	
How willing would you be to…	
a) pay higher prices in order to protect the environment	2.90 (1.17)
b) pay higher taxes in order to protect the environment	2.40 (1.17)
Index (mean score of previous variables)	2.65 (1.20)
Sociodemographic variables	
Female	50.4%
Age	44.34 (13.90)
Education	
Compulsory school	35.2%
Vocational training	11.6%
High school degree	23.9%
University degree	29.3%

For the Twitter data, we used the Twitter Search API[1] and matched the time period to the survey data (i.e., from July 30th, 2020, to August 10th, 2020). The search query was restricted to the German language containing terms related to the environment[2], which resulted in a total of 16,780 tweets. Furthermore, we considered two smaller subsets of these tweets containing the German and English terms for prices (i.e., 'preis' and 'price' = 275 tweets) and taxes (i.e., 'steuer' and 'tax' = 470 tweets). Subsequently, we conducted a sentiment

---

[1] We used twarc2 with the full-archive search using the academic research product track (https://twarc-project.readthedocs.io/en/latest/twarc2_en_us/#search).

[2] The list contains 'environment', 'climate', their German counterparts 'umwelt' and 'klima', as well as their corresponding hashtags (e.g., #environment).

analysis using the TextBlob library with the German language extension, which includes a sentiment polarity lexicon that we use for sentiment extraction. We chose this approach as it applies well to the analysis of short texts, such as tweets. After extracting the sentiment, we removed tweets that express no sentiment to exclude purely objective statements which would otherwise dominate the resulting distribution. The extracted sentiments are on a scale from −1 for negative to +1 for positive sentiment.

For the analysis at the individual level, we manually annotated all tweets mentioning the keywords 'environment'/'umwelt', 'climate'/'klima', or the accompanying hashtags from those respondents, who provided us with their Twitter handle and compared their overall opinion with their survey answers. This resulted in 60 tweets by 9 individuals, as only this small number of respondents who provided their account information used the selected keywords or hashtags in their tweets or retweets. Since they also shared only a few tweets, we expanded our time period from the 12 days of the survey to the entire year 2020. Our research required specific ethical considerations (Sloan et al. 2020). We informed the survey respondents about the content of our research, the voluntary nature of their participation, as well as the confidential treatment of their data. Hence, the archived dataset (Hadler et al. 2021) does not include any information that would allow the identification of individuals, including Facebook and Twitter account names or tweets.

3. Results

3.1. Social Media use and the willingness to provide account information

As shown in Table 2, our results indicate that almost 70% of the survey respondents use Facebook, whereas only 16% use Twitter. Among these, a similar proportion describes themselves as active users in each case (40% vs. 35%), as well as a similar number of users of both platforms have shared their account information (35% vs. 30%) which is in line with the numbers reported in previous studies (Al Baghal et al. 2020). However, Facebook's terms and conditions do not allow access to their data, which is why we could only analyze the Twitter data of our respondents. At last, we were able to access 79 Twitter accounts as 40 people (accidentally) provided a false account name or protected account.

Table 2. Overview on Social Media use and the willingness to provide account information

	Facebook	Twitter
Account holders[3]	1774 (69.3%)	404 (15.8%)
Active users[4] (% of account holders)	700 (39.5%)	141 (34.9%)
Account provided[5] (% of account holders)	617 (34.8%)	119 (29.5%)
Successfully accessed[6] (% of account holders)	N.A.*	79 (19.6%)
Total sample	2560	2560

*Account access is restricted by Facebook's terms and conditions.

We conducted two multinomial logistic regression models to estimate the influence of opinions on environmental measures and sociodemographics on the willingness to provide one's account information. The reference groups in Table 3 are respondents who provided their account information. Hence, the regression models show the results for survey respondents who do not have a Facebook or Twitter account, and for those who do have an account but did not share them.

Table 3. Regression analysis on Social Media use and the willingness to provide account information

	Facebook account (ref.: account shared) b-values		Twitter account (ref.: account shared) b-values	
	no account	account but not provided	no account	account but not provided
Intercept	-.629	.666*	.864	1.210
Acceptance of env. measures (index)	-.115*	-.154**	-.103	-.029
Gender (Female)	-.138	.162	.770***	-.052
Age	.031***	.004	.025**	-.002
Compulsory School	-.177	-.201	.275	-.178
Vocational training	-.225	-.152	.319	-.401
High school degree	.334*	.157	-.017	-.159
Cox-Snell	.040		.044	
Nagelkerke	.046		.067	
X^2 (df)	103.621***		112.674***	
n	2516		2516	

Note: Acceptance of env. measures low value = low acceptance; Gender 1 = male, 2 = female; Age numeric; Education reference category = University degree. $* = p < 0.05$, $** = p < 0.01$, $*** p < 0.001$.

---

[3] 'Do you have a private Facebook/Twitter account?' (yes/no)

[4] 'How would you describe the way you use Facebook/Twitter?' (actively, i.e., posting/passively, i.e. reading etc.)

[5] 'We would like to find out who is using Facebook/Twitter and for which purposes. If you provide access, we will keep all your information confidential. Do you agree to provide us with your personal Facebook/Twitter username as well as access to your data in order to link it to this survey data?' (yes/no)

[6] 'What is your username?'

As for Facebook, the results indicate that respondents without an account tend to be older and have a high school degree rather than a university degree. In terms of environmental willingness, they are more likely against the surveyed environmental measures. Similarly, respondents who do have a Facebook account but did not share it in the survey are more likely opposed to environmental measures. In comparison, opinions on environmental measures have no significant influence on the willingness to share one's Twitter account. Only some sociodemographic variables show significant effects in the sample, i.e., that women and older individuals are less likely to have a Twitter account. To evaluate the soundness of our results, we additionally performed binary logistic regression models (account shared vs. account not shared; account shared vs. account not shared and no account), whose results are very similar to those of the multinomial models.

3.2. Comparing public opinion and sentiments expressed on Twitter

We need to limit our second research question – do the opinions regarding environmental measures shared in the survey match those expressed on social media – to the Twitter data, as we do not have access to the Facebook data. The results regarding the overall Twitter data show a positive sentiment in tweets using the keywords or hashtags 'environment' (median = 0.7, mean = 0.21) or 'climate' (median = 0.7, mean = 0.27). Both keywords show a similar distribution, whereby the sentiment regarding the term 'environment' is slightly more dispersed. For the matching at the aggregate level, we first compare the survey responses to the sentiment analysis considering all tweets related to environmental measures 'pay higher prices' and 'pay higher taxes'. Second, we turn towards the individual comparison of those survey respondents that provided their Twitter handles.

Figure 1 shows the opinions on the two environmental measures expressed on Twitter (top) as well by all survey respondents (middle) and survey respondents who have a Twitter account (bottom). The Twitter boxplots (top) show that price (median = 0.7; mean = 0.32) receives way more positive sentiments than tax (median = -0.7; mean = -0.37). Regarding price, the opinions are more diverse as the wider quartile range indicates, yet show the most positive sentiment within all data sources.
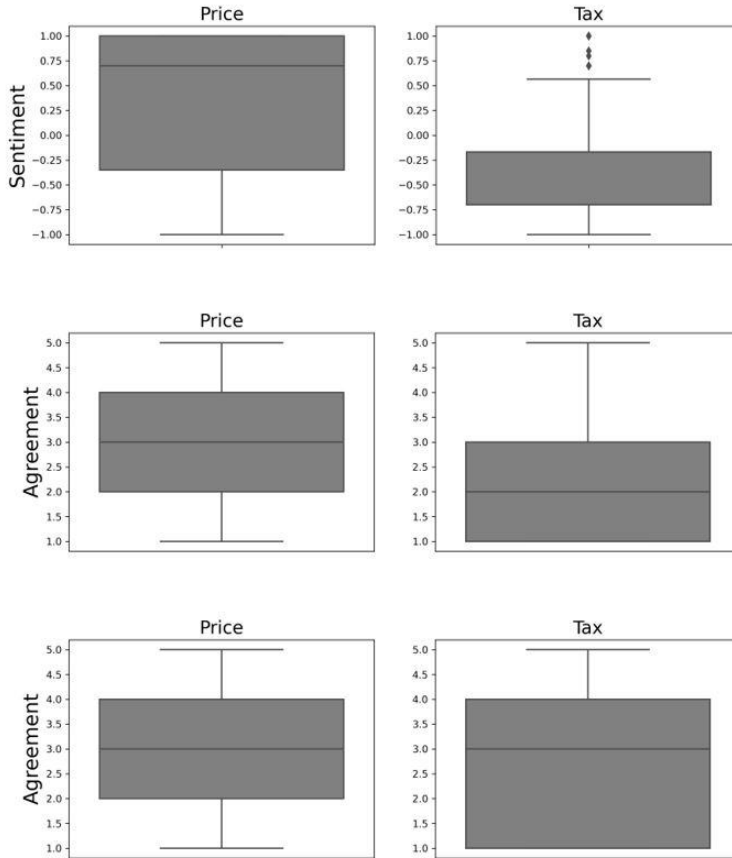
Figure 1. Opinions towards environmental measures expressed on Twitter and in the survey. Top: Twitter sentiments. Middle: All survey respondents. Bottom: Survey respondents with a Twitter account.

Note: Polarization in the Twitter data in terms of sentiment regarding the two environmental measures during the survey period. The sentiments are measured per tweet on a range from −1 for the maximum negative sentiment to +1 for the maximum positive sentiment. Tweets with neutral sentiment are excluded. n (price)=275, n (tax) = 470. Polarization in the survey data based on the question: 'How willing would you be to a) pay higher prices, b) pay higher taxes in order to protect the environment?'; 1=no acceptance and 5=high acceptance. n_{all} (price)=2512, n_{all} (tax)=2511; $n_{account}$ (price)=398, $n_{account}$ (tax)=397.

As for the survey respondents (middle and bottom boxplots), opinions regarding paying higher prices to protect the environment are distributed similarly (median$_{all}$ = 3, mean$_{all}$ = 2.90; median$_{account}$ = 3, mean$_{account}$ = 2.98), regardless of whether they have a Twitter account or not. Opinions regarding paying higher taxes, however, tend to be more positive in respondents holding a Twitter account (median$_{account}$ = 3, mean$_{account}$ = 2.55), compared to all survey respondents (median$_{all}$ = 2, mean$_{all}$ = 2.40), as the median is one value higher, and the dispersion grows towards agreement. Overall, the measure 'pay higher prices' receives more support in all datasets than 'pay higher taxes to protect the environment', and the comparison

of survey opinions and Twitter sentiments at the aggregate level shows a relatively good overlap concerning the tendency of opinions. However, on Twitter people can express their opinions using multiple tweets whereas in the survey there is only one answer per respondent and item. Therefore, we also take a look at the individual level and compare the survey responses of those who provided their Twitter accounts with their actual tweets regarding environmental measures.

Considering the match at the individual level, we were able to assess 60 tweets and retweets by 9 survey respondents using the keywords or hashtags 'environment' or 'climate'. The agreement with environmental measures was assigned manually, using the same scale (1 = no acceptance, 5 = high acceptance) as in the survey, and compared in terms of congruence to the opinions towards environmental measures shared by these respondents in the survey. Overall, the results show a relatively high congruence within the two data sources by these 9 people, although no assessment could be made for two respondents' opinions on Twitter due to differing or unclear content (such as using #climate in working atmosphere content).

4. Conclusion

To summarize our findings regarding our first research question, we find divergent results regarding the willingness to share one's social media account depending on the selected social media platform. As for Facebook, opinions regarding environmental measures have a significant effect on the willingness to share one's account information. Respondents who oppose those measures are more likely to refuse sharing their social media accounts. However, this effect cannot be proven for Twitter. These results are in line with our findings regarding COVID-19 measures (Hadler et al. 2022). Hence, we assume that Facebook is potentially a more polarized platform, whose users are more cautious about sharing their data and that the social desirability bias is stronger on Facebook than on Twitter, regardless of the investigated topic. This may be attributable to the fact that different social media platforms follow diverse agendas and have different aims, and users may differ accordingly. For Twitter, we assume, based on our findings, that users tend to be more scientifically minded and therefore agree with current policies as well as with making their data available for research purposes.

Regarding our second research question, we find a relatively high congruence between the opinions shared in the survey and those posted on Twitter, both at the aggregate and the individual level. Our findings differ in this respect from previous studies, such as Pasek et al. (2020), or Amaya et al. (2020), who both showed that social media content concerning political issues on Twitter or Reddit differs from public opinion, i.e., in the European Social Survey. This could be due to the fact that opinions in our data regarding environmental measures are less subject to social desirability bias than other, mainly political, topics.

However, our study faces some limitations. First, social media users are not representative of the overall population in terms of sociodemographics, as they tend to be younger. Second, comparisons between social media data and survey data at the aggregate level always face the bias of an unequal number of responses per individual, as social media users can express their opinions in multiple tweets or postings whereas there is only one answer per item per survey respondent. Finally, our congruence analysis at the individual level is based on a very small number of subjects, as individuals who provide their social media account must also tweet about the particular topic in order to be included in the analysis. Hence, the findings might not be suitable for generalization but can provide a first insight into the field and serve as a basis for further investigations.

References

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2020). Linking Twitter and Survey Data: The Impact of Survey Mode and Demographics on Consent Rates Across Three UK Studies. *Social Science Computer Review* 38 (5), 517-532. 10.1177/0894439319828011.

Amaya, A., Bach, R., Kreuter, F., & Keusch, F. (2020). Measuring the Strength of Attitudes in Social Media Data. In: Hill, C. A., Biemer, P. P., Buskirk, T. D., Japec, L., Kirchner, A., Kolenikov, S., Lyberg, L. E. (eds.) *Big Data Meets Survey Science: A Collection of Innovative Methods*, 163-192. 10.1002/9781118976357.ch5.

Grimm, P. (2010). Social desirability bias. *Wiley international encyclopedia of marketing, Part 2 Marketing Research*. 10.1002/9781444316568.wiem02057.

Larson, R. B. (2019). Controlling social desirability bias. *International Journal of Market Research* 61 (5), 534-547. 10.1177/1470785318805305.

Hadler, M., Klösch, B., Lex, E., & Reiter-Haas, M. (2021). *Polarization in public opinion: Combining social surveys and big data analyses of Twitter* (SUF Edition). 10.11587/OVHKTR, AUSSDA, V1, UNF:6:jPjxWXqS6RVg4uYo3Zplcw== [fileUNF].

Hadler, M., Klösch, B., Reiter-Haas, M., & Lex, E. (2022). Respondents' opinion on COVID-19 and their willingness to provide their social media info in a survey. Paper presented at the *Annual Meeting of the American Sociological Association*, Los Angeles, August 2022.

Pasek, J., McClain, C., Newport, F., & Marken, S. (2020). Who's tweeting about the president? What big survey data can tell us about digital traces? *Social Science Computer Review* 38 (5), 633-650. 10.1177/0894439318822007.

Reiter-Haas, M., Klösch, B., Hadler, M., & Lex, E. (2022). Polarization of Opinions on COVID-19 Measures: Integrating Twitter and Survey Data. *Social Science Computer Review*. 10.1177/08944393221087662.

Sloan, L., Jessop, C., Al Baghal, T., & Williams, M. (2020). Linking survey and Twitter data: informed consent, disclosure, security, and archiving. *Journal of Empirical Research on Human Research Ethics* 15 (1-2), 63-76. 10.1177/1556264619853447.