IFS

Institute for Fiscal Studies

Bram De Rock
Florine Le Henaff

Working paper 23/34

Walk the talk: Measuring green preferences with social media data

Economic and Social Research Council

Walk the talk: Measuring green preferences with social media data

Bram De Rock* Florine Le Henaff†

October 25, 2023

Abstract

We created a unique data set based on social media data by collecting and geo-localising all the tweets of 54 thousand Swedish citizens from January 2019 to June 2019. This allows us to construct an attractive individual-level measure of preferences for pro-environmental behavior. We demonstrate this by using our measure in two applications. We first document a subjective well-being gap between individuals with and without green preferences, using the average sentiment scores in tweets as a proxy of individuals' subjective well-being. We then investigate the existence of a gender gap in green preferences and the propensity to act for the environment, relating our measure to publicly available data on electric and hybrid car registrations and political support for environmental policies in Sweden.

*ECARES, Université Libre de Bruxelles and Department of Economics, University of Leuven (KU Leuven). Av. F. Roosevelt 50, B-1050 Bruxelles, Belgium. E-mail: bram.de.rock@ulb.be.
Bram De Rock gratefully acknowledges FWO and FNRS for their financial support.

†ECARES, Université Libre de Bruxelles. Av. F. Roosevelt 50, B-1050 Bruxelles, Belgium. E-mail: florine.lehenaff@ulb.be.
Florine Le Henaff gratefully acknowledges FNRS for their financial support.

Keywords: Individual preferences, social media, pro-environmental behavior, subjective well-being, gender identities.
JEL Classification: D01, Q5, C8.

1 Introduction

In their report published in April 2022, scientists of the IPCC underline again the emergency of taking drastic actions to mitigate climate change, “if we are to limit global warming to 1.5°C”. In this context, a rapidly growing body of literature underlines the role of citizen mobilization and individual actions as essential components for global greenhouse gas emissions (GHGEs) reduction, both in terms of pro-environmental behaviors (PEBs), such as changing consumption patterns and habits (Stehfest *et al.*, 2009; Williamson *et al.*, 2018; Ivanova *et al.*, 2020), and in terms of public support of green policies aiming for example at increasing the prevalence of renewables in the energy mix (Lan *et al.*, 2010). IPCC scientists estimate the mitigation potential of demand-side options to be between 40 and 70% of GHGEs by 2050 with respect to the baseline scenario.

Although preferences for the conservation of the environment are a central determinant of environmentally sustainable individual choices and support for green policies, their estimation has until now mostly been based on stated preferences in surveys. This paper joins the growing economic literature using social media data (Acemoglu *et al.*, 2018; Campante *et al.*, 2020; Müller & Schwarz, 2021), as it proposes an individual-level measure of green preferences based on the unsolicited opinions of individuals on Twitter. Our attractive measure enables us to analyze two important phenomena related to pro-environmental preferences. We first investigate the potential subjective well-being (SWB) costs of environmental concern at the individual level. In a recent report, the American Psychology Association found that more than two-thirds (68%) of the adults surveyed had “at least a little eco-anxiety” about climate change and its effects (Schreiber, 2021). In this context, we investigate the existence of a negative relationship between environmental preferences and individuals’ SWB. Second, we test the finding of a gendered responsibilization of environmental protection in the environmental sociology literature (Hunter *et al.*, 2004; Kennedy & Kmec, 2018; Dzialo, 2017). We investigate the relationship between gender and green preferences, and, via geo-localizing our tweets, the relationship between green preferences and propensities to partake in PEBs across genders. As such we go beyond stated preferences by linking our measure to revealed preferences.

Methodology and results. We created a unique dataset gathering 54 thousand individuals' tweets over 6 months from January to June 2019 in Sweden. Our measure of preferences for the conservation of the environment is based in this paper on the assumption that individuals get utility from their time spent on Twitter and that they maximize this twitting utility by choosing to discuss the topics they have strong preferences for. As such, an individual is said to reveal environmentally-friendly preferences if they discussed environment-related issues in their tweets and we define as preference strength the share of tweets an individual dedicates to topics related to the environment.

We then use unsupervised machine learning (specifically the Vader sentiment analysis algorithm, Hutto & Gilbert (2014)) to proxy SWB and use a simple ordinary least square (OLS) estimation of the relationship between individuals' green preferences and their SWB. Tweets of individuals with green preferences have, on average, a lower level of expressed happiness than tweets of individuals without. Higher shares of tweets dedicated to the environment are also associated with lower SWB for individuals with strong green preferences. Our results therefore suggest that having green preferences appears to be associated with a well-being cost.

Finally, we investigate green preferences and propensity to act for the environment across genders, using real choice data. Therefore we need to geo-localize each Twitter user and based on a name analysis we also determine their gender. Individual-level behavior data not being available to match the individuals in our sample, we then use municipal-level measures of PEBs to investigate the relationship between green preferences and PEBs across genders. More precisely, we collect from the Swedish Statistical Agency (SCB) information on the share of green cars registered and the share of votes dedicated to the Green party in each municipality, which we relate to a municipal-level measure of male and female green preferences. Women are found to be more likely than men to have green preferences, and their preferences are also found to be stronger than the latter. Investigating the relationship between green preferences in municipalities and aggregated PEBs, we find evidence supporting the existence of a gender gap in the use of green cars but not in voting behavior. As we discuss below, these results are in line with the nuanced findings in the literature.

Contributions to the literature. The value of our new measure stands from the fact that it is based in its entirety on the unsolicited and spontaneous views of individuals shared on social media, and is therefore free from survey bias concerns. As such, our paper makes three contributions to the existing literature.

First, at a more methodological level, we contribute to the economic literature using social media in their analysis, by using Twitter publications to elicit preferences for the conservation of the environment. Close to this paper is the work of Baylis (2020) and Loureiro *et al.* (2022), which aims at estimating individuals' life satisfaction with expressed happiness in their tweets using natural language processing (NLP) techniques, and relate it respectively to the weather in the US and forest wildfires in Spain and Portugal to estimate preferences (or welfare costs in the case of wildfire) for these non-marketable goods. In contrast, we do not adopt a thematic approach and collect the entirety of the tweets of 54 thousand individuals over a 6-month period, focusing the analysis at the individual level and adapting an ex-ante expectation of utility approach. Namely, individuals estimate the utility they will get from discussing given topics on Twitter and maximize it by tweeting on a topic they care about.

Second, at the empirical level, this paper is, to the best of our knowledge, among the first papers to document a potential negative relationship between environmental concern and SWB. Although the psychological and economic literatures increasingly focus on the negative consequences of climate change on mental health, coping mechanisms, and emotional responses to eco-anxiety, no paper in these literatures aims at quantifying the SWB gap caused by environmental-concern. Analyzing user-created content online (memes), Elgaaied-Gambier & Mandler (2021) find that individuals concerned about the environment experience eco-anxiety, perceived lack of control, and lack of faith in the future. In their recent paper, Whitmarsh *et al.* (2022) investigate the predictors of eco-anxiety, and find that younger individuals and individuals partaking in more information-seeking behaviors on climate change have higher levels of eco-anxiety. We contribute to the economic well-being literature by shedding light on a negative correlation between green preferences and SWB at the individual level and motivating the need to investigate further the consequences of this anxiety on individuals' economic outcomes.

Finally, the paper contributes to the environmental literature focusing on gender as a driver of green preferences and PEBs in two ways. Accounts of environmental preferences rely almost exclusively on individual answers in surveys related to specific topics: answers to questions on economic trade-offs (Dunlap & Riley, 1983), on ecological worldviews (Stern *et al.*, 1993), on participation in given pro-environmental activities (Hunter *et al.*, 2004), on policy preferences (Konisky *et al.*, 2008), etc..[1] Our measure adds to the literature, since, not only is it not based on solicited stated preferences, but it is also not restricted to one main theme: it considers a very large spectrum of environment-related topics. In a similar fashion as for green preferences, PEBs are rarely investigated using real individual actions, leading to significant misreporting and potentially biasing results of existing studies (Kormos & Gifford, 2014); see Blankenberg & Alhusen (2019) for a review of PEBs measures.

Findings in the environmental sociology literature do show overall a gender gap in PEBs, with women engaging in more environmentally-friendly behaviors (Hunter *et al.*, 2004; Kennedy & Kmec, 2018; Dzialo, 2017), although this is not robust to all specifications of PEBs. Particularly, the classification of PEBs as private sphere (e.g. energy use at home or purchase of sustainable goods) and public sphere (e.g. petitioning, protesting for climate, supporting green policies) behaviors lead to different conclusions. We therefore contribute to the literature by investigating this gender gap in PEBs both in the private sphere (purchase of an environmentally-friendly car) and public sphere (support for green policies), by using real data instead of stated PEBs.

[1] This classification of measures of environmental concern is based on McCright & Xiao (2014).

Outline. The remainder of the paper is structured as follows. In the next section, we provide details on the data used in the paper to estimate preferences and proxy PEBs. Section 3 provides the microeconomic foundation and the identification strategy for our individual measures of green preferences and SWB. In Section 4 we investigate the individual well-being costs associated with environmental concern, and differences in green preferences across gender. Section 5 documents the existence of gender gaps in PEBs. Section 6 concludes.

2 Data

We detail in this section the creation of the Twitter dataset we use to measure green preferences and estimate SWB. We also introduce the municipal-level data used to proxy PEBs.

2.1 Individual Twitter data

Collection. Created in 2006, Twitter is a leading social media in the world with close to 230 million daily active users. Twitter users voluntarily share their thoughts and opinions on a wide range of cultural, political, and societal topics in their tweets. We create our sample of individuals by collecting tweets between January and June 2019 using the academic track of Twitter's Application Programming Interface (API) without restrictions on the type of tweets.[2] Given that geo-localized tweets represent only 1 to 5% of the total universe of tweets (Schlosser *et al.*, 2021), we do not restrict the data collection to tweets that are geo-localized. We select tweets that are either geo-localized in Sweden, or that are written in Swedish.[3] We then keep in our sample accounts of private individuals for which we could identify the gender and their municipality in Sweden.

[2] Tweets, retweets, mentions and replies are collected.

[3] An additional requirement is necessary by construction of the Twitter API. We thus collect precisely all the tweets that are written in Swedish *and* that contain one of the most common 50 words in the Swedish language. The list of words is provided in Appendix A, and is based on word frequency in movie subtitles as provided by Hermit Dave, URL: www.101languages.net/swedish/most-common-swedish-words/, visited last on 15/07/2021.

Gender attribution of users is based on their name on Twitter. We exploit the Swedish National Statistics Agency (SCB)'s list of usual names given to newborns between 1998 and 2020 and the US Social Security Administration (SSA)'s list of registered baby names in the United States since 1880 to classify names' most attributed gender. We also use individuals' self-entered location to map them to a municipality in Sweden, using geographical data from the Humanitarian OpenStreetMap Team (HOTOSM), which links 68 thousand populated places (e.g. villages, isolated dwellings, cities) to a Swedish municipality. We provide more information on the algorithms used to determine individuals' gender and municipality in Appendix A. We keep users belonging to the middle 96% of the total number of tweets distribution over the period and drop duplicate tweets to account for the potential presence of bots in our data.

Descriptive statistics. Our final sample is composed of 54 740 users, for a total of 4 371 825 tweets. Table 1 provides statistics on the number of users, tweets, and tweets per user for each gender. Our sample is characterized by a gender imbalance: there are 2.1 times more men than women, and the former write more tweets on average than the latter. This gender imbalance is consistent with the demographics of Twitter users globally as, as of October 2019, the men-to-women ratio in the global population of Twitter users was of 1.9 (Kemp, 2019). The population of Twitter users is indeed usually found not to be representative of the general population. Mellon & Prosser (2017) find it to be younger, gender-imbalanced, and more educated than a nationally representative sample of individuals in the UK, while Blank & Lutz (2017) do not find the level of education as being decisive for Twitter adoption. Although the lack of representativeness of Twitter data should be kept in mind when drawing conclusions based on individuals' tweets, it should not be a major concern in our context of green preferences measurement, comparison across gender, and investigation of SWB differences between climate-concerned and unconcerned individuals as long as male and female Twitter users are comparable.

Table 1: Summary statistics

	Female users				Male users			
	Mean	Sd	Min-Max	Median	Mean	Sd	Min-Max	Median
Tweets per users	74.1	135.8	2-951	21	82.61	149.31	2-953	22
Number of users	17 672 (32.28%)				37 068 (67,72%)			
Number of tweets	1 309 487 (29.95%)				3 062 338 (70.05%)			

Notes: 54 740 users in total, for a total of 4 371 825 tweets collected from January to June 2019.

2.2 Municipal level data

Registered vehicles. The *Vehicles in Use by Region and Type of Vehicles Statistics* from the SCB provide us with the number of cars registered in a municipality (either privately owned cars, or taxis and company cars), classified by their type of fuel. Cars belong to one of the 8 following categories: gasoline, diesel, electric, hybrid, plug-in hybrid, ethanol, gas, and others. We compute for each municipality the share of environment-friendly cars among all the cars registered in a municipality:

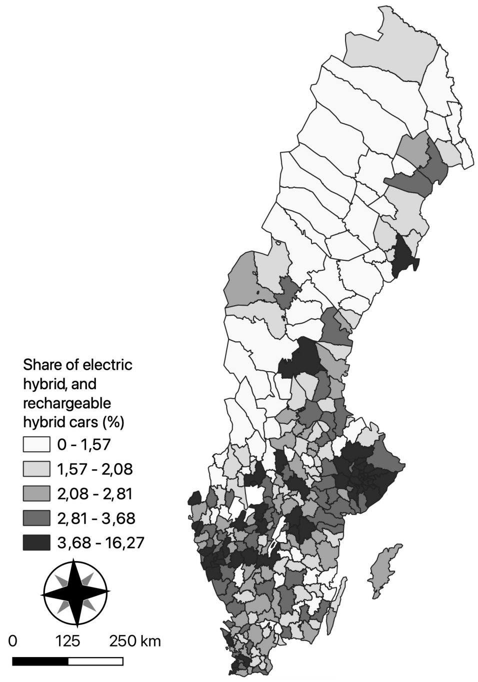
$$EFC_m = \frac{\#electric_m + \#hybrid_m + \#plugin_m}{\#cars_m}, \tag{1}$$

where m denotes the municipality. We opt for a strict definition of environment-friendly cars since we include only electric, hybrid, and plug-in hybrid cars. We however also checked the robustness of our results by including cars running on natural gas. Figure 1a depicts the share of environment-friendly cars in the 290 Swedish municipalities as defined in Equation (1), in 2019. The south is clearly characterized by higher shares of environment-friendly cars, while the shares in the north are amongst the lowest. In the analysis, we control for the density of population and for county fixed effects to account for unobserved regional differences in the shares of environment-friendly cars due to geographical components.

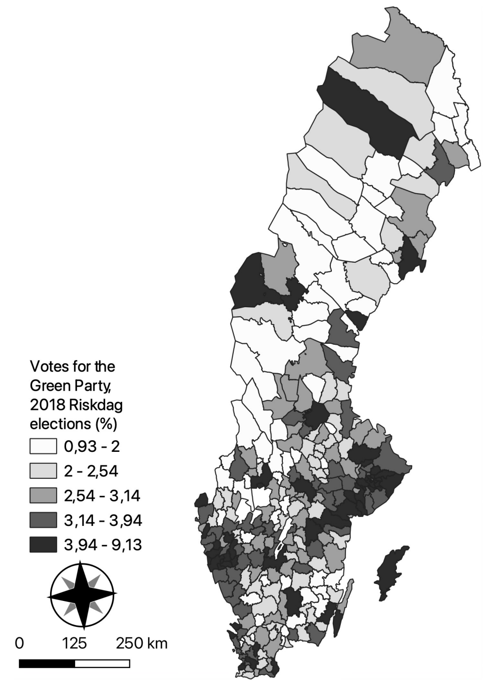
Votes for the Swedish Green Party. The SCB also provides us with voting data in the parliamentary elections of 2018 aggregated at the municipal level. Parliamentary elections are the national elections in Sweden and voter turnout

Figure 1: Geographical distribution of two PEBs in Sweden, municipal level.

(a) Share of environment-friendly cars



(b) Share of votes for the Green Party in the 2018 Riksdag elections



Notes: Categories given by the quintiles of respective distributions.
Source: Car registration and election results data from the SCB.

in 2018 was 87.18%. Figure 1b depicts the geographical distribution of the share of votes dedicated to the Green Party at the municipal level. As for the share of green cars, we observe quite some heterogeneity throughout the country with higher shares of votes for the Green Party in the south.

Controls. Additional municipal-level variables are collected from the SCB to account for potential confounding factors in the analysis of the relationship between our measure of green preferences and PEBs. We collect data on the median annual income of inhabitants, the density of population, the share of individuals between 20 and 35 years old, the share of men and women, and the share of company cars and taxis registered in the municipality.

3 Identification

Measurement of pro-environment preferences and sentiments is often based on answers to surveys or face-to-face interviews. We provide here a measure of preferences for the conservation of the environment using the unsolicited opinions of individuals shared on the social media platform Twitter. We start by providing a micro-foundation of our elicitation of preferences. Subsequently we present more details on our identification of green preferences and SWB.

3.1 Elicitation of preferences

Micro-foundation. The measure of individuals' pro-environmental preferences proposed in this paper relies on the topics discussed in their tweets and adopts a close-to-revealed preference approach. Assume a finite set of topics to be potentially discussed $\mathcal{T}$. Consider individual $i \in \{1, ..., N\}$, with set of tweets $\mathcal{D}_i = \{d_i^1, ..., d_i^{D_i}\} \subset \mathcal{D}$, and $\mathcal{D}$ the universe of tweets. The function $e : \mathcal{D} \times \mathcal{T} \rightarrow \{0, 1\}$, takes value $e(d, t)$ =1 if a tweet $d \in \mathcal{D}$ is related to topic $t \in \mathcal{T}$, 0 otherwise. Define as S_{it} the share of tweets individual i dedicates to topic t:

$$S_{it} = \frac{\sum_{d \in \mathcal{D}_i} e(d, t)}{D_i}. \tag{2}$$

An individual $i \in \{1, ..., N\}$, who chooses to tweet about topic $t_1 \in \mathcal{T}$ rather than $t_2 \in \mathcal{T}$ does so if the utility she gets from discussing t_1 is higher than the utility from discussing t_2. Importantly, we assume here that discussing topic t_1 more than topic t_2 is equivalent to preferring t_1 over t_2, e.g. an individual who spends all his time on Twitter discussing sport and never discusses music is assumed to prefer sport over music. An individual thus chooses $S_{it_1}, S_{it_2}, ..., S_{it_k}$ such that the utility he gets from using Twitter is maximized:

$$\max_{S_{it_1}, S_{it_2}, ... S_{it_k}} u(S_{it_1}, S_{it_2}, ... S_{it_k})$$

$$s.t. \quad S_{it_1} + S_{it_2} + ... + S_{it_k} = 1.$$

We base our green preference measure on the following assumption: if $S^*_{it_1} \geq S^*_{it_2}$, for $t_1, t_2 \in T$ and $t_1 \neq t_2$, then $t_1 \succcurlyeq t_2$. In words, if the chosen share of tweets dedicated to topic 1 is higher than the chosen share of tweets dedicated to topic 2 for individual i, then she reveals to prefer topic 1 over topic 2.

Climate scepticism. The method used to identify individuals with environmentally friendly preferences relies on the assumption that anyone discussing environment-related topics in their tweets cares about the environment. We relax this assumption in the context of green preferences in Appendix B, by considering the possibility of individuals discussing environmental topics on Twitter being climate-change deniers. Tweets denying the existence, anthropological causes, or consequences of climate change and environmental problems are identified using support vector machines (SVM), a classification machine learning algorithm, and the analyses in the paper are replicated dropping from the sample individuals who are potentially climate skeptic. On the randomly selected 2000 tweets related to environmental issues that have been manually classified to train the algorithm, only 3.15% have been found to criticize environmental policies or negate the anthropological aspect of climate change. Appendix B shows that all the empirical results of Sections 4 and 5 are robust to removing potentially climate skeptic individuals from the sample.

3.2 Identification of environment-related tweets

Identification. Our green preference measure is based on the frequency with which each individual discusses environmental issues. We classify tweets as being environment-related or not based on whether they include a uni- or bigram in connection to an environmental issue. We therefore created a list of 82 unigrams and bigrams that belong to one of the following categories: causes of climate change (e.g. greenhouse gas, pollution), consequences on the environment (e.g. oceans acidification, mass extinctions), and potential solutions to climate change (e.g. green energy, circular economy). Our list of environment-related keywords is based on global warming glossaries (see for example BBC's Climate Change glossary (2014), or UCDavis's Climate Change Terms and Definitions), as well as on Kurisu (2015)'s list of PEBs.[4] Tweets containing at least one of the keywords from the dictionary are classified as environment-related. Table 2 provides the unigrams and bigrams in the climate dictionary.[5]

Green preferences in Sweden. Among the 54 740 individuals in our dataset, 25.26% chose to dedicate a non-zero share of their tweets $S_{i,env}$ to environmental issues: 25.26% of individuals thus reveal to care about the conservation of the environment. Considering only individuals who at least once referred to an environmental issue in their tweets (13 825 individuals), Figure 2 shows the distribution of the share of tweets dedicated to environmental issues. $S_{i,env}$ can also be interpreted as a coefficient of strength of environmental preferences. The average share of tweets dedicated to environmental issues is 7.76%, and $S_{i,t_{env}}$ is below 22% for 90% of users with green preferences.

3.3 Identification of subjective well-being

Our measure of individual SWB is based on individuals' expressed level of happiness in their tweets. Doing so, we follow the recent literature that uses the experienced preferences method using social media data. This literature investigates the relationship between individuals' SWB or life satisfaction and non-marketable goods to elicit preferences for these goods, using social media data to measure SWB (see for example Baylis (2020) or Loureiro *et al.* (2022)). Following this new literature, we use the expressed happiness in a tweet measured by using natural language processing (NLP) methods to proxy SWB. We build our individual estimates in three steps.

[4] See, respectively, https://www.bbc.com/news/science-environment-11833685 and https://climatechange.ucdavis.edu/science/climate-change-definitions/

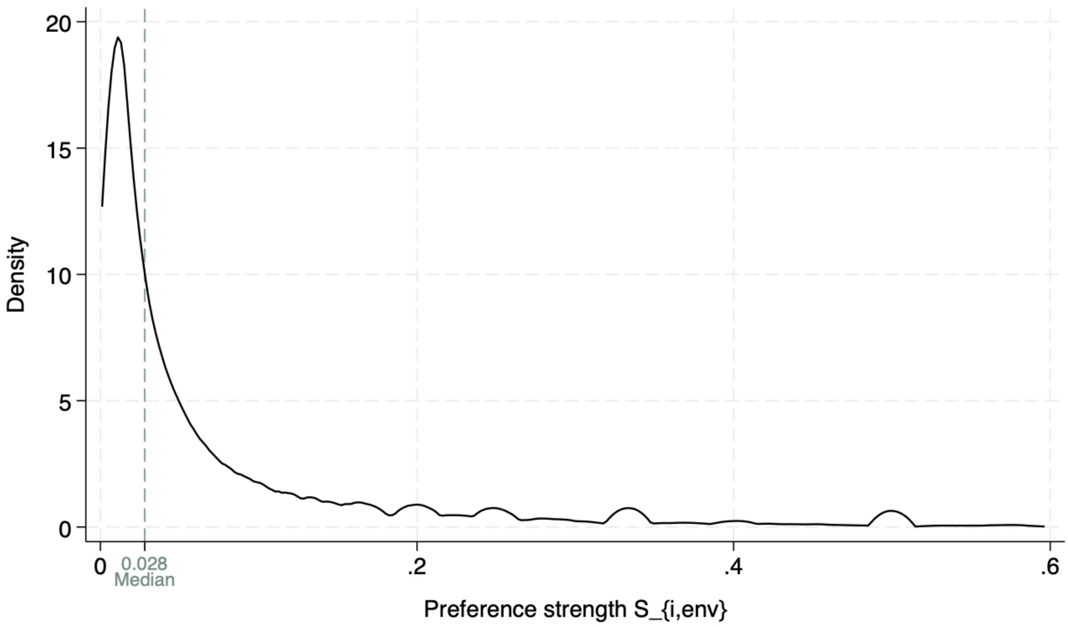
[5] We make the choice of excluding the Swedish Green Party from the set of keywords we use to identify environment-related tweets for two reasons. First, one might discuss politics and the Green Party on topics unrelated to the environment. Second, there would be a large probability of reverse causality in Section 5 when investigating the relationship between green preferences as measured with the tweets, and the share of votes obtained by the Green Party in the 2018 Riksdag elections.

Table 2: Environment-related unigrams and bigrams

Theme	Topic	Unigrams and bigrams
Causes	GHGEs	Greenhouse gas, GHG, emissions, carbon dioxyde, fossil fuel, carbon footprint, carbon cycle.
	Energy	Energy consumption, EACOP.
	Pollution	Pollution, plastic waste, waste generation, polluter, polluting.
	Over-exploitation	Deforestation, overfishing, overconsumption.
	Others	Greenwashing.
Consequences	Climate change	Climate change, climate, climatic, global warming, global average temperature, extreme weather, natural disaster.
	Sea levels	Sea level, melting ice, melting glaciers, ice loss.
	Biodiversity	Biodiversity, mass extinction, ocean acidification, extinction, ecosystems, threatened species, endangered species, wildlife.
	Resources	Water scarcity, resource scarcity.
	Others	Ozone depletion.
Solutions	Energy transition	Energy mix, energy transition, renewable energy, biofuels, clean energy, green electricity, solar panel, green energy, green transition.
	Mitigation	Mitigation, carbon neutrality, net-zero target.
	Sustainable Development	Sustainable, sustainability, organic farming , eco-tourism, circular economy.
	Reforestation	Reforestation, plant tree, planting tree.
	Institutions	IPCC, UNFCCC, kyoto protocol, green deal, Paris agreement.
	Activism	#Fridaysforfuture, #protectnature, #peacewithnature, climate strike, climate march, climate action, greenpeace, wwf, #nomoreemptypromises, #youth2030.
	Others	Recycling, ecology, organic food, waste sorting, eco-friendly, eco-label, green behavior, green consumption.

Source: Own classification of climate-related unigrams and bigrams, from climate change glossaries such as the BBC's (see www.https://www.bbc.com/news/science-environment-11833685) or UC-Davis' (see www.https://climatechange.ucdavis.edu/science/climate-change-definitions/), both last visited on the 04/08/2021.

Figure 2: Distribution of preference strength $S_{i,env}$ for individuals with green preferences



Notes: Kernel density estimate, bandwidth=0.0061. For clarity, the graph shows the distribution of individuals' green preference strengths for individuals with a strictly positive share of tweets dedicated to the environment, lower than 60% of their tweets. Individuals with a share of environment-related tweets higher than 60% represent less than 1% of the sample.

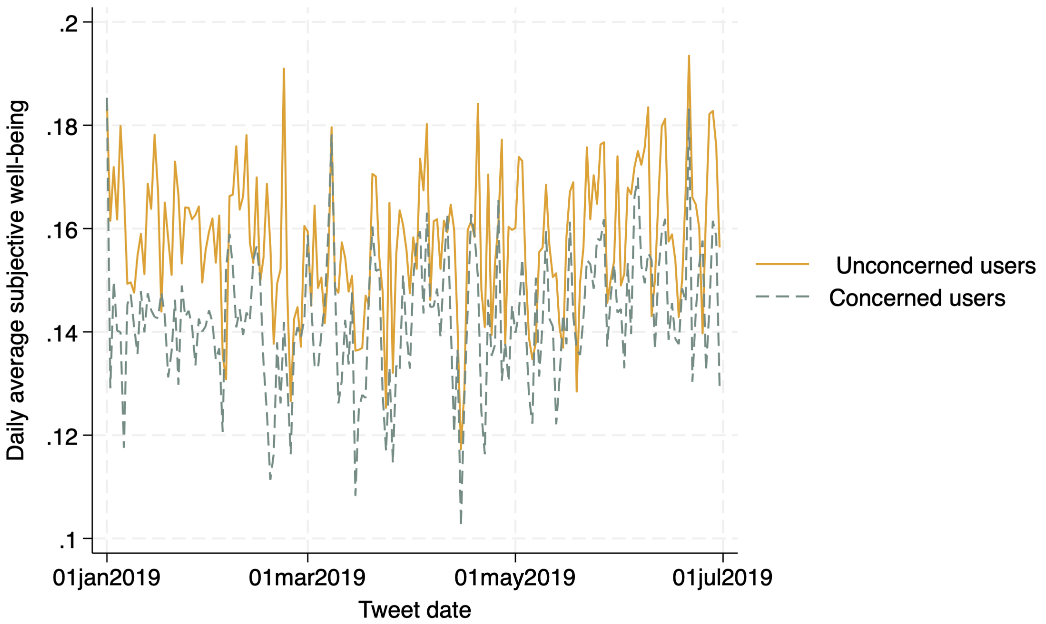
Sentiment analysis. Sentiment analysis is the process of identifying and extracting subjective information from textual data. It uses various techniques from NLP to analyze the tone, attitude, and emotions expressed in the text. The main goal of sentiment analysis is to determine the overall sentiment of a piece of text: whether it is positive, negative, or neutral. We base our SWB estimates on the VADER (Valence Aware Dictionary and sEntiment Reasoner, Hutto & Gilbert (2014)) algorithm, a rule-based sentiment analysis tool designed specifically for social media content. It uses a lexicon of words and phrases (including emojis and slang) with pre-determined sentiment scores to analyze the sentiment of a given text. The algorithm takes into account not only the sentiment of individual words but also the context and grammar of the text. The algorithm outputs a sentiment score ranging from -1 to +1, where -1 indicates extremely negative sentiment, +1 indicates extremely positive sentiment, and 0 indicates neutral sentiment. We denote this measure for tweet d by s_d. We run the VADER algorithm on the English translation of individuals' original tweets. We thus drop from the sample individuals whose only activity on Twitter is retweeting.

Daily SWB. In this step of the SWB estimation, we average the sentiment in each individual tweet daily. Importantly, tweets related to the environment are not considered in the estimation of SWB, to avoid it being biased by negative expressed sentiments in green tweets. We also drop from the sample days during which individuals tweet less than 3 times to avoid our SWB estimates being based on the tone of too few tweets. The average SWB of individuals i on day m is thus given in Equation (3) by the average expressed sentiment in all the tweets of i written on this day, $D_{i,m}$:

$$SWB_{i,m} = \frac{\sum_{d \in \mathcal{D}_{i,m}} s_d}{D_{i,m}}. \tag{3}$$

Figure 3 shows that, despite the huge variation in the daily average of individuals' SWB, there seems to be a clear difference between individuals who reveal to be concerned about the environment and those who do not.

Figure 3: Daily average SWB of individuals with and without green preferences.



Notes: Average daily SWB, proxied by the expressed sentiment of individuals revealing to care about the environment (concerned), and those not revealing to care (unconcerned). The expressed sentiment is computed using Vader sentiment algorithm.

Average daily SWB. To get an individual's average SWB on the period, we take the average daily SWB on the number of days between January and June 2019 for which an individual writes at least 3 tweets:

$$SWB_i = \frac{\sum_{m \in M_i} SWB_{i,m}}{M_i}, \tag{4}$$

with M_i the total number of days. We keep individuals for which at least 7 such days are available in the sample. This leads us to a sample size of 9786 individuals. Table 3 provides summary statistics for these individuals and demonstrates there is enough variation in the data to meaningfully conduct our empirical exercise.

To recall, we use our measure of environmental preferences in two exercises, at two different levels of aggregation. First, at the individual level, we investigate the potential existence of SWB costs of environmental concern, as well as the

Table 3: Summary statistics

	Mean	Sd	Min-max	Median
Tweets per users	299.89	210.94	29 - 953	231
Individual SWB	0.15	0.13	-0.53 - 0.80	0.14
Share of green tweets	0.026	0.052	0 - 0.69	0.01
Green preferences	54.33%			
Female users	27.21%			

Notes: This sample is the result of restricting the main sample to individuals with at least three tweets unrelated to the environment per day for at least 7 days in the period January to June 2019. Individual SWB refers to the individual average SWB on the period as given in Equation (4). Share of green tweets refers to the share of tweets dedicated to the environment for individuals with green preferences. Green preferences refers to the share of individuals with green preferences (a positive share of environment-related tweets). Female users refers to the share of female users.

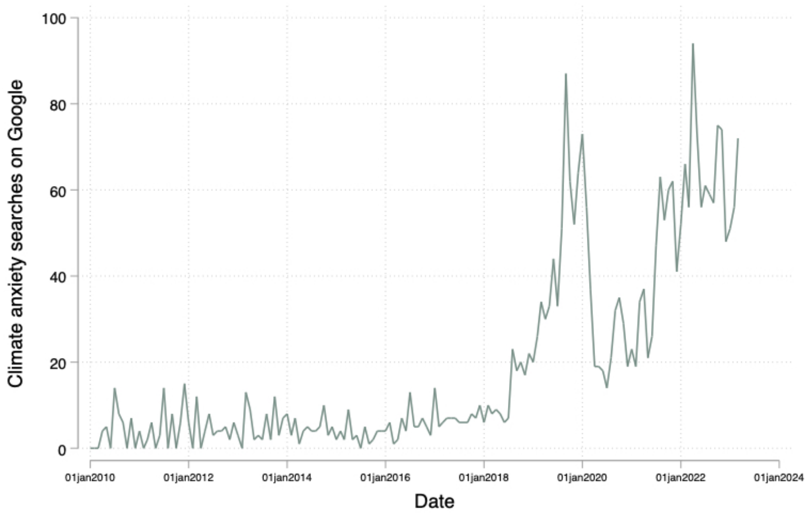
existence of a gender gap in green preferences. Second, we make use of municipal-level data on PEBs to test the gendered environmental responsibilization finding in the literature, that not only do women have stronger green preferences than men, but they also act more on these preferences than the latter.

4 Green preferences and subjective well-being

We use our individual measures of green preferences and SWB to shed further light on two empirical questions that got quite some attention in their respective literatures. First, a rising body of literature emphasizes the potential well-being effects of climate change. Individuals concerned about climate change, and aware of its potential consequences, increasingly experience stress and anxiety. To illustrate this, Figure 4 shows the number of Google searches for "climate anxiety", which has significantly increased in the last four years. In addition to being a concern for individuals' well-being *per se*, lower well-being for individuals with environmental preferences could also translate into an increase in well-being inequalities across genders given the gender gap in green preferences the environmental sociology and this paper shed light on. We make use of our individual-level preference measure to investigate the potential SWB cost associated with environmental concern by comparing individuals with and without green preferences.

Second, we test the finding from the environmental sociology literature that women are more concerned about the environment than men. Our green preferences measure being the fruit of individuals' revealed concern for the environment on social media platforms, we contribute to the literature on gender and environmental attitudes which is mostly based on stated preferences. Preferences at the extensive and intensive margins are considered respectively by investigating the share of men and women in our sample that ever reveal being concerned about environmental issues (extensive), and, focusing on individuals with revealed green preferences, by investigating whether men or women appear to care more (intensive).

Figure 4: Monthly number of searches for climate-anxiety on Google, 2010-2023



Source: Google trends.

4.1 Individual subjective well-being

Empirical results. We rely on ordinary least square estimates of the following relationship between average SWB and green preferences:

$$SWB_i = \alpha + \beta_1 GP_i + \beta_2 S_{i,env} + \beta_3 S_{i,env}^2 + \gamma X_i + \epsilon. \quad (5)$$

SWB_i refers to individuals' average well-being on the period as computed in Equation (4), GP_i is a binary variable equal to 1 if individual i has a positive share of tweets dedicated to the environment and $S_{i,env}$ refers to the individuals' share of green tweets. We include the quadratic term $S_{i,env}$ to capture potential non-linearities in the relationship between SWB and the intensity of green preferences. X_i is a vector of controls, which includes users' gender and their tweeting behavior (e.g. total number of tweets on the whole period). Controlling for the former is crucial as men are found to be less expressive than women on Twitter, resulting in lower SWB when computed using sentiment analysis. In the same fashion, the number of tweets may impact the SWB measure, making tweeting behavior an important control for the analysis. Month and municipality fixed effects are included, to control for seasonality in SWB and potential communication differences across municipalities.

Table 4 presents our results.[6] At the extensive margin, individuals who ever discuss the environment on Twitter have on average a mean daily SWB on the period that is 10.67% lower than individuals who never discuss environment-related topics. At the intensive margin, focusing now only on individuals with green preferences, the results indicate a positive relationship between the share of tweets an individual dedicates to the environment and her SWB for low shares of green tweets. This relationship however becomes negative for shares of environment-related tweets beyond 40.73% of the total number of tweets written by the individual. The decrease in SWB associated with being a man is expected and in line with the literature as explained above.

Discussion. The results above suggest that individuals who reveal to care about the environment in our sample are characterized on average by a lower SWB than individuals who do not, and, stronger green preferences at the intensive margin are associated with lower well-being for individuals with high environmental concerns.

[6] See Table 13 in Appendix C for all estimation results.

Table 4: Relationship between individuals' SWB and green preferences.

	(1) Average daily SWB
Env.-concerned	-0.016***
	(0.003)
Env.-concerned=1 × Share related	0.457***
	(0.071)
Env.-concerned=1 × Share related squared	-0.561***
	(0.191)
Male user	-0.026***
	(0.003)
Controls	Yes
Month FE	Yes
Municipality FE	Yes
Observations	9,786
R-squared	0.054

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the average SWB of individual i on the period. Env.-concerned is a binary variable equal to 1 if the individual reveals to care about the environment on Twitter. Share related refers to the share of tweets each individual dedicates to the environment. Male users is a binary variable equal to 1 if the individual is a man. Other controls include: the number of tweets written by the user and the number of days during which she wrote at least 3 tweets during the period.

However, we cannot talk about a negative impact of environmental concern on well-being, as the direction of causality from concern to lower expressed sentiment is unclear: awareness of environmental degradation, feeling powerlessness in front of climate change consequences, or the fear of rising uncertainty regarding one's future well-being could be leading individuals caring about the environment to have a lower SWB. However, individuals who have a lower level of SWB in general could also focus mainly on negative outcomes such as climate change, driving the difference in SWB between the two groups of individuals. We leave it to future research to identify the causal direction of this relationship and to quantify precisely the SWB cost of environmental concern and its potential consequences.

4.2 Gender gap in green preferences

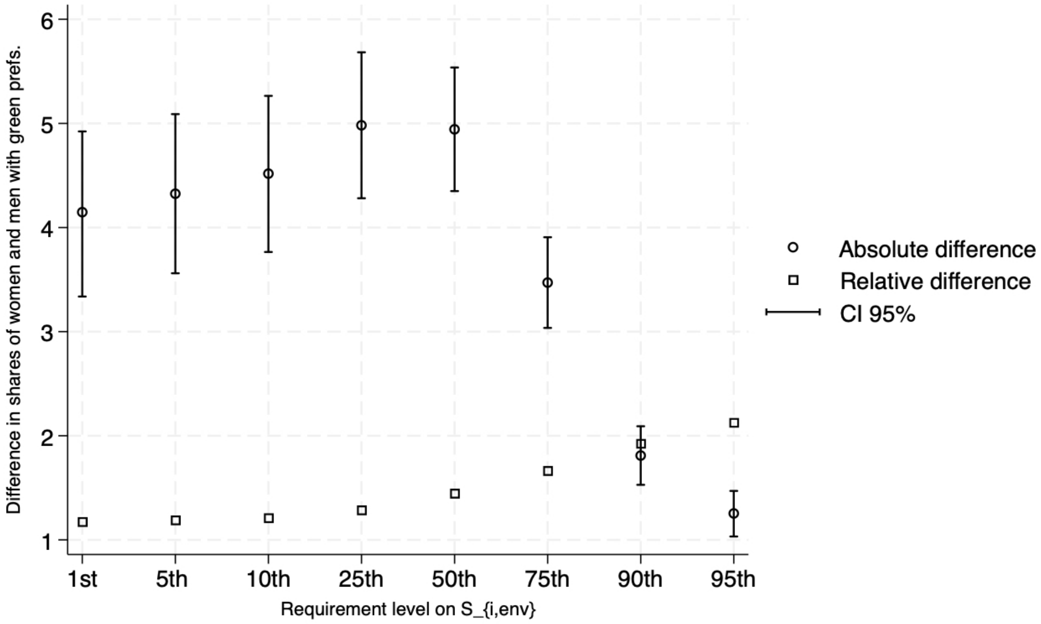
Extensive margin. The share of women choosing to dedicate a positive share of their tweets to environment-related topics out of all women in our sample is 28.05%, while the number of men discussing environmental issues in their tweets represents 23.92% of all men in our sample, indicating that women are more likely to care about the state of the environment than men. To show that this conclusion is robust to including some proxy of intensity, we present in Figure 5, the difference in percentage points for different minimum thresholds α on the share of tweets $S_{i,env}$ that must be dedicated to the environment.[7] For instance, if α is given by the 95th percentile, then an individual reveals to care about the state of the environment only if the share of related tweets belongs to the top 5% of the $S_{i,env}$ distribution.

We can conclude from this exercise that the superiority of the share of women concerned about environmental issues with respect to men holds for any level of minimal threshold for $S_{i,env}$, and the confidence intervals show that this difference is always statistically significant. We also observe an upward pattern in the relative difference in the shares for our different levels of α. This shows that the decrease observed in the gender difference in absolute terms after the 25th

[7]More precisely, $GreenPref_g = \frac{\sum_{i \in n_g} \mathbb{1}_{(S_{i,env} > \alpha)}}{n_g}$, with $g = (f, m)$ and n_g being the number of users of gender g in our sample.

percentile requirement is driven by the lower shares of men and women with such high shares of environment-related tweets, and is not due to a reduction in the gender gap in green preferences.

Figure 5: Extensive margin.

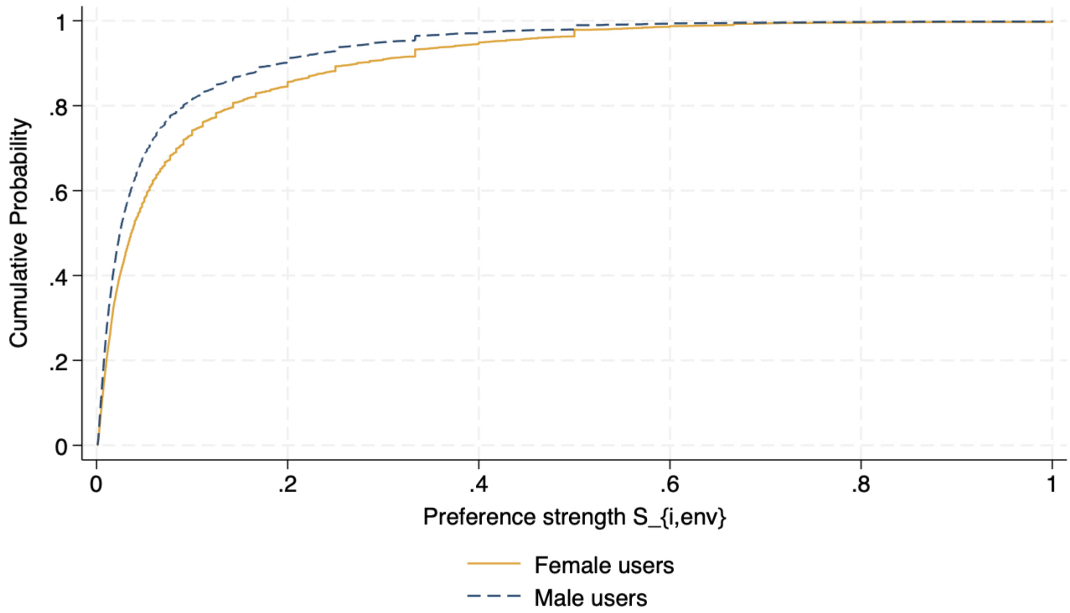


<u>*Notes*</u>: Difference between shares of women and men with green preferences for different levels of requirements on the share of tweets dedicated to the environment. Absolute differences are given in percentage points and indicated with dots with their confidence intervals at the 95% confidence level. The relative differences are computed as the ratio of the two shares and represented by squares. Requirement levels on $S_{i,env}$ are given by the percentiles of $S_{i,env}$'s distribution.

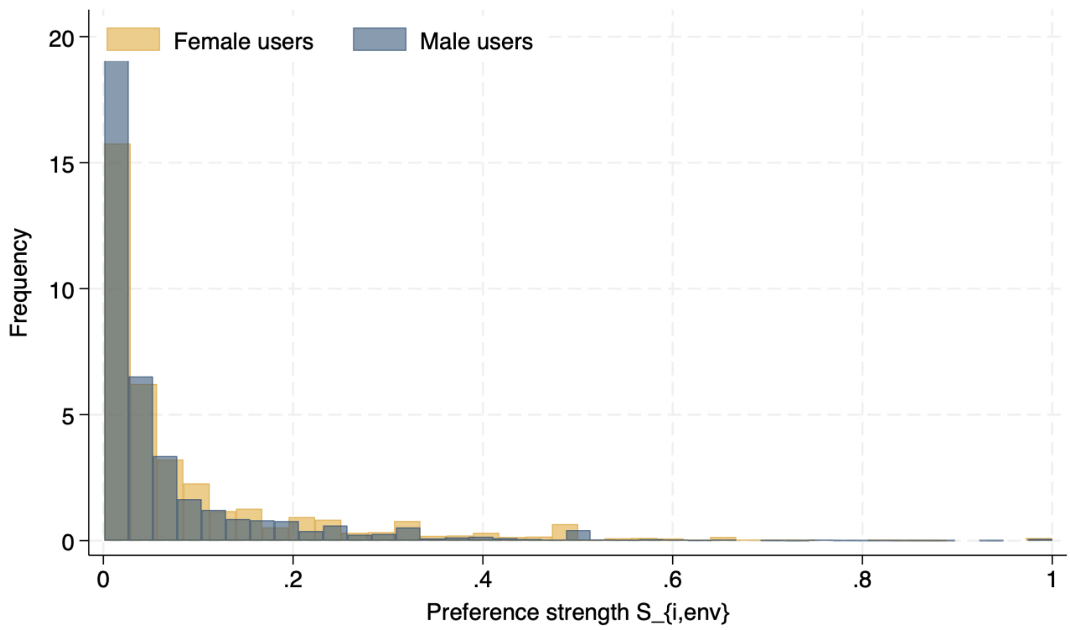
Intensive margin. We consider in this section only individuals who at least once referred to an environmental issue in their tweets (13 825 individuals) and compare individuals' green preference strength $S_{i,env}$ across genders. A first indicator of potentially stronger environmentally-friendly preferences for women than men is their average and median share of tweets in relation to the environment: men dedicate on average 6.8% of their tweets to environmental issues, for a median share of 2.44%, while these shares are respectively 9.5% and 3.64% for women. Women's average (resp. median) green preferences strength is thus 2.7 (resp. 1.2) percentage points higher than men's, and the difference in averages is statistically significant.

Figure 6 depicts the distributions and cumulative distributions of individuals' green preferences strength across genders, as measured by the share of tweets each individual dedicates to environment-related topics. The dominance of women's intensity of concern over men's is clearly illustrated in Figure 6a, depicting the cumulative distributions of men's and women's preference intensity respectively. At each level, the share of women with *at most* this preference strength is lower than the share of men, or, in other words, the share of women having *at least* this green preference strength is higher than men's one. This is further illustrated in Figure 6b, showing that the mass of male users with low preference strengths is higher than that of female users, while female users are more represented than male users for preference strengths larger than 10%.

Figure 6: Intensive margin.



(a) Cumulative distribution



(b) Distribution

Notes: Panel (a) depicts the cumulative distribution of males' and females' share of tweets related to the environment (i.e. $S_{i,env} > 0$). Panel (b) depicts the distribution of these $S_{i,env}$ for men and women in the sample.

Discussion. Our findings corroborate the conclusions of the environmental literature that women have stronger green preferences than men, both at the extensive and intensive margins. Relatively more women dedicate a positive share of tweets to the environment than men, and, comparing only men and women who do discuss environment-related topics, women dedicate a higher share of their tweets to the issue than men. Given our green preference measure, this translates into relatively more women having preferences for the conservation of the environment than men, and into their preferences being stronger than men's. Linking this to our results of the previous subsection, this therefore also leads to a gender gap in SWB.

We underline two main potential threats to our results here. First, and as underlined in Section 3.1, we make the rather strong assumption that all individuals who talk about green topics, e.g greenhouse gas emissions, energy transition, or renewable energy actually care about the environment.[8] To make sure that climate-skeptic tweets do not bias our results, we use supervised machine learning to identify climate-skeptic individuals in our sample and replicate the analysis of green preferences across genders in Appendix B. Among the individuals identified as potentially being climate skeptics, 69.56% are men, and 30.44% are women. Importantly, our findings are robust to removing potential climate skeptic individuals from our sample.

The second potential threat to our results would be the existence of systematic communication differences across genders on Twitter. Particularly, given that men write more tweets than women on average, the lower share of their tweets dedicated to environmental issues might be mechanical. Or, women might tend to talk only about things they know more about, explaining why women who discuss environmental issues discuss it more frequently than men. To rule out this second threat, we replicate the analysis in this section focusing on two other topics: immigration, and general politics. Results are provided in Appendix C, and no similar pattern across genders is found for these topics as for environment-related topics.

[8]A famous example of a climate change-denying tweet is Donald Trump's tweet, posted in October 2015 saying "It's really cold outside, they are calling it a major freeze, weeks ahead of normal. Man, we could use a big fat dose of global warming!"

5 Gender gap in pro-environmental behavior

Next to this individual analysis of green preferences and SWB, an obvious, and probably more important, question is then how this translates into actual pro-environmental behaviors. Findings in the environmental sociology literature show overall a gender gap in PEBs, with women engaging in more environmentally-friendly behaviors (Hunter *et al.*, 2004; Dzialo, 2017; Kennedy & Kmec, 2018). However, as mentioned in the Introduction, these findings depend crucially on being in the private or public sphere: women are in general found to partake in more private-sphere PEBs than men, while the literature is inconclusive regarding the existence of a gender gap in public sphere green behaviors.[9]

Testing the relationship between green preferences and PEBs at the individual level requires being able to match individual-level pro-environmental behavior data to our individual-level measure of green preferences. Unfortunately, such information is mainly available in surveys that are not linked to the respondents' Twitter accounts. As the second-best alternative, we therefore aggregate our measure of green preferences at the municipal level and investigate its conditional correlation with aggregated real data on the share of environmentally-friendly cars registered (i.e. private sphere PEB) and the share of votes for the Green Party in municipalities (i.e. public sphere PEB). Before discussing our empirical results, we briefly introduce our empirical strategy.

5.1 Empirical strategy

Data and robustness checks. We aggregate our measure of green preferences as given in Equation (2) to get the share of users of gender $g = (f, m)$ in the municipality m of population size N_m, who have a share of tweets dedicated to the environment $S_{i,env}$ higher than α:

[9] Mohai (1992) and Yates *et al.* (2015) find than men engage more than women in public sphere PEBs, while Hunter *et al.* (2004) and Xiao & McCright (2015) do not find significant difference across genders.

$$GP_{g,m} = \frac{\sum_{i \in N_{g,m}} \mathbb{1}_{(S_{i,env} > \alpha)}}{N_m}. \tag{6}$$

Our baseline specification is given by α equal to 0. In this case, any user who discusses at least once the environment in his tweets is considered to reveal preferences about the conservation of the environment. Figures 7a and 7b present the share of women and men with green preferences in the 290 Swedish municipalities. There is clearly still a lot of heterogeneity both across the country and across genders to meaningfully apply our analysis at the municipality level.

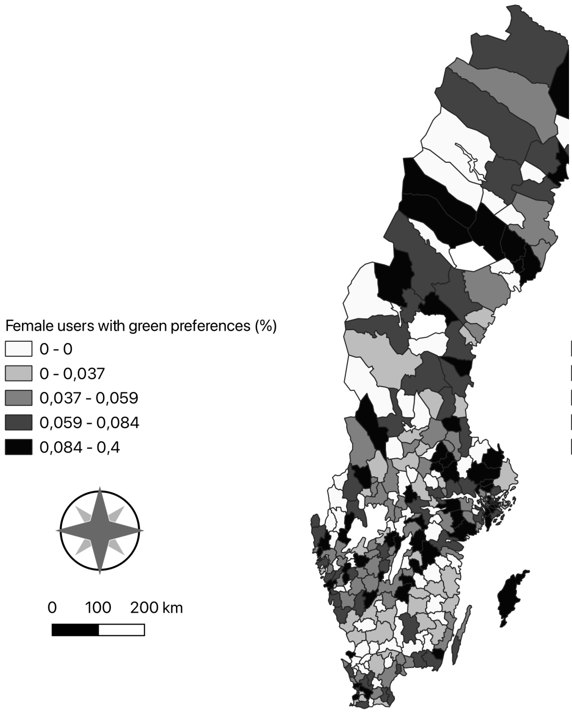
Next to our robustness exercises related to removing climate skeptic individuals from the sample, we have two more sets of robustness checks. First, we present in Appendix C the results of our empirical exercises for α equal to 0, 0.028 (which is the median preference strength for individuals with green preferences), and 0.05. As such we impose stricter requirements on individuals' tweeting behavior to be considered as caring about the environment. Next, a potential confounding factor in the analysis of the relationship between green preferences and PEBs in a municipality is the presence of local networks. Assuming the existence of local networks, it is indeed more likely that an individual will retweet or react to a tweet of another user in the network discussing environmental issues. In Appendix C, we therefore also replicate all the empirical exercises by dropping retweets from our sample. All the results in both appendices are in line with those discussed in the main text.

Empirical specification. In the following sections, we rely on ordinary least square estimates of the following relationship between green preferences and PEBs across genders:

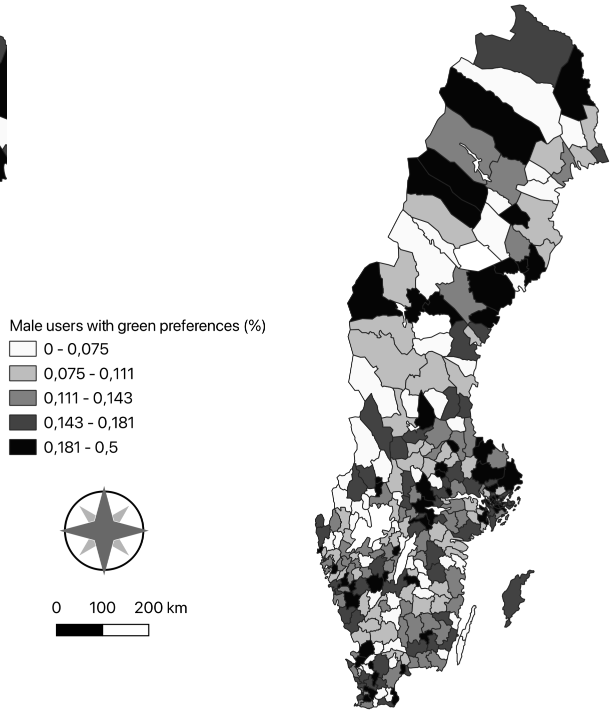
$$PEB_m = \alpha + \beta_1 GP_{F,m} + \beta_2 GP_{M,m} + \gamma X_m + \epsilon. \tag{7}$$

PEB_m refers to our municipal-level PEBs: either the share of votes for the Swedish Green party in municipality m in the 2018 Riksdag elections, or the share of registered electric and hybrid cars in the municipality in 2019. $GP_{g,m}$

Figure 7: Shares of men and women with green preferences in each Swedish municipality



(a) Female users



(b) Male users

Notes: Categories given by the quintile of the respective distributions. Shares are computed as the number of Twitter users (resp. female and male) in each municipality who ever talk about the environment in their tweets from 01/2019 to 06/2019, divided by the total number of users in the municipality.

denotes our measure of aggregated green preferences in municipality m, where $g = (F, M)$ denotes gender. X_m is a vector of controls at the municipal level. We control particularly for the median total annual earned income of men and women respectively aged between 20 and 65 years old in the municipality, and the share of young adults (between 20 and 35 years old) in the municipality as younger individuals are found in general to have stronger environmental preferences. We also control for the density of population and county fixed effects to account for geographical differences between municipalities (e.g. lowland versus mountains or rural versus urban area). By doing so, we account for a potential city effect on the choice of purchasing an electric car, and control for unobserved heterogeneity across counties. Finally, when investigating the share of green cars in a municipality, we also control for the gender division in the municipality to account for potential gendered preferences for electric and hybrid cars independent from environmental considerations, and the share of taxis and company cars in the municipality, since their type is out of individuals' control.

5.2 Green car purchases

Table 5 shows a subset of the estimation results for the relationship between aggregated environmental preferences of men and women in a municipality and the share of environment-friendly cars registered in the municipality for different specifications of the regression model.[10] Column 1 shows the baseline estimation results. Column 2 shows the estimation results for an alternative - softer - definition of green cars, including not only electric, hybrid, and plug-in hybrid cars but also cars working on natural gas. Finally, the last column of the table shows the estimated relationship between green preferences at the municipal level across genders and the share of more polluting cars in a municipality (namely diesel and gasoline cars).

It is clear from each column of the table that the relationship between green preferences in a municipality and the share of environment-friendly cars registered is much stronger for women than men. For instance, for the baseline specification,

[10]See Table 14 in Appendix C for all estimations results.

we get that an increase of one percentage point in the share of women with preferences for the environment in a municipality is related to a 0.023 percentage point increase in the share of green cars registered. To further interpret the size of this number, this is equivalent to the effect of an increase of 13 954 euros (164 286 SEK) in the yearly median income of men in the municipality. We also observe a strong and negative relationship between the share of women with green preferences and the share of polluting cars registered in a municipality: close to 1.5 times stronger than the estimated relationship between polluting cars and the share of men with green preferences. Our results thus clearly suggest that women act more in line with their environmentally-friendly preferences than men do when considering the purchase of a non-polluting car as a PEB.

In addition to the positive relationship between female concern and the share of green cars in a municipality, the share of young people and women in a municipality appears to be strongly positively correlated with the share of green cars. This finding indicates that either these people have intrinsic preferences for electric or hybrid cars independently of any environmental consideration, or, in line with our previous finding, the large estimated effect for the share of women in a municipality is due to environmental preferences not captured by our measure.

5.3 Swedish Green Party voting

The previous section confirms the findings of the literature on a gender gap in PEBs in the private sphere. As a final exercise, we also want to consider a public sphere setting, by replicating the analysis for the voting results for the Green Party in the 2018 Swedish Riksdag elections. Table 6 shows the estimated relationship between aggregated green preferences by gender in a municipality and the share of votes dedicated to the Green Party. We include the same controls as in the previous analysis, except for the share of company cars and taxis registered in a municipality.

Interestingly, the same pattern as for green cars is not observed in Table 6 for the relationship between the share of men with green preferences in a mu-

Table 5: Relation between male and female's green preferences and the share of green cars in Swedish municipalities.

	(1) Baseline	(2) + Gas	(3) Polluting
Green pr. Wom.	0.023***	0.023**	-0.028***
	(0.009)	(0.009)	(0.010)
Green pr. Men	0.011*	0.013*	-0.019**
	(0.006)	(0.007)	(0.008)
Income Women	0.007	0.011	-0.009
	(0.007)	(0.008)	(0.007)
Income Men	0.014**	0.013**	-0.011*
	(0.006)	(0.006)	(0.006)
% 20-35 yo	0.100***	0.160***	-0.205***
	(0.021)	(0.025)	(0.025)
Female pop	0.270***	0.322***	-0.430***
	(0.079)	(0.088)	(0.095)
Controls	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Observations	287	287	287
R-squared	0.830	0.840	0.828

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in (1) (resp. (2), (3)) is the share of electric, hybrid, and plug-in hybrid cars (resp. + cars working with natural gas in (2), and only diesel and gasoline in (3)) in municipalities in 2019. Green Pr. Wom. (resp. Men) gives the share of male (resp. female) users with a share of green tweets higher than $\alpha_1 = 0\%$. Income Women (resp. Income Men) is given by the yearly median income of men (resp. women) in the municipality measured in hundreds of thousands of Swedish Crowns in 2019. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old in the municipality. Female pop refers to the share of women in the municipality. Controls included: population density and share of company cars and taxis.

nicipality and the share of votes dedicated to the Green Party.[11] The estimated

[11] See Table 15 in Appendix C for all estimations results.

Table 6: Relation between men and women's green preferences and the share of votes for the Green Party in the 2018 Riksdag elections in Swedish municipalities.

	(1)	(2)
	GreenParty	*GreenParty*
Green pr. Wom.	2.196***	1.541**
	(0.799)	(0.733)
Green pr. Men	2.930***	2.638***
	(0.783)	(0.685)
Income Women	4.538***	3.384***
	(0.533)	(0.477)
Income Men	-1.530***	-1.525***
	(0.399)	(0.355)
% 20-35 yo	11.52***	10.41***
	(2.40)	(2.21)
Female pop		51.019***
		(5.877)
Controls	Yes	Yes
County FE	Yes	Yes
Observations	287	287
R-squared	0.671	0.738

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the share of votes for the Green party in the 2018 Riksdag elections. Green Pr. Wom. (resp. Men) gives the share of women (resp. men) with a share of environment-related tweets higher than $\alpha_1 = 0\%$. Income Women (resp. Men) is given by the median income of men (resp. women) in the municipality measured in hundred thousand Swedish Crowns. Female pop refers to the share of women in the municipality. Other controls included: population density and share of inhabitants aged 20 to 35 years old.

coefficient is higher for men than for women for both models, although the relative difference is smaller than before. Moreover, Table 15 in Appendix C shows that the sign of the difference is sensitive to the used threshold. This contradictory evidence does not support the existence of a gender gap in public sphere PEBs, which is in line with the literature.

The second column of Table 6 provides estimates of the relationship between green voting and municipal-level green preferences of men and women, investigating as well the effect of the share of the female population on the Green Party results in a municipality. The share of women in a municipality has a large effect on the share of green votes in this municipality: controlling for the female population reduces the magnitude of the relation between women's green preferences and the Green Party results. The share of women in a municipality being strongly correlated with municipal-level PEBs is coherent with stronger green preferences for women which would not be captured by our measure completely.

6 Conclusion

We propose in this paper a preference elicitation method at the intersection of revealed and stated preferences based on publications on the social media Twitter. We measure environmentally friendly preferences by investigating the share of tweets an individual dedicates to environmental issues out of all his/her tweets, based on the assumption that individuals reveal their preferences for a topic when they discuss it on Twitter.

The attractiveness of our measure is then demonstrated in two different empirical exercises. At the individual level, we document a potential cost of eco-consciousness, by investigating the relationship between individuals' green preferences and subjective well-being. We indeed show that not only have individuals with green preferences a lower SWB on average than individuals without, but also that an increase in the share of tweets dedicated to the environment is associated with a decrease in SWB for individuals with strong green preferences. Although we are not able to claim causality due to potential reverse causality issues, our evidence in the direction of a SWB cost of environmental concern clearly sheds light on the need to investigate its consequences on for instance between-group inequalities. Indeed, given that we find that women are more likely to be concerned about the state of the environment, a negative effect of environmental concern on productivity could, for instance, widen further gender economic inequalities. Obviously, our method could be used in further research to enable causal identification of the effect of eco-consciousness on SWB, by investigating, for example, the different effects of climatic events on individuals with and without preferences for the environment.

We then contribute to the environmental literature focusing on gender as a driver of green preferences and pro-environmental behaviors by corroborating the finding that women have stronger green preferences than men. Our results from the municipal-level analysis of the relationship between the share of women with green preferences (resp. men) and the share of green cars in Swedish municipalities suggests the existence of a gender gap in PEBs in the private sphere, while our results for voting behavior do not make us conclude the same for a public sphere PEB gender gap. In the current paper we only included basic robustness checks related to network effects. However, interesting future research could also use Twitter data more explicitly by uncovering the network amongst (influential) users and subsequently measuring its impact on private and public PEBs.

Next to the above suggestions directly related to the empirical questions of this paper, our approach for eliciting individual and unsolicited preferences based on social media data can also be used for a wide range of alternative possibilities for future research. The access to unlimited data anywhere and at any time allows for a detailed analysis of the impact of policy interventions, public events, natural disasters, etc. on individual preferences. This could uncover substantial heterogeneity across time, region, and individuals, and, in combination with network effects, could help improve our understanding of how all this impacts (economic) decisions taken by individuals and households.

References

Acemoglu, Daron, Hassan, Tarek A, & Tahoun, Ahmed. 2018. The power of the street: Evidence from Egypt's Arab Spring. *The Review of Financial Studies*, **31**(1), 1–42.

Baylis, Patrick. 2020. Temperature and temperament: Evidence from Twitter. *Journal of Public Economics*, **184**(Apr.), 104161.

Blank, Grant, & Lutz, Christoph. 2017. Representativeness of social media in great britain: investigating Facebook, Linkedin, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*, **61**(7), 741–756.

Blankenberg, Ann-Kathrin, & Alhusen, Harm. 2019. On the determinants of pro-environmental behavior: A literature review and guide for the empirical economist. *Center for European, Governance, and Economic Development Research (CEGE)*.

Campante, Filipe R, Depetris-Chauvin, Emilio, & Durante, Ruben. 2020. *The virus of fear: The political impact of ebola in the us*. Tech. rept. National Bureau of Economic Research.

Dunlap, & Riley. 1983. Male-female differences in concern for environmental quality. *International journal of Women's Studies*, **6**(4).

Dzialo, Liz. 2017. The feminization of environmental responsibility: A quantitative, cross-national analysis. *Environmental Sociology*, **3**(4), 427–437.

Elgaaied-Gambier, Leila, & Mandler, Timo. 2021. Me trying to talk about sustainability: Exploring the psychological and social implications of environmental threats through user-generated content. *Ecological Economics*, **187**, 107089.

Holmberg, Kim, & Hellsten, Iina. 2015. Gender differences in the climate change communication on Twitter. *Internet Research*, **25**(5), 811–828.

Hu, Yifan, Hu, Changwei, Tran, Thanh, Kasturi, Tejaswi, Joseph, Elizabeth, & Gillingham, Matt. 2021. What's in a name?–gender classification of names with character based machine learning models. *Data Mining and Knowledge Discovery*, 1–27.

Hunter, Lori M, Hatch, Alison, & Johnson, Aaron. 2004. Cross-national gender variation in environmental behaviors. *Social science quarterly*, **85**(3), 677–694.

Hutto, Clayton, & Gilbert, Eric. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Pages 216–225 of: Proceedings of the international AAAI conference on web and social media*, vol. 8.

Ivanova, Diana, Barrett, John, Wiedenhofer, Dominik, Macura, Biljana, Callaghan, Max, & Creutzig, Felix. 2020. Quantifying the potential for climate change mitigation of consumption options. *Environmental Research Letters*, **15**(9), 093001.

Kemp, S. 2019. *Digital 2019: Global digital report.*

Kennedy, Emily Huddart, & Kmec, Julie. 2018. Reinterpreting the gender gap in household pro-environmental behaviour. *Environmental Sociology*, **4**(3), 299–310.

Konisky, David M, Milyo, Jeffrey, & Richardson, Lilliard E. 2008. Environmental policy attitudes: Issues, geographical scale, and political trust. *Social science quarterly*, **89**(5), 1066–1085.

Kormos, Christine, & Gifford, Robert. 2014. The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology*, **40**, 359–371.

Kurisu, Kiyo. 2015. *What Are Pro-Environmental Behaviors (PEBs)?* Tokyo: Springer Japan. Pages 1–26.

Lan, Li, Lian, Zhiwei, & Pan, Li. 2010. The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings. *Applied Ergonomics*, **42**(1), 29–36.

Loureiro, Maria L., Alló, Maria, & Coello, Pablo. 2022. Hot in Twitter: Assessing the emotional impacts of wildfires with sentiment analysis. *Ecological Economics*, **200**(Oct.), 107502.

McCright, Aaron M, & Xiao, Chenyang. 2014. Gender and environmental concern: Insights from recent work and for future research. *Society & Natural Resources*, **27**(10), 1109–1113.

Mellon, Jonathan, & Prosser, Christopher. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, **4**(3), 205316801772000.

Mohai, Paul. 1992. Men, women, and the environment: An examination of the gender gap in environmental concern and activism. *Society & Natural Resources*, **5**(1), 1–19.

Müller, Karsten, & Schwarz, Carlo. 2021. Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, **19**(4), 2131–2167.

Schlosser, Stephan, Toninelli, Daniele, & Cameletti, Michela. 2021. Comparing Methods to Collect and Geolocate Tweets in Great Britain. *Journal of Open Innovation: Technology, Market, and Complexity*, **7**(1).

Schreiber, M. 2021. Addressing climate change concerns in practice. *Monitor on Psychology*, **52**(2).

Stehfest, Elke, Bouwman, Lex, Van Vuuren, Detlef P, Den Elzen, Michel GJ, Eickhout, Bas, & Kabat, Pavel. 2009. Climate benefits of changing diet. *Climatic change*, **95**(1), 83–102.

Stern, Paul C., Dietz, Thomas, & Kalof, Linda. 1993. Value Orientations, Gender, and Environmental Concern. *Environment and Behavior*, **25**(5), 322–348.

Usher, Nikki, Holcomb, Jesse, & Littman, Justin. 2018. Twitter makes it worse: Political journalists, gendered echo chambers, and the amplification of gender bias. *The international journal of press/politics*, **23**(3), 324–344.

Whitmarsh, Lorraine, Player, Lois, Jiongco, Angelica, James, Melissa, Williams, Marc, Marks, Elizabeth, & Kennedy-Williams, Patrick. 2022. Climate anxiety: What predicts it and how is it related to climate action? *Journal of Environmental Psychology*, **83**, 101866.

Williamson, Katie, Satre-Meloy, Aven, Velasco, Katie, & Green, Kevin. 2018. Climate change needs behavior change: Making the case for behavioral solutions to reduce global warming. *Rare: Arlington, VA, USA*.

Xiao, Chenyang, & McCright, Aaron M. 2015. Gender differences in environmental concern: Revisiting the institutional trust hypothesis in the USA. *Environment and Behavior*, **47**(1), 17–37.

Yates, Adam, Luo, Ye, Mobley, Catherine, & Shealy, Erin. 2015. Changes in public and private environmentally responsible behaviors by gender: Findings from the 1994 and 2010 general social survey. *Sociological Inquiry*, **85**(4), 503–531.

For online publication

Appendix A: Technical details

A.1 50 most common words in Swedish

Table 7: List of the 50 most common words in Swedish.

Swedish	English	Swedish	English
jag	I	till	to
det	it	kan	may
är	is	de	they
du	you	ni	you
att	to	ska	to
en	an	ett	an
och	and	men	but
har	has	av	by
vi	we	vill	to
på	on	nu	now
för	for	ja	yes
kommer	will	vet	know
han	he	nej	no
vad	what	bara	only
med	with	hon	she
mig	me	bra	good
som	as	när	when
om	on	ha	have
här	here	er	s
dig	you	ta	check
var	was	ut	out
den	it	då	then
så	as	skulle	would
din	your	kom	came
i	in		

Source: Hermit Dave, 2021. URL: http://www.101languages.net/swedish/most-common-swedish-words/, last visited on the 13/08/2021.

A.2 Gender attribution algorithm

Different methods can be used to determine demographics from Twitter data. Such information can be retrieved from the content written by users on the social media, as specific formulations or vocabulary can be specific of a demographic trait, as well as from the user-name, name, author's description (Twitter users can write a small biography to introduce themselves on the social media), and from their profile picture (Usher *et al.*, 2018; Hu *et al.*, 2021). Usernames and names are often used for gender detection, and it is the method used here.

In order to determine users' gender using their name and username, we create a Gender dictionary, containing names commonly used in Sweden and in the United States as well as the gender it most commonly refers to. We use data from the Swedish national statistics agency (SCB)'s list of usual names given to newborn boys and girls between 1998 and 2020, as well as from the US Social Security Administration (SSA)'s list of registered baby names in the United States since 1880 to create the dictionary.[12] We then create an algorithm going through the variables *authorname* or *authorusername*, which attributes the gender *female* if the author's name is listed as a woman's name in our dictionary, and *male* if the name is listed as a man's name. The functioning of the gender attribution algorithm is schematized in Figure 8. The algorithm is written in 3 main parts.

- The first part focuses on names that are written in the form "first name last name", as for instance "Elisa Stevens". In case of such a structure of the author's name on Twitter, if the algorithm just checks for the presence of a name from our dictionary, the attributed gender could be both *male* or *female*, as "Elisa", the author's first name, is classified as a female name, and "Stevens", the author's last name, is classified as a male name. The algorithm thus attributes a gender to author names that contain a name in the Gender dictionary followed by a space, to consider only the first name of the author for the gender attribution: "Elisa Stevens" contains "Elisa ", and is thus attributed the gender *female*.

- The second part focuses on names that are written in the form "first name, to be ignored", as for instance "William, biggest fan of Greta Thumberg". In that case, the second part of the author's name "biggest fan of Greta Thumberg" includes the first name "Greta", and could thus be assigned the gender *female*, while the first name of the user is "William". The algorithm thus attributes a gender to author names that contain a name of our Gender

[12] We noticed that many names of our Twitter users were American.

dictionary, followed by a comma, to avoid taking into account what is to be ignored in the author's name on Twitter: "William, biggest fan of Greta Thumberg" contains "William,"and is thus attributed the gender *male*.

- The third part takes care of other author names, whose structure is varying a lot from Twitter user to Twitter user. The algorithm considers the whole author name, and attributes to each author the gender corresponding to the name with the highest length of our Gender dictionary that fits in the author's name. Consider for instance the author name "Josephine SwedishGirl". The name "Joseph" is in the author's name and is classified as belonging to a male. However, "Josephine" is also in the author's name, and is longer than Joseph. The attributed gender is thus "female". This last step is repeated using the variable "author username" for the users whose gender could not be retrieved using the author name.

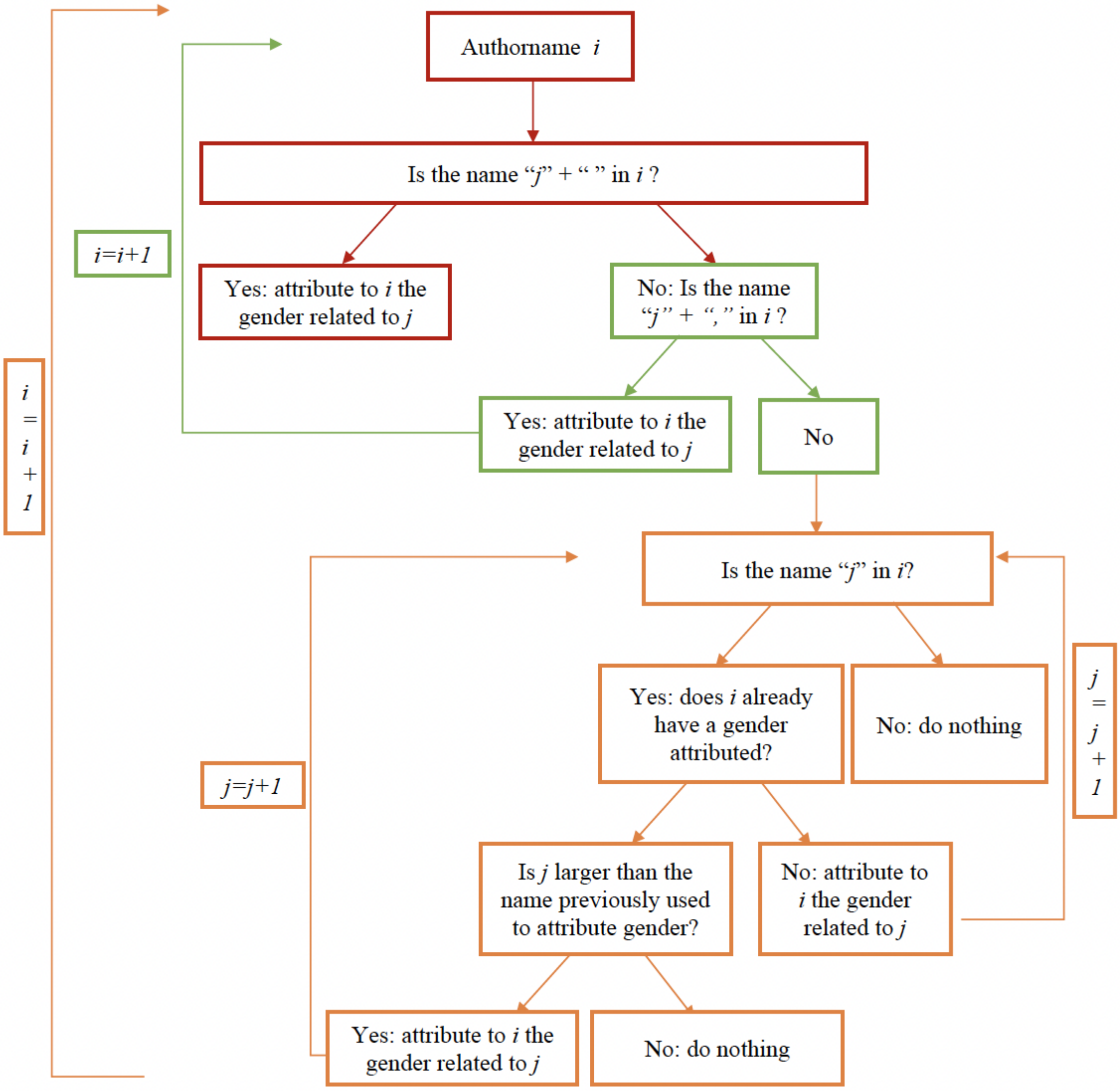


Figure 8: Functioning of the gender attribution algorithm, for an author name i, and a name in the gender dictionary j.

A.3 Municipality attribution algorithm

We base the localization of Twitter users on two variables: *authorlocation*, and *geoname*, where the former is the self-entered location of the user, and the latter is the location given by the geo-localization of tweets. We write an algorithm that goes through each value of the variables *authorlocation*, or *geoname*, and attributes to it a municipality. For the algorithm to be able to map a location to a municipality, we construct a Location dictionary based on SCB's regional data of municipalities and localities, and on geographical data provided by the Humanitarian OpenstreetMap Team (HOT).[13] The former provides us with a list of cities of 200 inhabitants or more and the municipality they belong to, and the latter with a list of places such as towns, villages, hamlets, cities and isolated dwellings, which we map to municipalities using Geographical Information Systems (GIS) tools. We thus create a dictionary that attributes to 68 600 places a municipality in Sweden, and use it to map each Twitter users to a municipality, using their self-entered location. Figure 9 schematizes the functioning of the algorithm. The algorithm is written in 3 main parts.

- The first part focuses on author locations that are written in the form "municipality name", as for instance "Goteborg". In this case the attributed municipality is set equal to the author location.
- The second part focuses on author locations that are written in the form "locality, to be ignored", as for instance "Porjus, but citizen of the world". Localities considered here are the cities of 200 inhabitants or more provided by the SCB. The algorithm attributes to the author location the municipality of the longest locality names it includes (we pay attention to the length of the string as we do in the gender algorithm).
- The third part takes care of other author locations, whose structure varies a lot from author to author. The algorithm considers the whole location name, and attributes to each author the municipality corresponding to the

[13]Respectively, https://www.scb.se/en/finding-statistics/regional-statistics/HOTOSM and https://www.hotosm.org/tools-and-data. Last visited on the 13/08/2021.

longest location of the places dictionary that fits in the author's location (as in the third step of the gender algorithm). In this third step, the attribution of a municipality can be more prone to error since they include neither the municipality name nor the name of a classified locality. It can include instead the name of a district, of a parish, of a street, of a county, etc. We manually double-checked all the municipalities attributed via this third step and corrected mismatches.

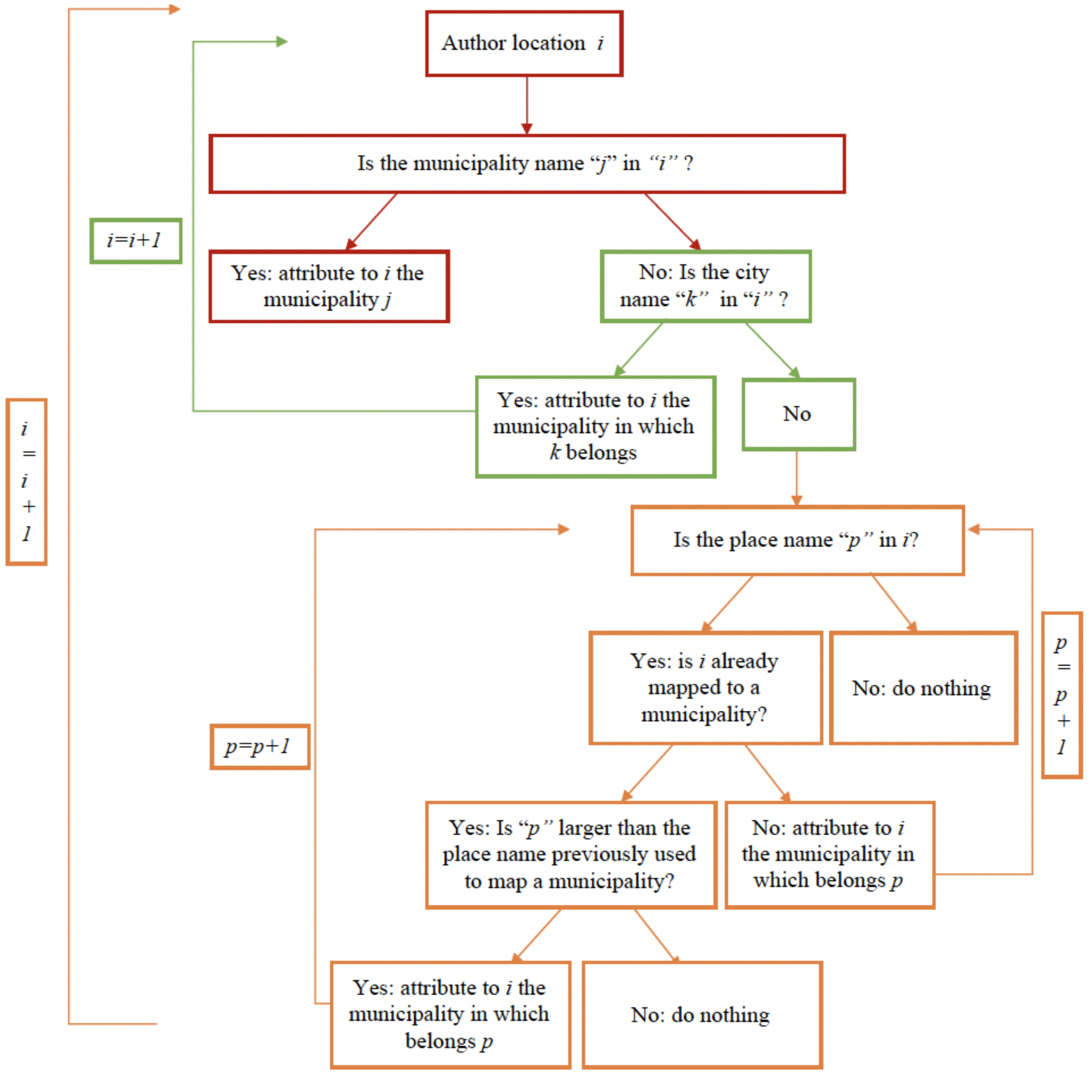


Figure 9: Functioning of the municipality attribution algorithm for an author location i, and a municipality j in the municipality dictionary.

Appendix B: The role of climate deniers

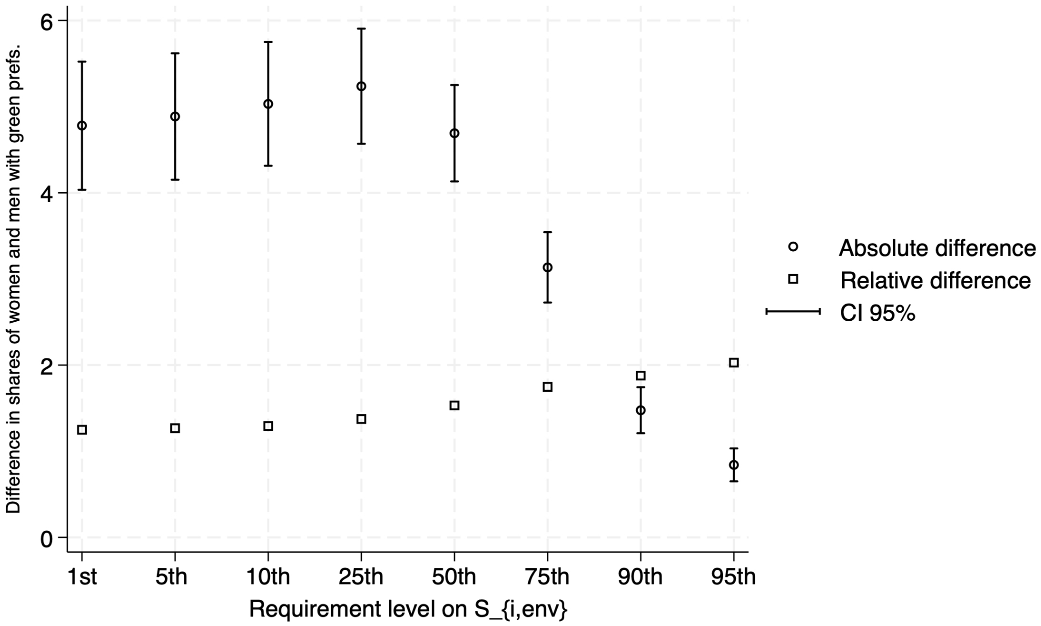
To make sure climate deniers do not bias our results, we use supervised machine learning to identify climate skeptic individuals in our sample. For this part of the analysis and to limit the computational cost of translating all the tweets from English to Swedish, we keep only the original tweets of users (we drop retweets). This leads us to classify 8996 individuals (out of the 13 825 individuals discussing climate on Twitter) and their 39571 climate tweets related to climate change as climate-denying or not.

We classify manually a training set of 2000 tweets, in which the share of tweets with a climate skeptic message is only 3.15%. This low share of climate-denying tweets is reassuring for the analysis, but challenging for the machine learning method to work, as this 3.15% of tweets hold a limited amount of information to identify climate deniers. We thus train the Support Vector Machine model with a linear kernel on a re-sampled classified set, where denying tweets have been over-sampled by 10%, neutral tweets have been sampled down to be 60% more than denying tweets, and asymmetric class weights have been attributed to each class in the cost function of the algorithm, such that it is twice as costly to estimate a denying tweet as neutral than to predict a neutral tweet as denying. We give as input to the model the TF-IDF (Term Frequency - Inverse Document) frequency for each tweet.[14] The model predicts true negative in 78% of the case. It classifies 19% of non-denying tweets as being denying, due to our restrictive approach (remember that we aim at removing potential climate deniers, so we prefer identifying rather too many deniers than not enough). We estimate true positives in 80.9% of the cases.

We reproduce the main results of Sections 4 and 5 on gender, with a reduced sample of individuals in which we drop all individuals who could potentially discuss environmental issues in negative terms. Given our restrictive approach, we estimate that, from our individuals with green preferences, 21.82% are potentially climate skeptics, of which 69.56% are men and 30.44% women. This finding goes in

[14]The TF-IDF corresponds to a measure of how often each uni-gram or bi-gram occurred within the tweet adjusted for frequency overall tweets.

Figure 10: Extensive margin, dropping potential climate skeptics individuals.



Notes: Difference between shares of women and men with green preferences for different levels of requirements on the share of tweets dedicated to the environment. Absolute differences are given in percentage points and are indicated with dots with their confidence intervals at the 95% confidence level. The relative differences are computed as the ratio of the two shares and represented by squares. Requirement levels on $S_{i,env}$ are given by the percentiles of $S_{i,env}$'s distribution. Requirement levels given by the percentiles of $S_{i,env}$'s distribution.

the same direction as Holmberg & Hellsten (2015), who found that tweets written by men tend to refer more often to climate skeptic private individuals than tweets written by women. As is clear from the figures and tables below, climate deniers do not appear to be a concern for our empirical conclusions.

Table 8: Relationship between individuals' SWB and green preferences, dropping potential climate skeptics individuals.

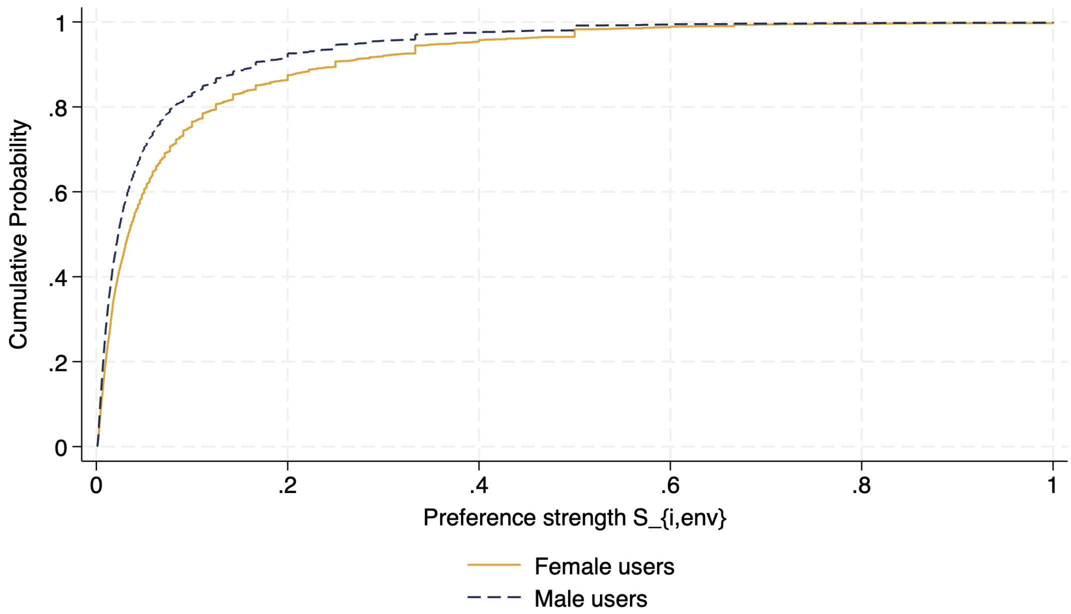
	(1) Average daily subjective well-being
Env.-concerned	-0.012***
	(0.004)
Env.-concerned=1 × share_related	1.009***
	(0.143)
Env.-concerned=1 × share_relatedsquared	-0.757
	(0.464)
User activity level	-0.000
	(0.000)
Active days	0.000
	(0.000)
group(gender)=2	-0.022***
	(0.003)
Month FE	Yes
Municipality FE	Yes
Observations	7,928
R-squared	0.056

Standard errors in parentheses

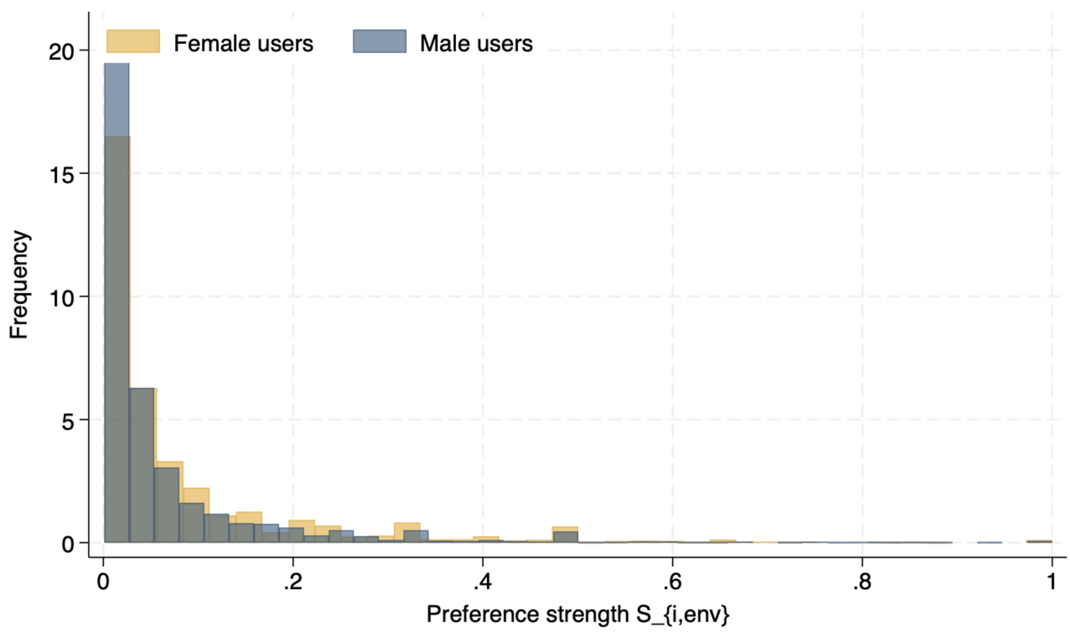
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the average SWB of individual i on the period. Env.-concerned is a binary variable equal to 1 if the individual reveals to care about the environment on Twitter. Share related refers to the share of tweets each individual dedicates to the environment. User activity level is the number of tweets written by the user. Active days is the number of days during which the user wrote at least 3 tweets. Male users is a binary variable equal to 1 if the individual is a man.

Figure 11: Intensive margin, dropping potential climate skeptics individuals.



(a) Distribution



(b) Cumulative distribution

Notes: Panel (a) depicts the cumulative distribution of males' and females' share of tweets related to the environment (i.e. $S_{i,env} > 0$). Panel b depicts the distribution of these $S_{i,env}$ for men and women in the sample.

Table 9: Relation between men and women's green preferences and the share of green cars in Swedish municipalities. Dropping potential climate skeptics individuals.

	(1) Baseline α_1	(2) Baseline α_2	(3) Baseline α_3	(4) + Gas α_1	(5) + Gas α_2	(6) + Gas α_3	(7) Polluting α_1	(8) Polluting α_2	(9) Polluting α_3
Green pr. Wom.	0.025***	0.028***	0.033***	0.024***	0.025**	0.031**	-0.032***	-0.038***	-0.050***
	(0.009)	(0.010)	(0.012)	(0.009)	(0.011)	(0.012)	(0.011)	(0.011)	(0.013)
Green pr. Men	0.009	0.012	0.007	0.013	0.016	0.012	-0.019*	-0.018	-0.013
	(0.007)	(0.009)	(0.009)	(0.008)	(0.010)	(0.010)	(0.010)	(0.012)	(0.013)
Pop density	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	-0.006***	-0.006***	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Company cars	0.066***	0.065***	0.065***	0.070***	0.070***	0.070***	-0.027	-0.026	-0.026
	(0.021)	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
% 20-35 yo	0.100***	0.104***	0.105***	0.160***	0.164***	0.165***	-0.204***	-0.210***	-0.212***
	(0.021)	(0.021)	(0.022)	(0.025)	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)
Income Women	0.007	0.007	0.007	0.010	0.011	0.011	-0.008	-0.009	-0.009
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Income Men	0.014**	0.014**	0.014**	0.013**	0.013**	0.013**	-0.011*	-0.011*	-0.011*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female pop	0.280***	0.286***	0.288***	0.332***	0.338***	0.343***	-0.441***	-0.451***	-0.454***
	(0.078)	(0.079)	(0.078)	(0.087)	(0.088)	(0.088)	(0.094)	(0.094)	(0.093)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	287	287	287	287	287	287	287	287	287
R-squared	0.830	0.829	0.829	0.839	0.838	0.838	0.828	0.826	0.827

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable from columns 1 to 3 (resp. 4 to 6, and 7 to 9) is the share of electric, hybrid, and plug-in hybrid cars (resp. + cars working with natural gas in 4 to 6, and only diesel and gasoline for 7 to 9) in municipalities in 2019. Green Pr. Wom. (resp. Men) gives the share of female (resp. male) users with a share of environment-related tweets higher than $\alpha_1 = 0\%$; $\alpha_2 = 2.8\%$ (median); and $\alpha_3 = 5\%$. Income Men (resp. Income Women) is given by the yearly median income of men (resp. women) in the municipality measured in hundreds of thousands of Swedish Crowns in 2019. Population density is measured as the number of inhabitants per square meter. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old. Company cars refer to the share of non-private cars in a municipality. Female pop refers to the share of women in the municipality.

Table 10: Relation between men and women's green preferences and the share of votes for the Green Party in the 2018 Riksdag elections in Swedish municipalities. Dropping potential climate skeptics individuals.

	(1) α_1	(2) α_2	(3) α_3	(4) α_1	(5) α_2	(6) α_3
Green pr. Wom.	1.97**	2.19**	2.83***	1.54*	1.83**	2.35**
	(0.89)	(0.99)	(1.09)	(0.81)	(0.91)	(1.01)
Green pr. Men	3.16***	2.94***	1.66	2.85***	2.50**	1.69
	(0.82)	(1.09)	(1.12)	(0.68)	(1.02)	(1.13)
% 20-35 yo	11.52***	12.24***	12.36***	10.41***	11.02***	11.10***
	(2.40)	(2.46)	(2.48)	(2.21)	(2.28)	(2.30)
Income Women	4.52***	4.58***	4.60***	3.34***	3.38***	3.37***
	(0.54)	(0.55)	(0.56)	(0.47)	(0.50)	(0.50)
Income Men	-1.50***	-1.48***	-1.47***	-1.50***	-1.48***	-1.47***
	(0.40)	(0.42)	(0.42)	(0.35)	(0.37)	(0.37)
Pop density	0.02	0.02	0.02	0.06	0.06	0.06
	(0.14)	(0.14)	(0.14)	(0.11)	(0.11)	(0.12)
Female pop				51.77***	52.69***	53.55***
				(5.87)	(6.31)	(6.40)
County FE	Yes	Yes	Yes	Yes	Yes	
Observations	287	287	287	287	287	287
R-squared	0.666	0.652	0.645	0.736	0.724	0.720

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the share of votes dedicated to the Green party in the 2018 Riksdag elections. Green Pr. Wom. (resp. Men) gives the share of female (resp. male) users with a share of environment-related tweets higher than $\alpha_1 = 0\%$; $\alpha_2 = 2.8\%$ (median); and $\alpha_3 = 5\%$. Income Women (resp. Men) is given by the median income of men (resp. women) in the municipality measured in hundred thousand Swedish Crowns. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old. Population density is measured as the number of inhabitants per square meter. Female pop refers to the share of women in the municipality.

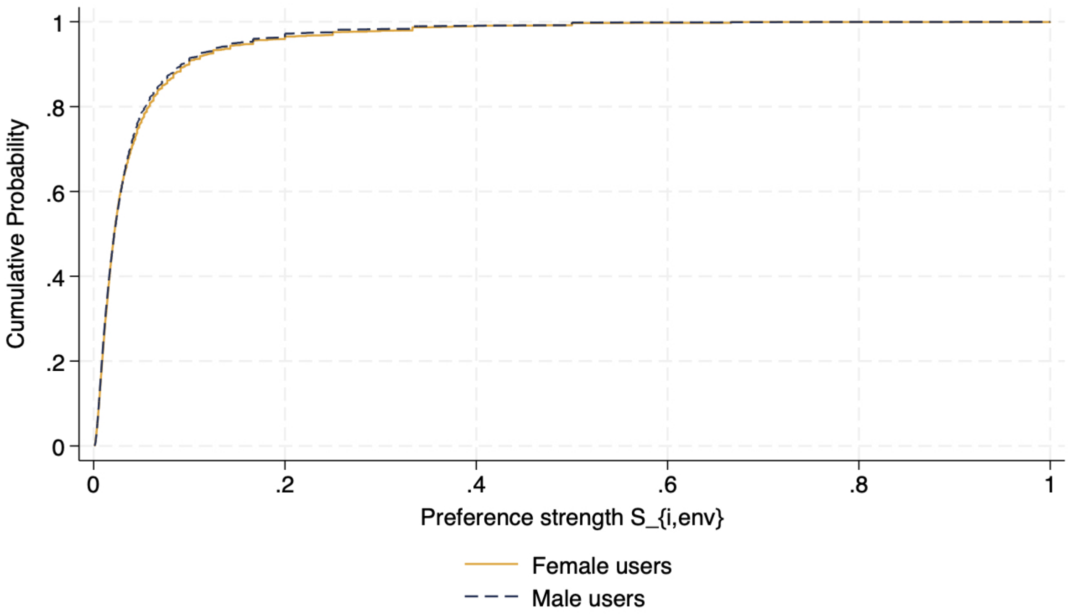
Appendix C: additional empirical results

C.1 Systematic communication gender differences

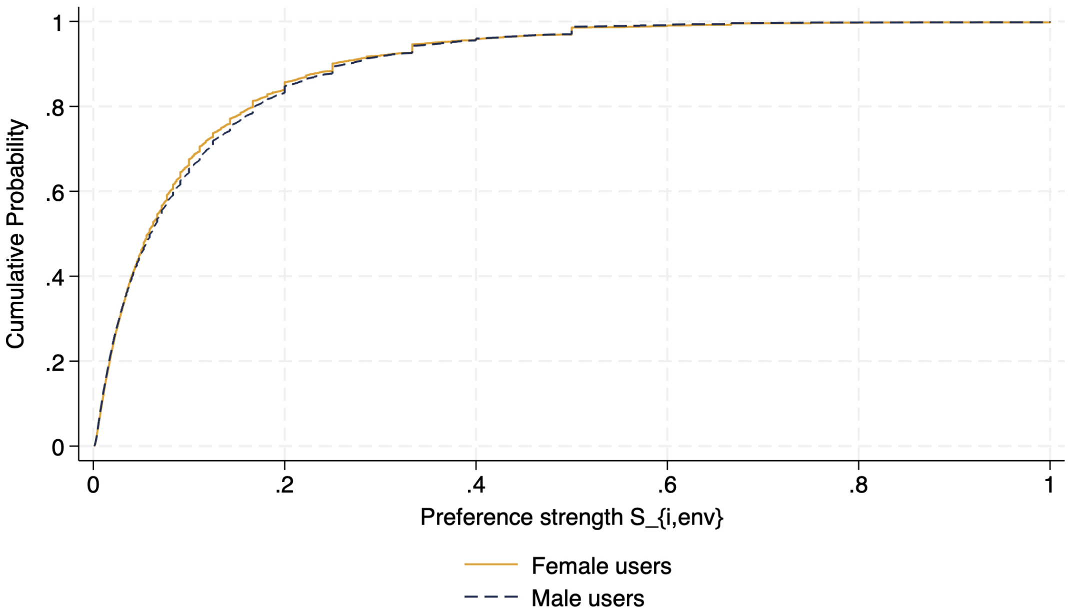
We reproduce the findings provided in Section 4.2 for different topics discussed on Twitter, to show that our finding that women are more climate concerned than men, both at the extensive and intensive margin, is not driven by communication differences on Twitter, and is specific to environmental topics. Particularly, we investigate discussions across gender about immigration and politics. The keywords used to identify related tweets are given in Table 11. Table 12 provides the share of male and female users that discuss environmental topics, immigration, and politics respectively in their tweets. There are clearly some differences but these are much smaller for immigration and politics.

Different patterns are observed for the three topics: while relatively more women discuss environmental topics than men and to a greater extent, relatively more men discuss immigration, although less extensively than women. For politics, relatively more women discuss related topics, but men discuss them more intensively. Looking at Figure 12, we also see that there is no clear dominance in the cumulative distributions of the share of tweets dedicated to immigration (Figure 12a) and that the cumulative distribution of the share of men's tweets dedicated to politics (Figure 12b) is, although very slightly, to the right of women's. All in all we can safely conclude that the differences in the main text are more outspoken.

Figure 12: Distribution of individuals' preference strength by gender for immigration and politics.



(a) Cumulative distribution of males and females' share of tweets dedicated to immigration



(b) Cumulative distribution of males and females' share of tweets dedicated to politics

Notes: Panel (a) depicts the cumulative distribution of shares of tweets related to immigration $S_{i,immig} > 0$ for men and women in the sample. Panel (b) depicts the cumulative distribution of shares of tweets related to politics $S_{i,pol}$ for men and women in the sample.

Table 11: Unigrams and bigrams used to identify tweets related to immigration and politics

	Unigrams and bigram
Immigration	migration, deportation, migrant, refugee, foreigner, citizenship, border, citizen, undocumented, migrate, temporary resident, unhcr, asylum, exodus, displacement, exile, expulsion, naturalization, xenophobia, racism, brain drain, brain gain, illegal stay, illegal entry, residence permit
Politics	social democrats, moderates, left party, Sweden democrats, green party, center party, christian democrats, liberals, feminist initiative, feminist party, European parliament, government, constitution, democracy, republic, politics, corruption, prime minister, extreme right, extreme left, left wing, right wing, election, freedom

Notes: Unigrams and bigrams based on the general lexicon of immigration and political discourses.

Table 12: Discussion of environmental issues, immigration and politics on Twitter across gender.

	Male users			Female users		
Topic	Users	Av. tweets	Med. tweets	Users	Av. tweets	Med. tweets
Env.	23.93	6.79	2.44	28.04	9.52	3.64
Immig.	28.17	4.28	2.15	26.62	4.53	2.15
Pol.	38.96	10.67	5.95	39.36	10.42	5.71

Notes: Env refers to Environmental issues, for which the keywords considered are given in Table 2. Immig refers to issues related to immigration and Pol refers to political issues. The keywords considered for both are given in Table 11. Users gives the share of female (resp. male) users with a positive share of tweets on these topics. Av. (resp Med.) tweets provides the average (resp. median) number of tweets per individual on these topics, among the concerned individuals.

C.2. Full tables and alternative levels of α.

Table 13 provides the estimation results given in Table 4, with details on the control variables. Table 14 (resp. Table 15) replicates Table 5 (resp. Table 6), with results provided for different levels of requirements α on the share of tweets an individual has to dedicate to the environment to be considered as revealing green preferences. The tables also provide details on the control variables.

Table 13: Relationship between individuals' SWB and green preferences.

	(1) Average daily SWB
Env.-concerned	-0.016***
	(0.003)
Env.-concerned=1 × share_related	0.457***
	(0.071)
Env.-concerned=1 × share_related squared	-0.561***
	(0.191)
User activity level	-0.000
	(0.000)
Active days	-0.000*
	(0.000)
Male user	-0.026***
	(0.003)
Month FE	Yes
Municipality FE	Yes
Observations	9,786
R-squared	0.054

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the average SWB of individual i on the period. Env.-concerned is a binary variable equal to 1 if the individual reveals to care about the environment on Twitter. Share related refers to the share of tweets each individual dedicates to the environment. User activity level is the number of tweets written by the user. Active days is the number of days during which the user wrote at least 3 tweets. Male users is a binary variable equal to 1 if the individual is a man.

Table 14: Relation between men and women's green preferences and the share of green cars in Swedish municipalities.

	(1) Baseline α_1	(2) Baseline α_2	(3) Baseline α_3	(4) + Gas α_1	(5) + Gas α_2	(6) + Gas α_3	(7) Polluting α_1	(8) Polluting α_2	(9) Polluting α_3
Green pr. Wom.	0.023***	0.023**	0.030***	0.023**	0.022**	0.029**	-0.028***	-0.032***	-0.044***
	(0.009)	(0.010)	(0.011)	(0.009)	(0.011)	(0.012)	(0.010)	(0.011)	(0.011)
Green pr. Men	0.011*	0.011	0.008	0.013*	0.013	0.012	-0.019**	-0.015	-0.014
	(0.006)	(0.008)	(0.008)	(0.007)	(0.009)	(0.008)	(0.008)	(0.010)	(0.011)
Pop density	0.007***	0.007***	0.007***	0.007***	0.007***	0.007***	-0.006***	-0.006***	-0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Company cars	0.065***	0.064***	0.064***	0.070***	0.069***	0.069***	-0.026	-0.025	-0.025
	(0.021)	(0.021)	(0.021)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
% 20-35 yo	0.100***	0.104***	0.106***	0.160***	0.165***	0.166***	-0.205***	-0.211***	-0.213***
	(0.021)	(0.021)	(0.022)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Income Women	0.007	0.007	0.007	0.011	0.011	0.011	-0.009	-0.009	-0.009
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Income Men	0.014**	0.014**	0.014**	0.013**	0.013**	0.013**	-0.011*	-0.011*	-0.011*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female pop	0.270***	0.276***	0.281***	0.322***	0.328***	0.334***	-0.430***	-0.438***	-0.442***
	(0.079)	(0.080)	(0.078)	(0.088)	(0.089)	(0.088)	(0.095)	(0.094)	(0.093)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	287	287	287	287	287	287	287	287	287
R-squared	0.830	0.828	0.829	0.840	0.838	0.838	0.828	0.825	0.827

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable from columns 1 to 3 (resp. 4 to 6, and 7 to 9) is the share of electric, hybrid, and plug-in hybrid cars (resp. + cars working with natural gas in 4 to 6, and only diesel and gasoline for 7 to 9) in municipalities in 2019. Green Pr. Wom. (resp. Men) gives the share of female (resp. male) users with a share of environment-related tweets higher than $\alpha_1 = 0\%$; $\alpha_2 = 2.8\%$ (median); and $\alpha_3 = 5\%$. Income Men (resp. Income Women) is given by the yearly median income of men (resp. women) in the municipality measured in hundreds of thousands of Swedish Crowns in 2019. Population density is measured as the number of inhabitants per square meter. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old. Company cars refer to the share of non-private cars in a municipality. Female pop refers to the share of women in the municipality.

Table 15: Relationship between men and women's green preferences and the share of votes for the Green Party in the 2018 Riksdag elections in a municipality.

	(1)	(2)	(3)	(4)	(5)	(6)
	α_1	α_2	α_3	α_1	α_2	α_3
Green pr. Wom.	2.20***	2.40***	3.12***	1.54**	1.78**	2.39***
	(0.80)	(0.88)	(0.98)	(0.73)	(0.81)	(0.89)
Green pr. Men	2.93***	3.07**	2.38*	2.64***	2.43**	2.13*
	(0.78)	(1.19)	(1.22)	(0.69)	(1.04)	(1.10)
% 20-35 yo	11.48***	12.27***	12.48***	10.40***	11.07***	11.21***
	(2.39)	(2.44)	(2.45)	(2.21)	(2.28)	(2.28)
Income Women	4.54***	4.58***	4.58***	3.38***	3.42***	3.38***
	(0.53)	(0.54)	(0.55)	(0.48)	(0.49)	(0.49)
Income Men	-1.53***	-1.50***	-1.47***	-1.52***	-1.49***	-1.47***
	(0.40)	(0.41)	(0.41)	(0.35)	(0.36)	(0.36)
Pop density	0.02	0.01	0.02	0.05	0.05	0.06
	(0.14)	(0.14)	(0.14)	(0.11)	(0.11)	(0.11)
Female pop				51.02***	51.28***	52.55***
				(5.88)	(6.22)	(6.32)
County FE	Yes	Yes	Yes	Yes	Yes	
Observations	287	287	287	287	287	287
R-squared	0.671	0.659	0.652	0.738	0.726	0.724

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the share of votes dedicated to the Green party in the 2018 Riksdag elections. Green Pr. Wom. (resp. Men) gives the share of female (resp. male) users with a share of environment-related tweets higher than $\alpha_1 = 0\%$; $\alpha_2 = 2.8\%$ (median); and $\alpha_3 = 5\%$. Income Women (resp. Men) is given by the median income of men (resp. women) in the municipality measured in hundred thousand Swedish Crowns. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old. Population density is measured as the number of inhabitants per square meter. Female pop refers to the share of women in the municipality.

C.3 Accounting for local networks existence

Tables 16 and 17 replicate respectively Tables 14 and 15, dropping retweets from the set of individuals' tweets considered.

Table 16: Relation between men and women's green preferences and the share of green cars in Swedish municipalities. No retweets.

	(1) Baseline α_1	(2) Baseline α_2	(3) Baseline α_3	(4) + Gas α_1	(5) + Gas α_2	(6) + Gas α_3	(7) Polluting α_1	(8) Polluting α_2	(9) Polluting α_3
Green pr. Wom.	0.015	0.020	0.014	0.016	0.020	0.016	-0.018	-0.025	-0.021
	(0.011)	(0.015)	(0.015)	(0.012)	(0.016)	(0.016)	(0.014)	(0.017)	(0.017)
Green pr. Men	0.008	0.016	0.007	0.009	0.017	0.010	-0.013	-0.009	0.000
	(0.008)	(0.011)	(0.010)	(0.009)	(0.011)	(0.011)	(0.010)	(0.012)	(0.012)
Pop density	0.007***	0.007***	0.008***	0.007***	0.007***	0.007***	-0.006***	-0.006***	-0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Company cars	0.065***	0.064***	0.065***	0.070***	0.069***	0.070***	-0.027	-0.026	-0.027
	(0.021)	(0.021)	(0.022)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
% 20-35 yo	0.103***	0.105***	0.105***	0.163***	0.166***	0.165***	-0.208***	-0.211***	-0.211***
	(0.021)	(0.022)	(0.022)	(0.025)	(0.025)	(0.026)	(0.025)	(0.025)	(0.025)
Income Women	0.007	0.006	0.007	0.010	0.010	0.010	-0.008	-0.008	-0.008
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)
Income Men	0.014**	0.014**	0.014**	0.012**	0.013**	0.013**	-0.011*	-0.011*	-0.011*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female pop	0.291***	0.283***	0.293***	0.344***	0.335***	0.346***	-0.460***	-0.456***	-0.466***
	(0.079)	(0.079)	(0.079)	(0.088)	(0.088)	(0.088)	(0.095)	(0.095)	(0.095)
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	287	287	287	287	287	287	287	287	287
R-squared	0.827	0.828	0.826	0.837	0.837	0.836	0.823	0.822	0.822

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Green preference measure based on the sample of individuals' tweets excluding retweets. The dependent variable from columns 1 to 3 (resp. 4 to 6, and 7 to 9) is the share of electric, hybrid, and plug-in hybrid cars (resp. + cars working with natural gas in 4 to 6, and only diesel and gasoline for 7 to 9) in municipalities in 2019. Green Pr. Wom. (resp. Men) gives the share of female (resp. male) users with a share of environment-related tweets higher than $\alpha_1 = 0\%$; $\alpha_2 = 2.8\%$ (median); and $\alpha_3 = 5\%$. Income Men (resp. Income Women) is given by the yearly median income of men (resp. women) in the municipality measured in hundreds of thousands of Swedish Crowns in 2019. Population density is measured as the number of inhabitants per square meter. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old. Company cars refer to the share of non-private cars in a municipality. Female pop refers to the share of women in the municipality.

Table 17: Relation between men and women's green preferences and the share of votes for the Green Party in the 2018 Riksdag elections in Swedish municipalities. No retweets.

	(1)	(2)	(3)	(4)	(5)	(6)
	α_1	α_2	α_3	α_1	α_2	α_3
Green pr. Wom.	1.82*	2.16*	1.99*	1.45	1.62	1.74*
	(0.99)	(1.16)	(1.12)	(0.94)	(1.09)	(1.05)
Green pr. Men	2.56**	2.86*	2.12	2.43***	2.25*	1.72
	(1.01)	(1.59)	(1.84)	(0.93)	(1.34)	(1.58)
% 20-35 yo	11.96***	12.47***	12.54***	10.74***	11.20***	11.26***
	(2.45)	(2.46)	(2.49)	(2.26)	(2.29)	(2.31)
Income Women	4.54***	4.51***	4.59***	3.32***	3.34***	3.37***
	(0.53)	(0.54)	(0.54)	(0.48)	(0.50)	(0.50)
Income Men	-1.53***	-1.47***	-1.50***	-1.52***	-1.47***	-1.50***
	(0.41)	(0.41)	(0.41)	(0.36)	(0.36)	(0.36)
Pop density	0.02	0.02	0.02	0.06	0.06	0.06
	(0.14)	(0.14)	(0.14)	(0.12)	(0.12)	(0.12)
Female pop				53.22***	52.45***	53.42***
				(6.15)	(6.25)	(6.31)
County FE	Yes	Yes	Yes	Yes	Yes	
Observations	287	287	287	287	287	287
R-squared	0.655	0.650	0.644	0.729	0.721	0.718

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Green preference measure based on the sample of individuals' tweets excluding retweets. The dependent variable is the share of votes dedicated to the Green party in the 2018 Riksdag elections. Green Pr. Wom. (resp. Men) gives the share of female (resp. male) users with a share of environment-related tweets higher than $\alpha_1 = 0\%$; $\alpha_2 = 2.8\%$ (median); and $\alpha_3 = 5\%$. Income Women (resp. Men) is given by the median income of men (resp. women) in the municipality measured in hundred thousand Swedish Crowns. % 20-35 yo gives the share of inhabitants aged 20 to 35 years old. Population density is measured as the number of inhabitants per square meter. Female pop refers to the share of women in the municipality.