Temperature and Temperament: Evidence from Twitter

Patrick Baylis*

December 23, 2019

Abstract

How do people value their climate? This paper demonstrates a new approach to estimating preferences for nonmarket goods using social media data. I combine more than a billion Twitter updates with natural language processing algorithms to construct a rich panel dataset of expressed sentiment for the United States and six other English-speaking countries around the world. In the U.S., I find consistent and statistically significant declines in expressed sentiment from both hot and cold temperatures. To better understand how preferences may adapt, I document heterogeneity in both regional and seasonal responses. I complete the U.S. analysis with a suite of validation exercises to understand the magnitude of these effects and two methods to estimate willingness-to-pay for climate amenities. Finally, I document similar relationships between temperature and expressed sentiment for four out of the six non-U.S. countries I examine.

*University of British Columbia, Vancouver School of Economics; Iona 113, 6000 Iona Dr., Vancouver, BC V6T 2E8; patrick.baylis@ubc.ca. I am grateful to Maximilian Auffhammer, Severin Borenstein, and Solomon Hsiang for their invaluable suggestions, as well as to Michael Anderson, Hunt Allcott, Judson Boomhower, Josh Blonz, Marshall Burke, Fiona Burlig, Tamma Carleton, Richard Carson, Aluma Dembo, Meredith Fowlie, Walter Graf, Sylvan Herskowitz, Nick Obradovich, Elizabeth Sadoulet, seminar participants at UC Berkeley, the AERE Annual Conference, the CU Environmental and Resource Economics Workshop, the Heartland Workshop at Illinois, and several anonymous referees. This paper was previously circulated under the title "Temperature and Temperament: Evidence from a billion tweets."

As the possibility of substantial changes to Earth's climate becomes more certain, economists have become increasingly interested in calculating the full scope of benefits and costs resulting from these changes. While much of the literature has focused on changes in welfare from increases in temperature on aggregate incomes, morbidity and mortality, civil conflict, and agricultural profits among others, relatively little work has examined individuals' underlying preferences for different climates and, by extension, for the amenity value implications of climate change itself. This gap is the result of a classic problem in nonmarket valuation: because ambient temperature is non-rival and non-excludable, there are no direct markets from which researchers might infer preferences for different climates.

Instead, the existing literature has relied on hedonic valuation approaches to indirectly estimate individuals' willingness-to-pay for different climate characteristics by observing how housing prices vary with climatic conditions. These approaches generally estimate that individuals would pay between 1% and 4% of their annual incomes to avoid projected end-of-century increases in temperature (Sinha, Caulkins, and Cropper 2018; Albouy et al. 2016). However, because the climate to date has varied relatively little across time, these values are necessarily identified using cross-sectional differences in climate. Preferences for climate that are robust to potential cross-sectional contamination are key parameters for the development of local, national, and global public policies relating to climate change.

This paper demonstrates a new method to estimate preferences over nonmarket goods that allows researchers to include controls for unobservables across both time and space: I construct a spatially and temporally rich dataset on daily expressed sentiment, or emotional state, and estimate the relationship between sentiment and outdoor ambient temperature.

Section 1 discusses how this approach relates to existing methods designed to elicit preferences for nonmarket goods. Section 2 describes the construction of the dataset, which begins with a geographically and temporally dense collection of more than a billion geocoded social media updates (hereafter, "tweets") from the online social media platform Twitter. I measure the expressed sentiment of each tweet using a set of natural language processing (NLP) algorithms designed to extract sentiment, or emotional state, from unstructured text data. For computational tractability and to account for noise in the estimation of expressed sentiment, the primary analysis takes daily averages for each Core-Based Statistical Area (hereafter, CBSAs) as the unit of observation, although I also estimate a model with individual tweets as the unit of observation to test for compositional effects. Because of the uncertainty inherent in estimating underlying emotional state from language, I compile four separate measures of sentiment using word lists constructed using previous research in NLP and, in three of four cases, specifically intended to extract sentiment from "microblogs" such as tweets.

The analysis in Section 3 then uses the geographic information attached to the tweets in my dataset to match measures of sentiment to daily weather conditions at the location of the user. The identifying assumption in the econometric model I estimate is that temperature realizations are as good as random after accounting for spatial and temporal fixed effects. Allowing temperature to enter the model flexibly, I find consistent evidence of an upside-down "U" shape: a roughly symmetric decline in sentiment away from moderate temperatures, with peak sentiment occurring roughly around 21.0 C (69.8 F). The point estimate of the difference in expressed sentiment between 20-25 C and above 40 C is a statistically significant and between 0.1 and 0.2 standard deviations (SD), depending on measure used. The responses of expressed sentiment to temperature are markedly similar across choice of measure, and both qualitatively and quantitatively consistent across a range of different specification choices. To discern the mechanism by which sentiment responds to temperature, I also estimate the relationship between online profanity and temperature and find a U-shaped relationship there as well, suggesting that aggression is at least part of the explanation for the decline in expressed sentiment in both hot and cold temperatures.

To better understand how preferences for climate are formed and updated, I extend the baseline results by examining regional and seasonal heterogeneity in the response of sentiment to temperature. I find notable differences in both. Regionally, areas that are colder tend to have stronger response to warm temperatures, and vice-versa. Seasonally, the responses suggest preferences for cooler temperatures in summer and fall and warmer temperatures in winter, with relatively little sentiment response to temperatures on the spring.

Section 4 documents a set of validation and valuation exercises designed to aid interpretation of the main results, as well as a projection exercise to estimate future damages. I begin by showing how average sentiment changes over the course of the week, and that the difference in expressed sentiment between a Sunday and Monday is roughly 0.1 SD. Next, I present additional empirical exercises identifying the effect of plausibly random variation in hurricanes, American football outcomes, quarterly wages, and receipt of parking or speeding tickets on observed social media sentiment. I conclude by using the value of expressed sentiment from the ticket estimate to value of the amenity losses due to climate change, which I find to be between 0.6% and 2.5% of present-day annual income by end of century.

Finally, Section 5 estimates the relationship between temperature and expressed sentiment in Australia, India, South Africa, the Philippines, Kenya, and Uganda, six English-speaking countries for which I am able to obtain sufficient data from Twitter and with adequate temperature variation. Compared to the United States, I estimate similar preferences for temperature in Australia, India, South Africa, and the Philippines, but not in Kenya or Uganda, suggesting that the effects I estimate are not isolated to the unique case of the United States, but that caution should be taken when extrapolating these results to the entire world.

This paper makes several contributions to the literature. It is the first to identify sentiment-analyzed social media posts as a source of information on latent individual preferences for environmental goods.[1] I identify a response function of sentiment to temperature that concurs qualitatively with existing work and is robust to a wide range of statistical specifications, and the methods I document provide a tractable roadmap for future work estimating preferences for nonmarket goods from social media posts. Second, the paper identifies sharply diverging seasonal and regional preferences for temperature, suggesting an important adaptive component to the baseline observed response. Third, the set of validation and valuation exercises I employ serve as novel empirical exercises in their own right and provide initial steps in valuing the impacts observed here. Fourth and finally, I provide estimates of similar patterns in four out of the six English-speaking countries for which I am able to obtain suitable data, suggesting the preferences for climate that I observe may well be global.

1 Background

Economists have studied the economic impacts of climate change for more than two decades (Nordhaus 1991; Cline 1992), but the increasing availability of a range of panel datasets have made possible the identification of the causal effects of changes in temperature on a diverse set of economic outcomes, including crop yields, economic production, civil conflict, mortality, migration, and many others (Carleton and Hsiang 2016). In the absence of historical changes in long-run climate, researchers have used estimates of the changes in these outcomes resulting from plausibly exogenous historical variation in temperature in order to predict future damages from climate change (Dell, Jones, and Olken 2014). The assumptions required for this extrapolation are formally derived in Hsiang (2016), but intuitively the central concern is whether or not adaptation behaviors require large fixed costs. This question is likely to be answered differently for different sectors, but for those in which it is econometrically possible to estimate a "long-differences" ap-

1. In a follow-on paper, we extend this method to include Facebook posts as well (Baylis et al. 2018).

proach, e.g., Burke and Emerick (2015), few differences have been found between estimates produced using short-run or long-run variation in temperature.

This work has had an impact on public policy. Many of the estimated outcomes contribute, directly or indirectly, to aggregate measures of the total cost of climate change produced by summary reports (Stern 2006; Houser et al. 2014) and integrated assessment models (IAMs), which in turn are inputs to the United States government's estimate of the social cost of carbon, or SCC (Interagency Working Group on Social Cost of Carbon 2013).[2] By 2014, the then-central value of \$36 per ton of CO_2 equivalent had been incorporated into 79 U.S. regulations as part of required benefit-cost analyses conducted in the course of the federal rule-making (United States Government Accountability Office 2014).[3]

Different areas of the world will experience climate change in very different ways. Coastal areas will face rising sea levels and major economic impacts from typhoons or hurricanes (Hsiang 2010). Farmers are likely to experience substantial changes in the yields of major crops (Schlenker and Roberts 2009), and many areas in the developing world where subsistence farming is a major source of calories could experience catastrophic droughts and resulting food security crises (Lobell et al. 2008). For others, the impacts of climate change will be more subtly felt: instead of increases in large-scale natural disasters or acute economic crises, most of the world will simply experience a steady increase in average temperatures (IPCC 2014). Prior work has projected the impact of these gradual changes on income (Deryugina and Hsiang 2014), crime (Ranson 2014), mortality (Deschênes and Greenstone 2011), and other outcomes. This paper focuses instead on the welfare cost of changes in amenity values resulting from rising outdoor temperatures.[4]

Traditional approaches to calculating the welfare impact of a policy change date back as far as Marshall (1890) and rely on knowledge of either the demand curve, the supply curve, or both. For private goods with well-established markets, the shapes of these curves can be estimated using plausibly exogenous supply or demand shifters and from those the change in welfare due to a change in policy can be calculated. Estimating changes in welfare due to changes in the allocation of public goods, or nonmarket goods more generally, has proven to be more challenging due to the absence of available markets. Nevertheless, a handful of approaches to this problem have emerged, many within the

2. Three IAMs are used to derive the social cost of carbon: DICE, FUND, and PAGE. Diaz (2014) provides a more detailed comparison of how these models are built.

3. The Environmental Protection Agency, under direction from the Trump administration, reduced the SCC to between \$1 and \$6 per ton, largely because it used only damages to the United States in its accounting for the costs of carbon emissions.

4. To date, only the DICE model directly incorporate estimates of the costs or benefits from climate as an amenity.Nordhaus and Boyer (2000) finds that 2.5 C warming will result gain of about 0.3% of GDP.

environmental economics literature (Pearce 2002).

Climate can be viewed as a public amenity[5]: it is non-rival (a single person's consumption of climate does not reduce the amount of climate available to anyone else) and non-excludable (no person cannot be prevented from consuming climate), and although individuals can alter their local climates at home and at work, the outdoor ambient temperature is determined by factors outside of their control. Hedonic price approaches provide a method to value amenities like climate: recent work by Sinha, Caulkins, and Cropper (2018) and Albouy et al. (2016) identify implicit values for different climates using observed household decisions about where to live. These approaches are useful in part because it is straightforward to back out monetary valuations of different climates from the model estimates. However, since historical changes in climate have thus far been fairly modest, the estimates from these models must be identified using cross-sectional variation. As a result, unobserved spatial variation such as cultural norms, geographic factors like proximity to oceans or mountains, or other unobserved amenities that correlate with climate could bias these estimates in unknown directions.

A related approach to understanding preferences is to use surveys of subjective well-being (SWB) to estimate preferences over temperature. These surveys ask respondents to assess their well-being on a single dimensional scale (Diener 2000; Dolan, Peasgood, and White 2008). Kahneman and Krueger (2006) and Mackerron (2012) discuss the merits and weaknesses of these studies: a common challenge is that measurements of SWB are by definition subjective and likely to include unobserved variation across time and space. For example, responses to questions about one's well-being may depend on regional dialects or norms, or could be driven by the interaction between the interviewer and the interviewee, which may itself be affected by temperature.

The estimates of the effect of temperature on SWB vary widely within the literature. Most studies use cross-sectional variation or follow a very small group of individuals over time. To my knowledge, only two control for unobservable cross-sectional variation using panel data methods. Feddersen, Metcalfe, and Wooden (2012) use nearly 100,000 observations from Australian SWB surveys to compare the effects of short-term weather and long-term climate on life satisfaction. Since individuals are observed more than once in their data, they are able to control for individual fixed effects for some specifications. They find that weather affects reported life satisfaction through solar exposure, barometric pressure, and wind speed, but do not find impacts from changes in temperature itself. Dennisenn et al. (2008) uses an online survey to find that weather impacts are

5. I use the term "public amenities" here to indicate that changes in the climate can be either goods or bads, depending on the region and sector of the world in which they occur.

variable across individuals, but that those variations do not correspond to observable characteristics.

A small literature attempts to assign a monetary value to environmental goods using self-reported happiness data (Welsch and Kühling 2009). For example, Rehdanz and Maddison (2005) estimate the relationship between climate and self-reported happiness, and include a valuation method based on country-level GDP. Levinson (2012) conducts a similar exercise to estimate WTP to avoid pollution using happiness data, but includes weather as a covariate. Because these studies implicitly rely on income as an exogenous driver of happiness, this approach could induce bias if that assumption does not hold (Mackerron 2012).

The method of assessing preferences for nonmarket goods I describe in this paper relies on the assumption that contemporaneous changes in expressed sentiment deliver insights into individuals' underlying preferences for these goods. I have described previous work that assesses these preferences and the challenges they face in controlling for unobservable sources of cross-sectional variation. The approach in this paper mitigates the problem of unobserved correlates over time and space, allows for flexible estimation of non-linear effects, and even provides sufficient data richness to examine geographic and seasonal variation in the response of expressed sentiment in order to better understand adaptation.

Conceptually, one way to view expressed sentiment is as an estimate of "experienced utility". The concept of experienced utility predates the modern neoclassical definition of utility, which for clarity and following Kahneman and Sugden (2005), I refer to hereafter as "decision utility". Whereas decision utility is an ordinal description of the value obtained from bundles of goods, experienced utility follows is an instantaneous measure pleasure and pain Bentham (1789). Discussions of which measure of utility is the appropriate metric for welfare analyses is beyond the scope of this paper, but for the purposes here it is sufficient to view expressed sentiment, like subjective well-being or experienced utility, as a useful proxy for individual preferences. In the following section, I describe how this paper estimates expressed sentiment and document descriptive statistics that suggest its relationship to underlying preferences.

2 Data

While it would be prohibitively expensive to estimate daily sentiment across the United States using a survey, publicly available updates on social media provide a low-cost alternative. By combining a large set of geolocated tweets with sentiment analysis algorithms (NLP algorithms designed to elicit emotional state), I am able to measure daily variation in expressed sentiment across the United States. In this paper, I combine this data with meteorological observations to estimate the sentiment response to temperature. Previous work in computer science has estimated models that related expressed sentiment to meteorological variables (Hannak et al. 2012), but to the best of my knowledge this is the first paper to do so in a causal framework and in order to elicit underlying preferences for temperature. The following section describes the construction of the measures of sentiment and the weather covariates. Table 1 summarizes the variables included in the empirical model. The first panel shows the count, mean, median, minimum, and maximum of the measures of sentiment, the second panel describes the weather data used, and the third panel summarizes the number of tweets by CBSAs and by individuals in the data.

[Table 1]

2.1 Twitter data

Created in 2006, Twitter is a social media platform where users exchange brief updates, otherwise known as tweets. Since its founding, Twitter has become one of the most popular such platforms worldwide, with 288 million active users sending over 500 million tweets per day as of 2015.[6]

Twitter's Streaming API is designed to give developers access to the massive amount of data generated on the Twitter platform in real time. Starting in June 2014, I began collecting geolocated Twitter updates from within the continental United States using a client that is continuously connected to the Streaming API.[7] I collect the vast majority of geolocated tweets produced within my sample period, which ends in October 2016.

Geolocated tweets are those for which the user has consented to have his or her location information shared for that post. The location information is either produced using the exact latitude and longitude or from a reverse-geocoding algorithm that derives the latitude and longitude from a general location (e.g., a neighborhood) entered by the user. Geolocations are assigned on a by-post basis. In principle, Twitter limits the total number of tweets delivered through the Streaming API to 1% (Morstatter, Pfeffer, Liu, and Carley 2013) of the total tweets created. Since I request only geolocated tweets from

6. Per Twitter's website, accessed September 2015.

7. More details on the collection process are given in Appendix A.1.1. There are two substantial gaps in the time series, from June 26th to July 12th, 2014, and from September 18th to October 27th, 2014, and a small number of gaps of a few days. These gaps correspond to periods of time when the streaming client was unable to connect to the Streaming API.

within the United States, this total infrequently comes to more than 1% of the total tweets worldwide (geocoded and otherwise). As a result, over the course of the days in which the streaming client was operational, the percentage of missed tweets is fewer than 0.01% of the total geolocated tweets within the United States. I include tweets only from users who tweet less than 25 times a day or less to reduce the proportion of non-human accounts in the sample. The left panel of Figure 1 maps the total tweet volume in my sample across the United States, where pixel shading represents the logged volume of tweets. There is considerable spatial variation in Twitter activity, and most activity occurs in cities. The map also captures the extent to which this activity follows human movement patterns: along with cities, major highways and roads are readily visible in the map.

[Figure 1]

The translation of unstructured text data into quantitative data is known as "natural language processing", or NLP. Within NLP, the set of techniques designed specifically to quantify expressed sentiment is called "sentiment analysis." At the time of this writing, there are more than fifty publicly available algorithms and/or wordlists used to conduct sentiment analysis (Medhat, Hassan, and Korashy 2014). Because the method by which these measures are constructed can differ substantially, analyses using expressed sentiment should ideally demonstrate reasonable consistency across multiple measures. In this paper, I translate tweet content into four measures of expressed sentiment derived from prior work: AFINN (Nielsen 2011), Hedonometer (Dodds and Danforth 2010), LIWC (Pennebaker et al. 2015), and VADER (Gilbert and Hutto 2014). The underlying machinery for each measure is similar: each contains a word list, or dictionary, which contains sentiment scores that correspond to English-language words. The overall measure of sentiment in a piece of text is the average of all scored words within that piece of text.

Panel A in Table 1 describes the unstandardized sentiment measures in the sample, although I standardize the measures prior to analysis for comparability. Following prior work, I pre-process each tweet before scoring in order to increase the precision of the NLP algorithms (Pak and Paroubek 2010). I remove punctuation, URLs, hashtags (e.g., "#job"), and mentions (e.g., "@person") to isolate the word selection in the tweet. Because the independent variable of interest is weather, I remove tweets that contain any weather-related terms (see Table A.4 for the list of weather teams I exclude) to ensure that the responses do not capture the sentiment of observations about the weather, only changes in general sentiment due to weather. Once the tweets have been pre-processed, I score them for sentiment using the pre-existing dictionary (AFINN, Hedonometer, and LIWC), with the exception of VADER, which contains its own pre-processing routines. Appendix A.1.2 gives background and additional detail for each measure. Finally, in addition to the sentiment measures I include a profanity measure intended to capture the use of online vulgarity, changes in which are reported as a percentage of average profanity used in the sample.[8]

Table 2 shows the correlations between the five measures at the CBSA-day level.[9] All of the measures are strongly positively correlated with each other, except the measure of profanity which is negatively correlated with all measures. The measures capture substantial geographic heterogeneity: the right panel of Figure 1 documents average sentiment (as measured by VADER) by CBSA across the country. As in the state averages, urban CBSAs and CBSAs in the northern part of the United States show higher average sentiment relative to rural CBSAs and CBSAs in the southern part of the country.

[Table 2]

2.2 Weather data

To obtain local estimates of daily weather across the contiguous United States, I use the PRISM Climate Group's AN81d gridded weather dataset. These data provide daily measures of minimum temperature, maximum temperature, and precipitation at roughly 4×4 kilometer grid cells for the entire United States. The data are produced using a model that interpolates measurements from more than 10,000 weather stations and corrects for altitude and other influences on local climate (Daly et al. 2002). The second panel in Table 1 describes sample statistics for the PRISM data. I aggregate the gridded data to the CBSA level using population weights to ensure that the weather covariates reflect the average weather experienced by individuals within each CBSA.

Prior work suggests that other weather variables besides temperature and precipitation may be drivers determinants of emotional state (Dennisenn et al. 2008). Accordingly, I also gather daily data on the proportion of day that was overcast, relative humidity, station pressure, and wind speed from 2,162 weather stations included in the NOAA Quality Controlled Local Climatological Data, or QCLCD.

8. Figure A.2 shows state average expressed sentiment, ordered from lowest to highest. While cross-sectional differences in expressed sentiment provide limited causal insights, the plot suggests that colder regions express higher sentiment on average: the top six states are Montana, New Hampshire, Vermont, Wyoming, Minnesota, and South Dakota.

9. **??** shows cross-sectional correlations between the measures at the state level, as well as comparisons to other measures of subjective well-being.

3 Estimating preferences for temperature

This section empirically estimates the expressed sentiment response to temperature in order to identify preferences. Section 3.1 describes the empirical approach and identification strategy, while Section 3.2 documents the findings across a range of measures of expressed sentiment, empirical specifications and sampling frames. Section 3.3 uses the richness of the data to explore whether climate preferences are likely to adapt over time.

3.1 Empirical approach

I identify the causal effect of temperature on expressed sentiment using a panel fixed effects model, with temperature entering the regression using a flexible functional form. This flexibility is justified for the following reasons: first, prior work estimating temperature has documented non-linearities across a wide array of responses to temperature (Carleton and Hsiang 2016), second, an appropriate flexible functional form should reveal the shape of the underlying response function, linear or otherwise (Hsiang 2016) and third, intuition suggests that there is some bliss point for temperature, if only because temperatures which threaten human survival are clearly not preferable. The value of the panel nature of the dataset is that it allows me to control for unobservable cross-sectional and seasonal variation. Specifically, I estimate the following statistical model:

$$\overline{S}_{cd} = f(T_{cd}) + P_{cd} + \phi_c + \boldsymbol{\phi}_{\text{time}} + \varepsilon_{cd} \tag{1}$$

Let c and d index CBSA and date. $\overline{S}_{cd}$ is the CBSA-day average of one of the four measures of sentiment described in Section 2. T_{cd} is the maximum daily temperature in a CBSA, and $f(T_{cd})$ is a flexible function of temperature, which I implement in practice using a binned model specification to allow for nonparametric responses of expressed sentiment to temperature. In particular, I let $f(T_{cd}) = \sum_b^B \beta_b T_{cd}^b$, where T_{cd}^b is an indicator variable equal to one if T_{cd} falls in the given bin b. P_{cd} is daily precipitation. ϕ_c represents CBSA fixed effects and $\boldsymbol{\phi}_{\text{time}}$ represents a set of additional temporal controls, including month, year, day of week, and holiday fixed effects, as well as state-specific time trends. ε_{cd} is the idiosyncratic error term, clustered by both CBSA and date. I estimate the model using weighted least squares, where the weights are the average number of tweets in the CBSA.

T_{gd}^b specifies one, three, or five degree bins running between 0 to 40 degrees C, with edge bins for all observations with maximum temperature less than 0 or greater than 40.[10]

10. Because three does not multiply evenly into 40, the upper limit for the three degree bin specification is 39 C.

I include both three and five degree versions of this model as part of the main results I present in the paper, and a comparison of all three bin widths in the appendix. For all bin widths, I choose the bin that contains 22.5 C as the omitted category. This choice does not alter the shape of the estimated response function, since relative differences between conditional means are preserved, but it does reflects the prior finding that Americans prefer 65 F (18.3 C) average daily temperature (Albouy et al. 2016). In my sample, because I use daily maximum temperatures rather than average temperatures, this corresponds to the omitted bin that I choose.

As shown earlier, the right panel of Figure 1 documents cross-sectional variation in sentiment. Although all regions have a mix of high and low-sentiment CBSAs, visual inspection suggests that there is substantial regional variation in expressed sentiment. Additionally, prior evidence suggests that individuals with higher incomes tend to experience higher levels of life satisfaction and can afford to locate in areas with generally pleasant climate (Easterlin 2001). If this regional variation, which may result from cultural or economic factors, correlates with regional weather differences, a naïve estimate of the relationship between weather and expressed sentiment is likely to be biased. To account for this regional variation in sentiment, I include CBSA fixed effects ϕ_c. These fixed effects ensure that the model is estimated on deviations from CBSA averages rather than on cross-sectional differences in climate, which could correlate with average sentiment or lexical patterns that register as different sentiments. Intuitively, the implication of this modeling choice is that the estimates represent a weighted average of within-CBSA comparisons, e.g., the difference in sentiment in Dane County, WI on a hot day versus a cold day.

A second concern addressed by this identification strategy is the seasonality of both sentiment and temperature. To account for this possibility, I include month of year fixed effects as part of $\boldsymbol{\phi}_{\text{time}}$. Intuitively, this choice of fixed effects implies that the model coefficients represent a weighted average of the differences in sentiment on hot days versus cold days within, e.g., Chicago in June. State time trends and year fixed effects account for potentially correlated trends in both temperature and sentiment that are shared across the sample, while day of week and holiday fixed effects remove statistical noise related to within-week and by-holiday variation in expressed sentiment.

The combination of these fixed effects defines the identification strategy: at most, I assume that deviations in weather are as good as random after accounting for unobserved variation by CBSA, month of year, and year. This assumption is typical of the climate impacts literature (Hsiang 2016). I also estimate alternative specifications with differing sets of fixed effects. Conditional on the assumptions given above, the coefficients of interest β_b can be interpreted as the average change in sentiment resulting from replacing a day in the omitted bin with a day in temperature bin b.

3.2 Findings

I find statistically significant declines in expressed sentiment resulting from both low and high temperatures. Section 3.2.1 documents the baseline findings from Equation (1) across a range of measures of expressed sentiment and specification choices. Section 3.2.2 undertakes a disaggregated analysis using tweets as the unit of observation in order to test for compositional sorting, and Section 3.2.3 identifies the degree to which the use of profanity responds to temperature. I discuss each in turn.

3.2.1 Baseline estimates

For expositional clarity, I first present the main result for each sentiment measure in Figure 2. I show that the shape of the response functions is remarkably similar across the different measures of sentiment. Second, Table 3 tabulates the response of the VADER measure under a range of different choices of fixed effects.

[Figure 2]

Figure 2 documents the temperature response of all four measures of sentiment estimated using Equation (1). Because each outcome measure is standardized to have mean zero and unit standard deviations, the point estimates β_b represent the change in the conditional mean of expressed sentiment, measured in standard deviations, expected as a result replacing a day with a high of 21-24 C with a day with a high in bin b. I include a histogram underneath each plot to demonstrate the support of the temperature distribution. Each panel includes all four sets of point estimates, with the darker line indicating the central estimate and the dotted lines indicating the 95% confidence interval around that estimate indicating the measure given in the subtitle. The other estimates are included as light gray lines without confidence intervals for comparison.

The upper-left panel documents a decline in the AFINN sentiment measure below 12 C and above 30 C. The difference in sentiment between days with the coldest temperatures (< 3 C) and days in the omitted bin is around 0.15 SD, similar to the difference in sentiment between very hot days (> 39 C) and days in the omitted bin. Confidence intervals are slightly wider for cooler temperature estimates but the point estimates are statistically different from zero at both ends of the temperature range. The AFINN measure estimates the second largest cold-weather effect and the largest warm weather effect.

The upper-right panel estimates a similar response shape for the Hedonometer measure, both in shape and in magnitude. There is a slight uptick in sentiment at 12-15 C, but this is not a statistically significant difference. Point estimates are statistically different from zero on both ends of the temperature scale. The Hedonometer measure estimates the second largest warm-weather effect and the largest cool weather effect.

The bottom-left panel documents the response for the LIWC measure. While still within the confidence intervals of the other estimates, LIWC documents more limited impacts of cold temperatures on sentiment. The point estimates are statistically significant and similar to the other measures for warm weather temperatures, but smaller and not statistically significant for colder temperatures. This difference from the other measures may result from the measure's lack of suitability for the microblogging format.

The bottom-right panel documents the response as measured using VADER. Like the AFINN and Hedonometer measures, VADER estimates a statistically significant decline in sentiment below 12 C and above 30 C that reaches about 0.2 SD at maximum. VADER estimates a similar response to AFINN or Hedonometer and but a larger response than LIWC.

Each outcome measure in Figure 2 documents a statistically significant negative relationship between sentiment and hot temperatures, relative to a day with moderate temperatures. The magnitudes of the effect sizes differ, ranging from about 0.1 SD to more than 0.2 SD for the hottest temperature bin. The relationship between sentiment and cold temperatures is slightly less precisely estimated, and one of the four measures fails to reject the null of no difference between cold and moderate temperatures, although the consistent decline of the point estimates provides suggestive evidence of a negative effect in low temperatures. Despite these differences, the results of this exercise are markedly similar across measures: each exhibits the same upside-down "U" shape, each reaches similar magnitudes on both the cold and warm temperature ends of the temperature spectrum, and each is statistically significant at both of those ends (with the exception of LIWC in cooler temperatures).

Because the response functions are consistent across measures, the remainder of the paper focuses on results obtained using VADER. Table 3 estimates the effect of temperature on expressed sentiment using five degree C bins and across a range of choices of fixed effects. All columns include CBSA fixed effects.

[Table 3]

Column (1) reflects the baseline specification, which also includes month and year fixed effects. As in Figure 2, I observe negative and statistically significant effects below 10 and above 30. Column (2) adds day of week and holiday fixed effects to absorb weekly variation in sentiment (see Figure 5) and variation related to holidays, but the estimated effects are virtually unchanged. Column (3) introduces state-by-month fixed effects to account for regionally distinct seasonal trends. This set of fixed effects is particularly restrictive, as the model is now identified solely off of variation within a state-month and because most states have only a few CBSAs. Qualitatively, I find that the point estimates again identify an upside-down U-shape impact of temperature on expressed sentiment. However, the impact of cooler temperatures is not statistically different from zero under this specification, and the point estimates for both cold and hot temperatures are partly attenuated towards zero. These estimates likely result from the influence of the more restrictive state-month fixed effects, which reduce the residual variation available in the model more than any other specification (see Figure A.6). As described in Angrist and Pischke (2008), classical measurement error in fixed effect models can lead to a biasing of the coefficients towards zero as more and more restrictive fixed effects groupings are considered. To further investigate this claim, column (4) replaces the state-by-month fixed effects with month of sample fixed effects. Under this specification, the cold temperature estimates are again large and significant. Finally, column (5) replaces the state-by-month fixed effects from column (3) with a more flexible set of state trends and finds largely similar results, although the point estimates are somewhat larger for the cold temperature estimates.

The negative relationship between temperature and sentiment below 12 C and above 30 C resembles that estimated by Albouy et al. (2016), who find that individuals pay to avoid warm temperatures in summer and cold temperatures in winter. The preferred model estimates the magnitude of the difference between a moderate day and an extremely cold or hot day to be around 0.1 SD and 0.2 SD, respectively.

Broadly, I find qualitatively similar results across a range of specifications. Both hot and cold temperatures have a negative effects on expressed sentiment. In addition to the test for composition sorting described in Section 3.2.2, the appendix includes more sensitivity checks, including the inclusion of additional weather variables (Table A.7), variations on bin width (Figure A.5), different sampling frames (Table A.8, and weighting choices (Table A.9) none of which qualitatively alter the baseline results.

3.2.2 Compositional sorting

Because Twitter users choose when — and when not — to tweet, the selection mechanism into the sample could induce a compositional bias in the estimates observed in Figure 2, a sample selection effect akin to that described by (Heckman 1979). This can also be viewed as a form of the ecological fallacy: the observation that the properties of aggregated groups may not reflect properties of the individuals in the underlying populations (Robinson 1950). To fix ideas, imagine two types of Twitter users: positive and negative. Positive users create only positively-scored tweets, while negative users create only negative tweets. Because neither type will change the content of its tweets in response to temperature, the true underlying effect of temperature on their sentiment in zero. However, suppose as well that positive users choose to put their phones away when it's very cold or very hot, whereas negative users are unaffected. An econometric approach using CBSA averages that does not control for the type of user will in fact pick up this change in the sampling frame rather than the true effect.

Since the data I collect include an identifier for the tweet creator, I can account for compositional sorting in my sample using post-level data and user fixed effects. To do so, I estimate the following model:

$$E_{id} = \sum_{b \neq 20\text{-}25}^{B} \beta_b T_{cd}^{b} + \phi_i + \phi_m + \phi_y + \varepsilon_{id} \tag{2}$$

This model replaces CBSA fixed effects with user fixed effects ϕ_i in equation Equation (1). The model requires the use of the disaggregated sample of tweets in my dataset; for computational reasons, I focus on users who tweet during 25% of the days in my sample but who produce fewer than 25 tweets per day, resulting in 432 million tweets. Table 4 compares the results between models using CBSA averages and individual posts as the unit of observation.

[Table 4]

The fourth column documents the estimates using the individual sample. For computational reasons, this estimate is obtained using user, CBSA, month, and year fixed effects, but does not include state trends or day of week and holiday fixed effects. In order to compare this estimate to the main sample, the first three columns document results using the CBSA-by-date aggregated dataset, where the first column in Table 4 is equivalent to the fifth column in Table 3. The second column drops the state trends and day of week and holiday fixed effects, but maintains the baseline sample. The third column restricts the CBSA sample to the sample used in the individual-level estimate.

I find qualitatively similar results for these measures using the individual sample, although the estimates for higher temperatures are slightly attenuated in the individual fixed effects model relative to the models using CBSA-date averages. It is possible that this is evidence of some compositional sorting at higher temperatures, but could also be the result of measurement error driven by using a sparser source of variation as a result of the user fixed effects. In either case, the shape of the results is largely similar across the two models and the results do not appear to be primarily driven by compositional changes resulting from temperature variation.

3.2.3 Profanity response

A large literature has documented the impact of climate on conflict (Burke, Hsiang, and Miguel 2015). One possible mechanism is the finding that warm temperatures encourage aggressive behavior (Kenrick and MacFarlane 1986). To understand whether the expressed sentiment response to temperature is due in part to this aggression mechanism, I estimate the relationship between temperature and expressions of profanity. Using a list of more than 300 profanities, I estimate Equation (1) with the occurrence of tweets in a CBSA-day that contain a profanity as the outcome of interest. One concern with this approach may be that if users are simply using more profanities to reflect their mood, these effects could represent same decline in emotional state captured by Figure 2. To investigate this possibility, I also construct an "aggressive profanity" metric that counts only the number of tweets that included a popular, aggressively profane phrase. Figure 3 plots the results, scaling each measure by its own mean to obtain the percent change in tweets using the selected profane terms.

[Figure 3]

I find that use of profanity and aggressive profanity rises in both hot and cold temperatures. Previous work on both conflict (Burke, Hsiang, and Miguel 2015) and on violent crime (Ranson 2014) find that both increase during periods of high temperatures. That I document a similar effect for hot temperatures aligns with the hypothesis that increases temperature induce violence by making individual more aggressive. However, I also find that cold temperatures induce more profane text than moderate temperatures. This finding is in contrast to previous work on temperature and aggressive behavior, which has not typically found an increase in crime or conflict during periods of cooler temperatures (Ranson 2014; Burke, Hsiang, and Miguel 2015). It may be that aggressive impulses increase in response to temperature discomfort of both kinds, but that cooler temperatures limit opportunities to act on that aggression.

3.3 Understanding adaptation

As I discuss in Section 1, the climate impacts literature has identified a range of settings in which variation in temperature has had both statistically and economically significant impacts on economic outcomes of interest. The question of whether and to what extent these impacts can be extrapolated to climate change is critically important for projecting cumulative economic impacts. The spread of humans across the planet suggests that, in the long run at least, humans are highly capable of surviving in a wide range of environments. The relatively slow pace of climate change invites the possibility that many of the measured impacts could be partly mitigated by either adaptive responses or by sorting.[11] Empirical estimation of adaptation has presented substantial challenges for researchers working in this area: direct, causally identified models usually rely on long-differences methods as in Burke and Emerick (2015), which in turn rely on sufficient long-run variation in temperature and the outcome of interest across a large geographical area. For most studies, including this one, the requirement of a multi-decadal panel dataset for proper estimation of long-run effects is unattainable. Even for studies with such a dataset available, the research design effectively reduces the number of available observations to the number of observed geographical units, which restricts statistical power and reduces the ability of researchers to strongly reject large portions of the parameter space.

As an alternative to providing direct evidence on adaptation or sorting, in this section I take advantage of the richness of the dataset to run two empirical tests designed to suggest whether preference adaptation or sorting is likely to occur in this setting. Below, I estimate the degree of heterogeneity in temperature-sentiment responses both across the four quartiles of average annual temperatures and across the four seasons of the year, finding important differences in the sentiment response across these dimensions. While these are not sharp tests of adaptation, they do help to inform the extent to which temperature preferences do or do not adapt over time.

Figure 4 estimates separate splined models by region, where each panel identifies the response for the given region. Regions are split by quartiles of average annual temperature, with labels given in order as "Coldest", "Cold", "Warm", and "Warmest". In order to mitigate the loss of statistical power that results from estimating regional models, I use a splined model of temperature (separately estimated for the full sample in Figure A.3). To document uncertainty, I bootstrap these estimates with 1000 iterations of the same specification, sampling from the full dataset with replacement. For each region, the

11. There is a burgeoning literature on understanding adaptation. For more complete discussions of the subject across a range of areas, see Auffhammer et al. (2013), Houser et al. (2014), Graff Zivin, Hsiang, and Neidell (2018), Auffhammer (2013), and Dell, Jones, and Olken (2014).

red line indicates the response identified for the entire sample for that season and the lighter gray lines indicates the bootstrapped responses. All lines are normalized such that their maximum value is equal to 0. Clear differences in the sharpness of the sentiment response to temperature can be observed across regions: colder regions have attenuated responses to cold temperatures, while warmer regions have attenuated responses to warm temperatures. Conversely, colder regions respond more to high temperatures and warmer regions less to cold temperatures.

[Figure 4]

The regional heterogeneity in responses is evidence of either preference adaptation, technological adaptation, sorting, or some combination of all three but I cannot distinguish between these. Individuals may have adapted their preferences to accommodate their climatic zones, they may have a greater degree of technologies available to mitigate extreme temperatures to which they have become accustomed (e.g., air conditioning and indoor heating), or individuals with stronger preferences around lower or higher temperatures may have chosen to vote with their feet, so to speak.

To help distinguish between these explanations, I also estimate the differences in seasonal responses, where the possibilities for technological adaptation and sorting are more limited. Appendix A.2.2 documents differences that are consistent with common intuition: cooler temperatures are preferred in fall and summer, warmer temperatures are preferred in winter, and moderate temperatures are preferred in spring. The combined regional and seasonal dependencies of preferences for temperature are suggestive of adaptive possibilities for temperature, and articulate the importance of projecting climate impacts in a way that allows preferences to change in response to shifting average temperatures, as I do in Section 4.3.

4 Interpreting changes in expressed sentiment

By using expressed sentiment as a proxy for underlying preferences for temperature, I am able to mitigate the identification concerns that arise when using hedonic or discrete choice models to estimate the value of climate, as described above. The dataset I construct also allows me to estimate the underlying relationship between temperature and expressed sentiment non-parametrically, and to estimate region- and season-specific responses. These benefits must be weighed against the major drawback of this approach:

the challenge of interpretation. This is a problem also faced the body of literature that uses measures of life-satisfaction as the outcome of interest (Mackerron 2012): how much is one unit of expressed sentiment or reported life satisfaction worth?

The advantage of the hedonic and discrete choice approaches is that the derivation of a dollar value for preferences, is straightforward (Albouy et al. 2016; Sinha, Caulkins, and Cropper 2018). By contrast, backing out estimates of the the value of climate damages from changes in expressed sentiment requires additional assumptions. However, doing so is important for several reasons: first, assigning a monetary value grounds the size of these effects in a metric that is more likely to be consistently interpreted by different readers; second, monetary calibration of the effect of changes in temperature on emotional state allows researchers and policy analysts to compare the size of these estimates to other documented effects of climate change; third, monetary estimates are critical for inclusion in the three Integrated Assessment Models currently used by the United States Government to estimate the social cost of carbon (Rose 2014).

As a first step towards identifying willingness-to-pay for climate amenities, I provide a range of approaches designed to give meaning and context to the magnitude of the results I observe and to guide future work in this area. Section 4.1 describes three validation exercises: these demonstrate how other types of external variation impact the measures of expressed sentiment that I record. Section 4.2 demonstrates two valuation exercises to identify the monetary value of the shifts in sentiment I observe. Section 4.3 combines one of these estimates with predictions of future climate to project the damages from climate changes.

4.1 Validation exercises

This section documents the relationships between three non-temperature sources of variation with expressed sentiment. Section 4.1.1 documents changes in expressed sentiment by day of week, while Sections 4.1.2 and 4.1.3 identify the causal impacts of hurricanes and football game outcomes respectively on expressed sentiment.

4.1.1 Expressed sentiment by day of week

First, I conduct a validation exercise that examines how sentiment changes over the course of the days of the week. First, Figure 5 shows the standardized measures by day of week. The weekly variation in matches prior work (Dodds et al. 2011) and common intuition: weekends and Fridays are preferred to non-Friday weekdays, with the lowest measures of affect occurring on Mondays and the highest on Saturdays. To calibrate to the results shown earlier, note that the average difference in sentiment measure between Sunday and Monday is between 0.1 SD and 0.2 SD across the measures, or roughly the difference between experiencing a day with maximum temperature between 20 and 25 C and a day with maximum temperature between 35 and 40 C in Table 3.

[Figure 5]

4.1.2 Impact of hurricanes on expressed sentiment

The high winds, heavy rains, storm surges, and the distress and uncertainty hurricanes create results in both substantial economic losses and difficult-to-quantify human hardship (Hsiang and Jina 2014). And although hurricanes mostly affect areas on the Eastern seaboard between June and November, their appearance and path of destruction tend to both be unpredictable in the short term. For these reason, estimating the impact of hurricanes on expressed sentiment serves as a useful benchmark for the baseline results in this paper. I collect data on hurricane occurrence from the Atlantic Hurricane Database (Landsea and Franklin 2013) and combine it with CBSA-date averages of expressed sentiment, measured using the VADER. To identify the causal impact of a nearby hurricane on expressed sentiment, I estimate a model similar to Equation (1), but replace the weather variables with an indicator for proximity to a hurricane. Table 5 documents the results of this estimation for all hurricanes (first two columns) and for hurricanes with an average wind speed greater or equal to 40 m/s (second two columns), using two different sets of fixed effects.

[Table 5]

On average, CBSAs experience a daily reduction in expressed sentiment of nearly 0.4 SD from any nearby hurricane, while high speed hurricanes cause a reduction of expressed sentiment of around 0.7 SD. This effect is of the same sign and between 2 and 7 times are large as the effect estimated in Figure 2, indicating that the impact of nearby hurricanes is more pronounced than the impact of extremely hot or cold temperatures.

4.1.3 Impact of football game outcomes on expressed sentiment

I estimate a similar model to **??** using the outcomes of NFL football games during my sample, following in the spirit of Card and Dahl (2011), who shows that unexpected football losses correlate with family violence. I map each CBSA to a nearby football team, where being within either 100 or 25 kilometers of a team's home stadium qualifies as "nearby." The model I estimate also includes an indicator for whether the team was playing on a given day. Table 6 documents the findings, where the first two columns take a nearby team as one within 100 km and the second two take a nearby team as one within 25 km, with two different sets of fixed effects for each.

[Table 6]

As expected, a nearby team suffering a loss causes a negative impact on expressed sentiment. The size of the impact is roughly equal to the negative impact of a very hot day in Figure 2, and is slightly larger when estimated using teams that are within 25 km relative to 100 km.

4.2 Valuing expressed sentiment

This section describes two approaches to convert the estimates of changes in expressed sentiment presented above into monetary value in order to estimate the value of different climates. This approach follows in the spirit of Levinson (2012), who converts reported life satisfaction into a dollar value by dividing the response of life satisfaction to pollution levels by the response of life satisfaction to changes in income.[12] Conceptually, this method can be understood as follow: suppose an individual derives utility from a nonmarket good (such as temperature) which they do not choose and income. To estimate the extent to which these individuals value the nonmarket good, it is sufficient to divide the change in expressed sentiment caused by a one unit increase in income.

I note at the outset that obtaining credible estimates of the value of expressed sentiment is important but requires strong assumptions. First, it must be the case that these changes in income are exogenously determined with respect to expressed sentiment, and second, that expressed sentiment represents an appropriate proxy for utility. With respect to the first assumption, the following two approaches represent attempts to isolate plausibly exogenous variation in incomes. The second assumption is more is difficult to test and the results that follow should be interpreted in light of that consideration.

To identify the relationship between income and expressed sentiment, I first estimate the degree to which quarterly expressed sentiment responds to changes in quarterly wages

12. Allcott, Lockwood, and Taubinsky (2019) conduct a similar exercise in their investigation of the optimal soda tax.

from the Quarterly Census of Employment and Wages (QCEW) and use this response to estimate a per-SD value of $196.77. Appendix A.4.1 documents the construction and estimation procedure in detail. The second approach identifies plausibly exogenous variation in income by focusing on the population of users in my dataset who received parking or speeding tickets and who noted that receipt on Twitter. Using this approach, I find a per-SD value of $78.60. See Appendix A.4.2 for more details on the construction and estimation of this approach.

To estimate the value of these temperature changes, I multiply the per-SD values from the two approaches above with the base estimates of the change in expressed sentiment caused by temperature from Table 3, column (5). Table 7 documents the results. Interpreted literally, these estimates imply, for example, that the average individual in my sample would be willing to pay $11.94 or $4.77 (depending on whether the wage regression or the parking ticket response value is used) to exchange a day between 30 and 35 degrees with a day between 20 and 25 degrees.

[Table 7]

Because I estimate a higher per-SD value using the relationship between sentiment and quarterly wages, all of the estimates in the first column are larger (in absolute value) than those in the second column. Given the uncertainties outlined above in these valuation strategies, I proceed conservatively: the following section, which projects the costs of climate change using expected changes in expressed sentiment, relies on the smaller per-SD value ($78.60) obtained using the parking ticket approach.

4.3 Projecting damages from climate change

This section combines the estimates from Section 3 with projections of future climate and the valuations obtained in Section 4.2 to value the change in expressed sentiment we might expect from climate change. First, I project the annual amenity cost of rising temperatures across the United States on amenity value, measured in the change in SD of sentiment. I multiply these estimates with the per-SD value of changes in expressed sentiment to obtain the dollar value of these changes.

More specifically, the nature of the projection exercise can be described mathematically as follows:

$$\int^{T} f(t)\Delta g(t) v\, dt \tag{3}$$

where $f(\cdot)$ represents the damage function (valued in SD of expressed sentiment), such as the one estimated in Figure 2, $\Delta g(\cdot)$ is the change in the distribution of climate, and v is the value of a single SD change in dollars. By integrating the product of f and g over the range of temperature T and multiplying by v I obtain the total damages. Empirically, I estimate the shape of $f(\cdot)$, combine climate and weather data to obtain $\Delta g(\cdot)$, and draw v from Section 4.2 (the parking ticket-based estimate of the value of sentiment) in order to numerically approximate Equation (10).

With this framework, I conduct two exercises, referred to hereafter as the "baseline" and "adaptive" projection exercises. The baseline exercise projects damages using a single function for f, the estimate obtained by the splined model in Figure A.3. The adaptive exercise follows Auffhammer and Aroonruengsawat (2011) and uses the regionally-specific damage functions estimated in Figure 4 to project changes due to adaptation. For each projection, I assume that CBSAs respond with the regional damage function that corresponds to the annual average temperature it currently experiences. In other words, if climate change are projected an area to the extent that it moves from the "Warm" quartile into the "Warmest" quartile, then its damages are estimated using the "Warmest" quartile's damage function.

$g(\cdot)$ is estimated using the ensemble average from the output of 20 downscaled climate models[13], I compile average projections for each CBSA for the years 2006-2099. In order to de-bias the projections, I follow the prescriptions of Auffhammer et al. (2013) and add the difference between projected monthly decadal averages starting in 2026 and projected monthly averages from 2006-2025, then add those differences to the historical weather data from 2006-2015 to simulate future weather regimes for each decade while retaining historically observed variance in temperature. I estimate the difference in the distributions between baseline climate and the given future climate to obtain $\Delta g(\cdot)$.

Figure 6 documents the evolution of per-person annual damages over time, averaged over CBSAs and presented separately for RCP 4.5 and RCP 8.5, two different climate forcing scenarios of intermediate and high warming, respectively (IPCC 2014) and using both the baseline and adaptive methods described above.

[Figure 6]

I estimate annual damages of increasing over time across all scenarios, with damages

13. Climate forcings drawn from a statistical downscaling of global climate model (GCM) data from the Coupled Model Intercomparison Project 5 (Taylor, Stouffer, and Meehl 2012) using the Multivariate Adaptive Constructed Analogs (MACA; Abatzoglou and Brown 2012) method with the Livneh (Livneh et al. 2013) observational dataset as training data.

from RCP8.5 exceeding damages from RCP 4.5. For the baseline scenario, I estimate per-person annual damages of $297 under RCP 4.5 and $810 under RCP 8.5. Allowing for the possibility of adaptation shifts the estimates downward, to $195 and $313 respectively. In all cases, however, these are notable estimates: the annual median income in the United States was $31,786 in 2017, meaning that the estimates I obtain are between 0.6% and 2.5% of present-day annual income. Less conservative modeling decisions, such as using the per-SD value from the relationship between CBSA and income in Section 4.2, would obtain larger damages. Broadly speaking, the set I give above are in line with other estimates of the amenity cost of climate change. For example, Albouy et al. (2016) estimate end-of-century damages of between 1% and 4% of annual income. Figure A.12 documents the distribution of damages across space, estimating damages by end of century for each CBSA under RCP 8.5. Under the baseline scenario, damages are increasing in average temperate and are highest for CBSAs in the South, while under the adaptive scenario, damages are greatest for both the South and for parts of the upper Midwest.

The difference between the baseline and adaptive approaches is also notable. This estimate is made possible by the richness of the data I use, since I am able to estimate separate response functions for each region in Figure 4. As expected, the ability to adapt one's preferences for temperature as the climate warms mitigates the impact of climate change substantially. Whether this degree of adaptation is likely remains an open question, since these cross-sectional differences in responses to temperature may or may not be driven by factors that are able to change in the long run.

5 Impact of temperature on sentiment around the world

For reasons of data availability, the analysis thus far has focused on the United States. However, while all countries will be affected by climate change, extrapolating preferences for climate estimated earlier in the paper to the result of the world could be inappropriate. In order to provide the first estimates (of which I am aware) of the impact of temperature on sentiment outside of the United States, I collect a similar dataset for six English-speaking countries for which there is adequate temperature variation.[14]

To do so, I again use the Twitter Streaming API, but this time request tweets located inside a bounding box that includes all of Africa, parts of Southeast Asia, and Australia. These data, whose collection began later, span from October 2015 until March 2019. For each first-level administrative unit in these countries (equivalent to a U.S. state, hereafter

14. I define "adequate" as requiring that at least four of the five degree bins between 0 and 40 degrees C constitute more than 1% of the total observations.

referred to as "states" for brevity), I estimate the average sentiment score given by the VADER sentiment measure described above for all tweets in that administrative unit and day. I keep only countries for which I am able to observe at least one million tweets over the course of the sample.

Because the PRISM dataset is not available outside of the United States, I instead use gridded datasets of maximum daily temperature and total daily precipitation from NOAA.[15] I take population weighted-averages of these data for each administrative unit in my sample to obtain the temperature and precipitation experienced by an average individual in that administrative unit on that day. After combining these data, I estimate the following specification:

$$\overline{S}_{sd} = f(T_{sd}) + P_{sd} + \phi_s + \phi_m + \phi_y + \varepsilon_{sd} \tag{4}$$

Let s and d index administrative unit, or state, and date. $\overline{S}_{cd}$ is the state-day average of the VADER measure of sentiment described in Section 2. T_{cd} is the maximum daily temperature in a state, and $f(T_{sd})$ is a flexible function of temperature, which I implement in practice using five degree bins of maximum daily temperature. ϕ_s, ϕ_m, and ϕ_y represent state, month, and year fixed effects, respectively. ε_{sd} is the idiosyncratic error term, clustered by both state and date of sample, and the regression is weighted using the average number of scored tweets in each state.[16] Figure 7 documents the findings.

[Figure 7]

For these countries, I find evidence that moderate temperatures tend to be preferred. Australia, India, and to a limited degree Kenya all exhibit evidence of a preference for moderate temperatures over very hot or very cold temperatures. In South Africa and the Philippines, I find evidence of preferences for moderate temperatures over hot temperatures alone, but no evidence that cold temperature are less preferred. Australia's response seems to most closely resemble that of the United States, while India's response shows a gradual but not sharp distaste for increasingly warm temperatures. Uganda shows no discernible difference in expressed sentiment in response to various temperature, and

15. CPC Global Temperature and Precipitation data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at `https://www.esrl.noaa.gov/psd/`.

16. Because there are substantially fewer tweets in these countries compared to the United States, the specifications I estimate in this section are necessarily simpler than those given in Equation (1).

Kenya's response is fairly limited as well.[17]

I provide these results as evidence of differences in responses for ambient temperature across the world. However, since all of these countries differ from the United States in significant ways, whether these differences also reflect differences in preferences is, to some degree, subject to speculation. Twitter users in other counties may vary in important ways relative to both their own general populations and to the Twitter population in the United States. For example, if only relatively wealthy users participate on Twitter in Uganda, then their exposure to extreme ambient temperatures may be limited due to the use of expensive air conditioning and heating technologies. Because Twitter does not make demographic information available for their users (whether they have it or not), I am not able to directly evaluate the comparability of samples across countries. Nevertheless, the broad strokes of these findings is that there is some congruence, if not total agreement, in the sentiment response to temperature around the world. I view further pursuit of this question as a valuable avenue for future work.

6 Discussion

By using the contemporaneous responses of expressed sentiment on social media to temperature variation as a proxy for underlying preferences for temperature, I provide an alternative to traditional nonmarket valuation techniques. This approach allows me to estimate nonlinear responses of sentiment to temperature and to account for unobserved variation across both space and time and to identify spatial and seasonal differences of preferences for temperature. Finally, the set of validation and valuation exercise I demonstrate provide an interpretive baseline and way forwards for future valuation exercises following this method.

This new approach is not without its drawbacks. The formation of emotional state is undeniably complex: the physical, biological, and psychological bases for human emotions remain only partly understood (Russell 1980), and the distillation of that complexity to a single dimensional affective scale abstracts away from important nuances regarding the formation of emotion, not to mention its relationship between economic definitions of experienced utility. And although I am able to show that the users who choose to geolocate their tweets are not observationally different than the larger set of Twitter users (Appendix A.3.1), I am only able to provide suggestive evidence with respect to the degree to which these findings would generalize to the full population (Appendix A.3.2).

17. Figure A.13 replicates Figure 7 but includes the U.S. response as well for the purpose of direct comparison.

Finally, the estimates of the value of expressed sentiment Section 4.2, while important for improving our understanding of climate impacts, necessarily rely on strong assumptions.

Despite these limitations, this paper makes several contributions to the literature. It introduces a new method and data source to estimate preferences for and valuations of public goods while simultaneously accounting for possible unobservable cross-sectional and seasonal variation. It reveals previously unobservable geographic and seasonal of preferences for temperature and provides suggestive evidence of adaptive capacity in this area. It also demonstrates how NLP can facilitate previously intractable economic analyses and suggests a psychological channel through which other impacts of climate change may operate. Finally, it provides the first estimates of similar impacts in countries other than the United States. Broadly, this work provides additional evidence that changes in the amenity value of climate are an important component of the cost of climate change.

References

Abatzoglou, John T., and Timothy J. Brown. 2012. "A comparison of statistical downscaling methods suited for wildfire applications." *International Journal of Climatology* 32 (5): 772–780.

Albouy, David, Walter Graf, Ryan Kellogg, and Hendrik Wolff. 2016. "Climate Amenities, Climate Change, and American Quality of Life." *Journal of the Association of Environmental and Resource Economists* 3 (1): 205–246.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax." *The Quarterly Journal of Economics* 134, no. 3 (May): 1557–1626.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.*

Auffhammer, Maximilian. 2013. "Quantifying intensive and extensive margin adaptation responses to climate change: A study of California's residential electricity consumption." *Working Paper.*

Auffhammer, Maximilian, and Anin Aroonruengsawat. 2011. "Simulating the impacts of climate change, prices and population on California's residential electricity consumption." *Climatic Change* 109:191–210.

Auffhammer, Maximilian, Solomon Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change." *Review of Environmental Economics and Policy* 7 (2): 181–198.

Baylis, Patrick, Nick Obradovich, Yury Kryvasheyeu, Haohui Chen, Lorenzo Coviello, Esteban Moro, Manuel Cebrian, and James H Fowler. 2018. "Weather impacts expressed sentiment." *PLoS ONE* 13, no. 4 (April): e0195750.

Bentham, Jeremy. 1789. "An Introduction to the Principles of Morals and Legislation." In *The Collected Works of Jeremy Bentham: An Introduction to the Principles of Morals and Legislation.* Oxford University Press, January.

Burke, Marshall, and Kyle Emerick. 2015. "Adaptation to climate change: Evidence from US agriculture." *American Economic Journal: Economic Policy.*

Burke, Marshall, Solomon Hsiang, and Edward Miguel. 2015. "Climate and Conflict." *Annual Review of Economics* 7:577–617.

Card, David, and Gordon B. Dahl. 2011. "Family violence and football: The effect of unexpected emotional cues on violent behavior." *Quarterly Journal of Economics* 126 (1): 103–143.

Carleton, Tamma, and Solomon Hsiang. 2016. "Social and economic impacts of climate." *Science* 353 (6304).

Cline, William R. 1992. *The Economics of Global Warming.* Peterson Institute for International Economics.

Daly, Christopher, Wayne P. Gibson, George H. Taylor, Gregory L. Johnson, and Phillip Pasteris. 2002. "A knowledge-based approach to the statistical mapping of climate." *Climate Research* 22 (2): 99–113.

Dell, Melissa, Benjamin F Jones, and Benjamin A Olken. 2014. "What do we learn from the weather? The new climate-economy literature." *Journal of Economic Literature* 52 (3): 740–98.

Dennisenn, J., Ligaya Butalid, Lars Penke, and Marcel A.G. Van Aken. 2008. "The effects of weather on daily mood: A multilevel approach." *Emotion* 8 (5): 662–667.

Deryugina, Tatyana, and Solomon Hsiang. 2014. "Does the Environment Still Matter? Daily Temperature and Income in the United States." *NBER Working Paper.*

Deschênes, Olivier, and Michael Greenstone. 2011. "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US." *American Economic Journal: Applied Economics* 3 (4): 152–185.

Diaz, Delavane B. 2014. "Evaluating the Key Drivers of the US Government's Social Cost of Carbon: A Model Diagnostic and Inter-Comparison Study of Climate Impacts in DICE, FUND, and PAGE." *Working Paper.*

Diener, Ed. 2000. "Subjective Well-Being." *American Psychologist* 55 (1): 34–43.

Dodds, Peter Sheridan, and Christopher M. Danforth. 2010. "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents." *Journal of Happiness Studies* 11 (4): 441–456.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. 2011. "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter." *PLoS ONE* 6 (12): e26752.

Dolan, Paul, Tessa Peasgood, and Mathew White. 2008. "Do we really know what makes us happy? A review of the economic literaure on the factors associated with subjective well-being." *Journal of Economic Psychology* 29 (1): 94–122.

Easterlin, Richard A. 2001. "Income and Happiness: Towards a Unified Theory." *The Economic Journal* 111 (473): 465–484.

Feddersen, John, Robert Metcalfe, and Mark Wooden. 2012. "Subjective Well-Being: Weather Matters; Climate Doesn't." *SSRN Electronic Journal*, no. 627.

Gilbert, CJ, and Eric Hutto. 2014. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." In *Eighth International Conference on Weblogs and Social Media.*

Graff Zivin, Joshua, Solomon Hsiang, and Matthew Neidell. 2018. "Temperature and human capital in the short and long run." *Journal of the Association of Environmental and Resource Economists* 5 (1): 77–105.

Hannak, Aniko, Eric Anderson, Lisa Feldman Barrett, Sune Lehmann, Alan Mislove, and Mirek Riedewald. 2012. "Tweetin'in the rain: Exploring societal-scale effects of weather on mood." In *Sixth International AAAI Conference on Weblogs and Social Media.*

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161.

Houser, Trevor, Robert Kopp, Solomon M. Hsiang, Michael Delgado, Amir Jina, Kate Larsen, Michael Mastrandrea, Shashank Mohan, Robert Muir-Wood, and D. J. Rasmussen. 2014. *American Climate Prospectus: Economic Risks in the United States.*

Hsiang, Solomon. 2010. "Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America." *Proceedings of the National Academy of Sciences* 107 (35): 15367–15372.

———. 2016. "Climate Econometrics." *Annual Review of Resource Economics.*

Hsiang, Solomon, and Amir Jina. 2014. "The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth." *NBER Working Paper.*

Interagency Working Group on Social Cost of Carbon. 2013. *Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866.* Technical report. United States Government.

IPCC. 2014. *IPCC Fifth Assessment Report.* Technical report. Cambridge, United Kingdom and New York, NY, USA: IPCC.

Kahneman, Daniel, and Alan B. Krueger. 2006. "Developments in the Measurement of Subjective Well-Being." *Journal of Economic Perspectives* 20 (1): 3–24.

Kahneman, Daniel, and Robert Sugden. 2005. "Experienced utility as a standard of policy evaluation." *Environmental and Resource Economics* 32 (1): 161–181.

Kenrick, D. T., and S. W. MacFarlane. 1986. "Ambient Temperature and Horn Honking: A Field Study of the Heat/Aggression Relationship." *Environment and Behavior* 18 (2): 179–191.

Landsea, Christopher W., and James L. Franklin. 2013. "Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format." *Monthly Weather Review* 141, no. 10 (October): 3576–3592.

Levinson, Arik. 2012. "Valuing public goods using happiness data: The case of air quality." *Journal of Public Economics* 96, no. 910 (October): 869–880.

Livneh, Ben, Eric A. Rosenberg, Chiyu Lin, Bart Nijssen, Vimal Mishra, Kostas M. Andreadis, Edwin P. Maurer, and Dennis P. Lettenmaier. 2013. "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and extensions." *Journal of Climate* 26 (23): 9384–9392.

Lobell, David B., Marshall Burke, Claudia Tebaldi, Michael D. Mastrandrea, Walter P Falcon, and Rosamond L Naylor. 2008. "Prioritizing climate change adaptation needs for food security in 2030." *Science* 319 (5863): 607–610.

Mackerron, George. 2012. "Happiness Economics from 35000 Feet." *Journal of Economic Surveys* 26 (4): 705–735.

Marshall, Alfred. 1890. *Principles of Economics.* 1–323.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. 2014. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal* 5, no. 4 (December): 1093–1113.

Morstatter, Fred, Jürgen Pfeffer, Huan Liu, and Kathleen M Carley. 2013. "Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose." In *Seventh international AAAI conference on weblogs and social media.*

Nielsen, Finn Årup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." *arXiv.*

Nordhaus, William D. 1991. "To Slow or Not To Slow: The Economics of the Greenhouse Effect." *The Economic Journal:* 920–937.

Nordhaus, William D, and J.G. Boyer. 2000. *Warming the World: Economic Models of Global Warming.* 232. MIT Press.

Pak, Alexander, and Patrick Paroubek. 2010. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." In *LREC,* 10:1320–1326.

Pearce, David. 2002. "An Intellectual History of Environmental Economics." *Annual Review of Energy and the Environment* 27 (1): 57–81.

Pennebaker, James W, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015.* Technical report.

Ranson, Matthew. 2014. "Crime, weather, and climate change." *Journal of Environmental Economics and Management* 67 (3): 274–302.

Rehdanz, Katrin, and David Maddison. 2005. "Climate and happiness." *Ecological Economics* 52 (1): 111–125.

Robinson, W S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15 (3).

Rose, S. 2014. *The Social Cost of Carbon: A Technical Assessment.* Technical report. Electric Power Research Institute.

Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology.*

Schlenker, Wolfram, and Michael J. Roberts. 2009. "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." *Proceedings of the National Academy of Sciences* 106 (37): 15594–15598.

Sinha, Paramita, Martha L Caulkins, and Maureen L Cropper. 2018. "Household location decisions and the value of climate amenities." *Journal of Environmental Economics and Management* 92:608–637.

Stern, Nicholas. 2006. *The Economics of Climate Change,* 662.

Taylor, Karl E., Ronald J. Stouffer, and Gerald A. Meehl. 2012. *An overview of CMIP5 and the experiment design.*

United States Government Accountability Office. 2014. *Development of Social Cost of Carbon Estimates.* Technical report July. United States Government Accountability Office.

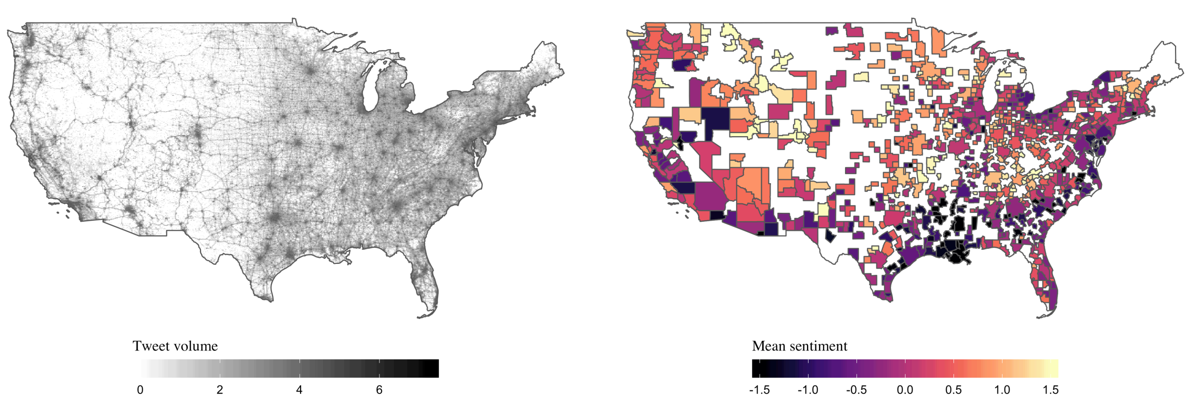
Welsch, Heinz, and Jan Kühling. 2009. "Using happiness data for environmental valuation: Issues and applications." *Journal of Economic Surveys* 23, no. 2 (April): 385–406.

Table 1: Sample characteristics

	Count	Mean	Median	Min	Max
A: Sentiment measures					
AFINN-111	598,750,185	0.5	0.4	-5	4
Hedonometer	1,061,098,510	5.5	5.5	2.6	8.3
LIWC	1,092,491,096	0.3	0.2	-5	5
Vader	1,160,617,577	0.1	0.1	-1	1
B: Weather covariates					
Minimum temperature (C)	1,160,617,577	9.4	11.3	-34.7	31.9
Maximum temperature (C)	1,160,617,577	21.2	24.2	-23.5	47.4
Precipitation (mm)	1,160,617,577	2.9	0	0	434.2
C: Twitter updates per...					
CBSA	908	1,395,932	205,419	15,224	78,506,336
User	11,659,619	106	11	1	17,957

Notes: First panel summarizes unstandardized measures of expressed sentiment: AFINN-111, Hedonometer, LIWC, and VADER. Second panel summarizes weather covariates obtained from PRISM. Third panel summarizes the number of tweets per CBSA and user.

Figure 1: Tweet density and average sentiment by CBSA



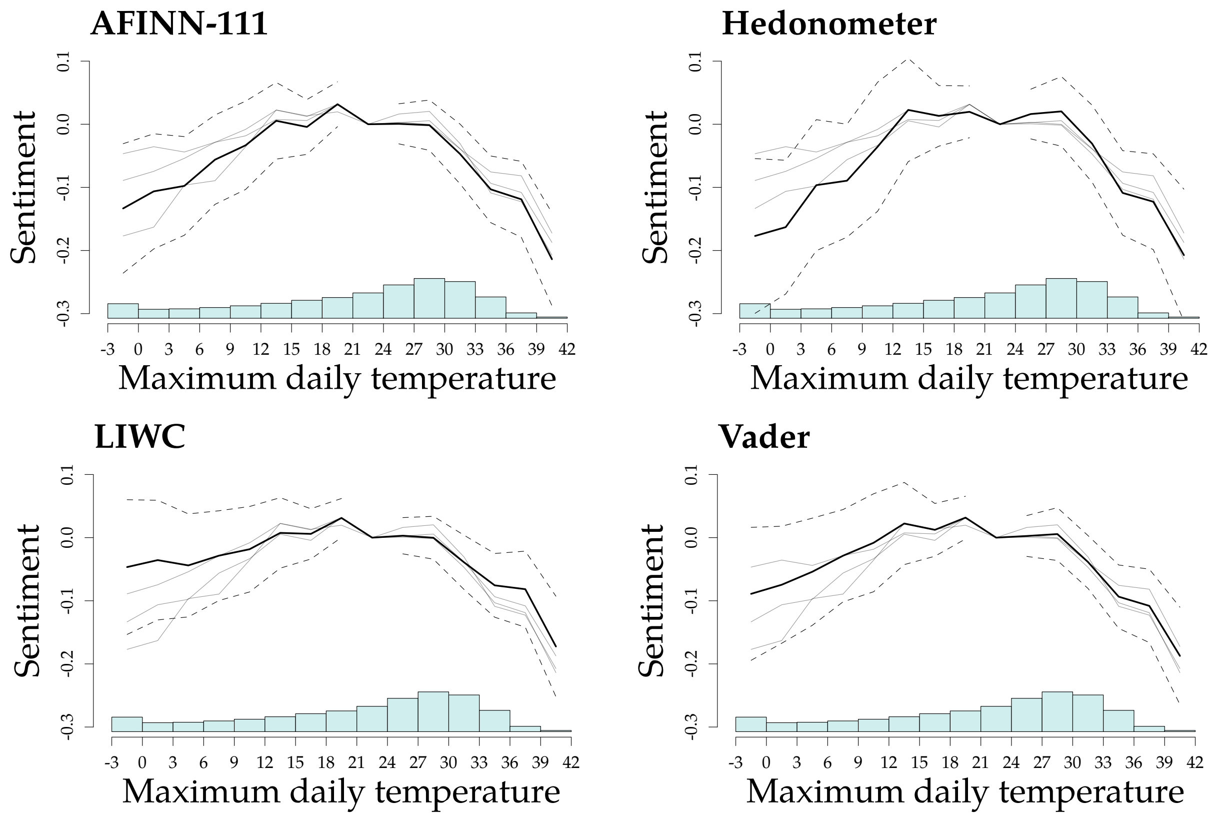
Notes: Left panel: pixel shading represents log (base 10) of count of tweets in sample. Right panel: Mean standardized VADER score by CBSA for CBSAs with more than 100 tweets in sample.

Table 2: Measure correlations: CBSA-date means

	AFINN-111	Hedonometer	LIWC	Vader	Profanity
AFINN-111	1.00	0.65	0.73	0.77	-0.56
Hedonometer		1.00	0.59	0.72	-0.34
LIWC			1.00	0.76	-0.38
Vader				1.00	-0.39
Profanity					1.00

Notes: Pairwise correlations of CBSA-date means of measures of standardized expressed sentiment and profanity measure.

Figure 2: Effect of temperature on expressed sentiment by measure



Notes: Panels document the temperature response for each of the four standardized measures of sentiment. Solid lines show the regression coefficients on temperature and represent the difference (measured in standard deviations) in CBSA-day sentiment for the temperature bin T_b relative to 21-24 C, controlling for state time trends and fixed effects for CBSA, day of week, holiday, month, and year fixed effects. Dotted lines show 95% confidence intervals estimated using cluster robust standard errors by CBSA and date.

Table 3: Effect of temperature on expressed sentiment by specification

	(1)	(2)	(3)	(4)	(5)
Maximum daily temperature T					
$T \leq 5$	−0.15	−0.16	−0.05	−0.19	−0.09
	(0.06)	(0.05)	(0.07)	(0.05)	(0.05)
$T \in (5, 10]$	−0.09	−0.07	−0.005	−0.11	−0.03
	(0.05)	(0.04)	(0.05)	(0.04)	(0.04)
$T \in (10, 15]$	−0.03	−0.02	0.02	−0.03	0.002
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
$T \in (15, 20]$	−0.01	0.004	0.02	−0.003	0.01
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
$T \in (25, 30]$	−0.001	0.01	−0.01	0.01	−0.003
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$T \in (30, 35]$	−0.04	−0.03	−0.04	−0.03	−0.06
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
$T \in (35, 40]$	−0.11	−0.10	−0.09	−0.09	−0.12
	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)
$T > 40$	−0.23	−0.20	−0.11	−0.18	−0.21
	(0.04)	(0.05)	(0.06)	(0.04)	(0.04)
Other controls					
CBSA	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes			Yes
Year	Yes	Yes	Yes		Yes
DOW, Hol		Yes	Yes	Yes	Yes
S×M			Yes		
MOS				Yes	
State trends					Yes
N (millions)	0.7	0.7	0.7	0.7	0.7
Tweets (millions)	1160.6	1160.6	1160.6	1160.6	1160.6

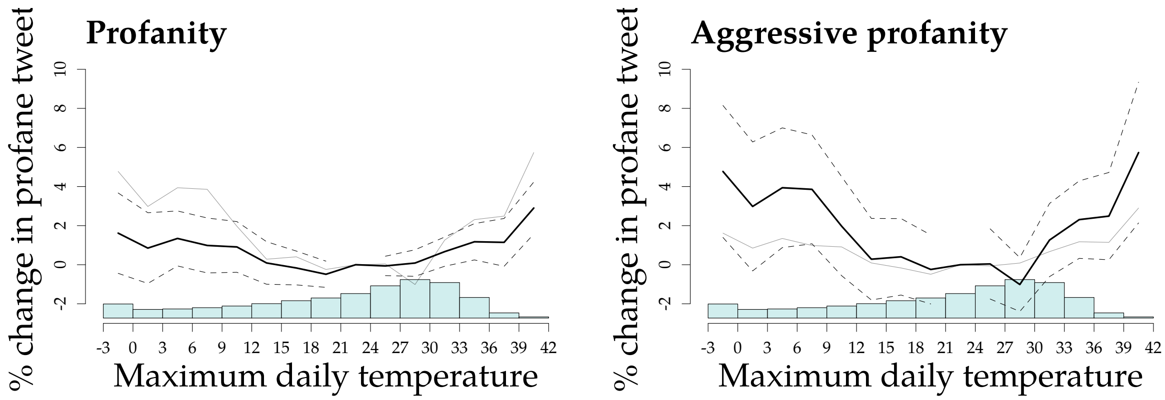
Notes: Table shows estimates of weather on expressed sentiment using different statistical specifications. Coefficients (in cells). Expressed sentiment measured using the VADER sentiment analysis algorithm. Statistical specifications vary according to choice of fixed effects and other controls, as given in the "Other controls" panel. All specifications include daily precipitation. Coefficients represent the difference (in standard deviations) of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. N (in millions) is the number of observations (CBSA-dates). Tweets (in millions) is the number of posts represented by each regression. Standard errors (in parentheses) clustered by CBSA and date.

Table 4: Impacts of temperature on expressed sentiment by unit of observation

Unit	CBSA-date averages			Posts
	Baseline	+Spec.	+Sample	Individual
Maximum daily temperature T				
$T \leq 5$	−0.09	−0.15	-0.08	−0.08
	(0.05)	(0.06)	(0.03)	(0.02)
$T \in (5, 10]$	−0.03	−0.09	-0.06	−0.04
	(0.04)	(0.05)	(0.02)	(0.01)
$T \in (10, 15]$	0.002	−0.03	-0.02	0.01
	(0.03)	(0.04)	(0.02)	(0.01)
$T \in (15, 20]$	0.01	−0.01	-0.01	0.01
	(0.01)	(0.02)	(0.01)	(0.01)
$T \in (25, 30]$	−0.003	−0.001	-0.01	−0.001
	(0.02)	(0.02)	(0.01)	(0.01)
$T \in (30, 35]$	−0.06	−0.04	-0.01	−0.02
	(0.02)	(0.03)	(0.02)	(0.01)
$T \in (35, 40]$	−0.12	−0.11	-0.07	−0.04
	(0.03)	(0.03)	(0.02)	(0.01)
$T > 40$	−0.21	−0.23	-0.15	−0.10
	(0.04)	(0.04)	(0.03)	(0.02)
Other controls				
CBSA FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
DOW, Hol FE	Yes			
State trends	Yes			
User FE				Yes
N (millions)	0.7	0.7	0.7	432.6
Tweets (millions)	1160.6	1160.6	432.6	432.6

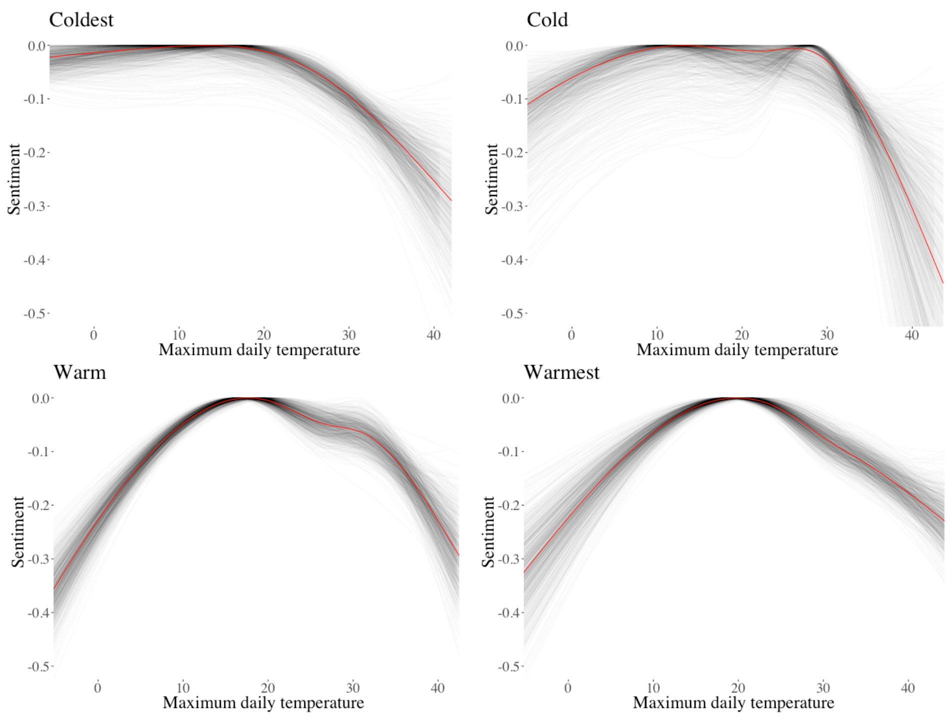
Notes: Table shows estimates of weather on expressed sentiment using different units of observation. Coefficients represent the difference (in standard deviations) of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. First three columns use CBSA-date averages as the unit of observation, the final column uses individual posts. "Individual" is estimates on tweets from users who posted on at least 25% of the days in the sample and averaged no more than 25 posts per day. "Baseline" column reproduces column (5) from Table 3. "+Spec." column matches the "Individual" column specification. "+Sample" additionally matches the "Individual" column sample. All specifications include daily precipitation. N (in millions) is the number of observations (CBSA-dates or posts). Tweets (in millions) is the number of posts represented by each regression. Standard errors (in parentheses) clustered by CBSA and date, except in the case of the "Individual" column, where they are clustered by CBSA.

Figure 3: Effect of temperature on profanity



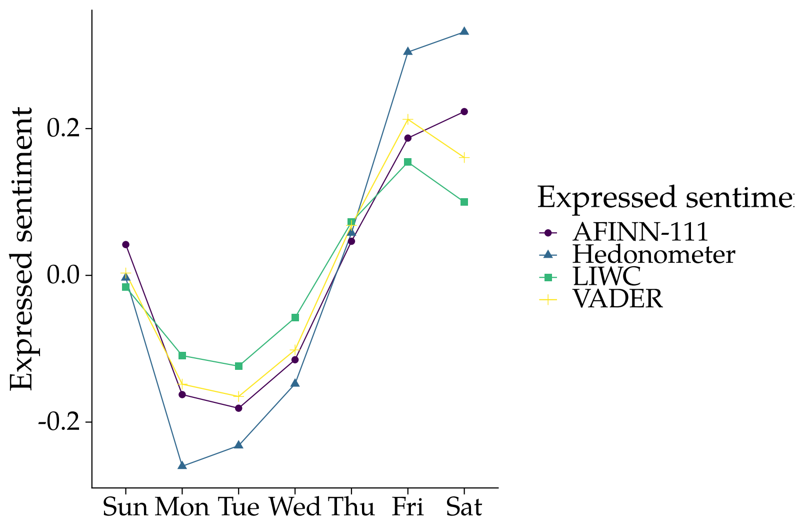
Notes: Figure documents occurrence of profanity in response to temperature, using a list of 311 profanities. Point estimates represent the difference (measured in standard deviations) in CBSA-day profanity occurrence for the temperature bin T_b relative to 21-24 C, conditional on CBSA and state by month of sample fixed effects. 95% confidence intervals estimated using two-way cluster robust standard errors on CBSA and date.

Figure 4: Sentiment responses to temperature differ by region



Notes: Panels document the response of the expressed sentiment to temperature for each of the four regions, where regions are defined by quartiles of average daily maximum temperature during the sampling frame. Dark red line is the splined response function estimated from the entire dataset, with knots at the 25th, 50th, and 75th percentile of experienced daily maximum temperatures. Lighter gray lines are splines estimated similarly using bootstrapped samples. All lines are normalized s.t. the highest point of the spline has y = 0. Regressions include CBSA, month, and year fixed effects. 95% confidence intervals estimated using cluster robust standard errors by CBSA and date.

Figure 5: Expressed sentiment by day of week



Notes: Lines represent average standardized measures of expressed sentiment by day of week. Standardization is conducted using the weighted mean and variance of the CBSA-date averages.

Table 5: Impact of hurricanes on expressed sentiment

	All	All	40+	40+
Hurricane in area	−0.41	−0.42	−0.65	−0.63
	(0.15)	(0.19)	(0.29)	(0.20)
Fixed effects				
CBSA	Yes	Yes	Yes	Yes
DOW	Yes	Yes	Yes	Yes
Month	Yes		Yes	
Year	Yes		Yes	
MOS		Yes		Yes

Notes: Table documents impact of nearby hurricanes expressed sentiment for nearby CBSAs (metro areas). Outcome variable is standardized CBSA daily average of expressed sentiment. All regressions control for day of hurricane along with listed fixed effects. First and second columns include all hurricanes, third and fourth columns only includes those with a wind speed of 40 m/s. Standard errors clustered by CBSA and date.

Table 6: Impact of football outcomes on expressed sentiment

	100 km	100 km	25 km	25 km
Nearby team won	0.10	0.10	0.08	0.08
	(0.03)	(0.03)	(0.05)	(0.05)
Fixed effects				
CBSA	Yes	Yes	Yes	Yes
DOW	Yes	Yes	Yes	Yes
Month	Yes		Yes	
Year	Yes		Yes	
MOS		Yes		Yes

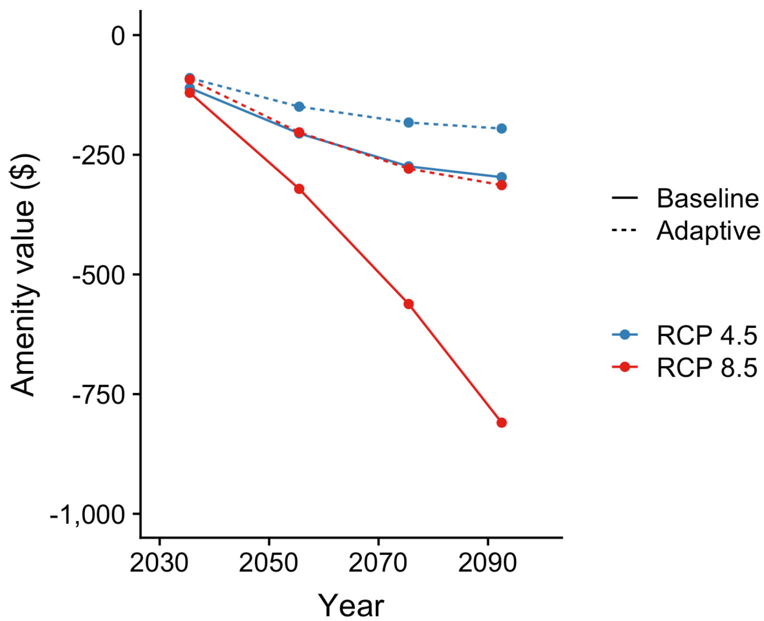
Notes: Table documents impact of football outcomes on expressed sentiment for nearby CBSAs (metro areas). Outcome variable is standardized CBSA daily average of expressed sentiment. All regressions control for day of game along with listed fixed effects. First and second columns include CBSAs within 100 km of a football team's stadium, third and fourth columns include CBSAs within 25 km. Standard errors clustered by CBSA and date.

Table 7: Value of temperature

	Value from wage regression	Value from parking ticket response
Maximum daily temperature T		
$T \leq 5$	−17.72	−7.08
	(8.90)	(3.55)
$T \in (5, 10]$	−6.73	−2.69
	(7.04)	(2.81)
$T \in (10, 15]$	0.46	0.18
	(6.42)	(2.56)
$T \in (15, 20]$	2.27	0.91
	(2.92)	(1.17)
$T \in (25, 30]$	−0.62	−0.25
	(3.15)	(1.26)
$T \in (30, 35]$	−11.94	−4.77
	(4.03)	(1.61)
$T \in (35, 40]$	−23.61	−9.43
	(5.69)	(2.27)
$T > 40$	−42.02	−16.78
	(8.53)	(3.41)

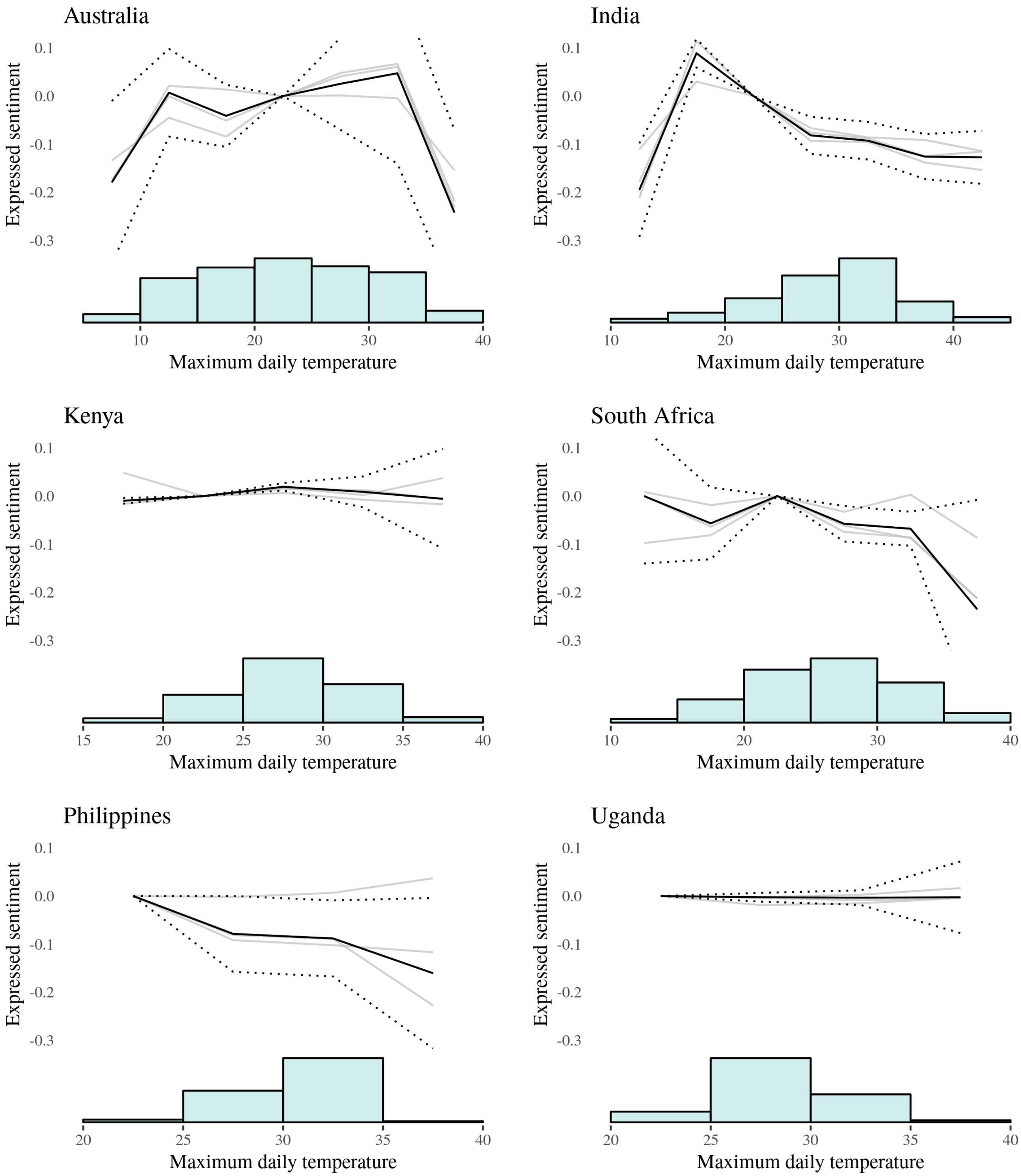
Notes: Reproduces column (5) in Table 3 with coefficients multiplied by valuation of a unit SD change in standardized VADER sentiment. Value of an SD change in the first column uses regression of expressed sentiment on quarterly CBSA wages, as described in Appendix A.4.1. Value of an SD change in the second column from dividing the median observed ticket cost ($100) by the total cumulative sentiment loss from a parking or speeding ticket (1.27 SD) in Figure A.11, as described in Appendix A.4.2.

Figure 6: Projections of changes in amenity value over time



Notes: Projections of average change in amenity value over time, measured in annual per-person change in SD of expressed sentiment. Line color indicates the warming scenario, or Representative Concentration Pathway, used in the the projection data. Line type indicates whether the projection method was the baseline or adaptive method, as described in Section 4.3.

Figure 7: Effect of temperature on expressed sentiment around the world (U.S. included)



Notes: Panels document temperature response for six countries. Solid lines show the regression coefficients on temperature and represent the difference (measured in standard deviations) in CBSA-day sentiment for the temperature bin T_b relative to 20-25 C, controlling for state, month, and year fixed effects. Light and solid lines represent estimates use alternative specifications of fixed effects including: state plus month of sample, state by month of year plus year, and state plus date. Dotted lines show 95% confidence intervals, clustered by state.

A Appendix

Contents

A.1 Construction of expressed sentiment from Twitter data

A.1.1 Accessing Twitter's Streaming API

In this section I describe the process by which I obtain the Twitter data used in this paper. I access these data through the Twitter Streaming API, which provides a real-time stream of public Twitter posts to users upon requests.[18]

First, I registered for an API key through Twitter's Developer program[19] to obtain a set of keys (API Key, API secret key, Access token, Access token secrete) that allow me to access the data. Users are required to agree to Twitter's conditions[20] in order to use these data, and are advised to undertake a careful reading.

Second, I wrote a Python script using the Python package "tweepy".[21] This script requests all Twitter data geolocated within the area given by Figure A.1. Generally speaking, there is a limit on the quantity of data available through the streaming API:

18. Twitter's license agreement for this API forbids sharing the raw Twitter data publicly.
19. Accessible here: https://developer.twitter.com/en/docs.
20. Accessible here: https://developer.twitter.com/en/developer-terms.
21. A sample version of the script to access the API, along with instructions, can be obtained here: https://tinyurl.com/txqstrb. Users will need to create their own API key.

Figure A.1: Tweet collection map

Notes: Map of bounding box used to collect posts from Twitter Streaming API given in red. State boundaries in black for reference.

at most, around 1% of the total number of tweets produced can be downloaded. Because I impose this geographic restriction, however, the script is rarely subject to this limitation, and according to the output from the API I am able to collect more than 99% of the geolocated posts requested.

To conserve hard drive space, I compress and preserve these posts in real time, retaining the date and time posted, the user who posted it, their location, and the tweet itself. This process involves converting the raw output data, which are given in JSON (JavaScript Object Notation) format, to a flattened CSV (comma-separated value) file, which I subsequently compress. This step is optional: users with sufficient hard drive space may simply store the raw JSONs, which may also be compressed to some degree. The following sections describe how each measure of expressed sentiment is constructed from the post text.

A.1.2 Measuring expressed sentiment

The AFINN measure is constructed using an expert-created dictionary that maps words to measures of emotional state. The AFINN-111 dictionary contains 2,477 words scored using integers between -5 and 5, where -5 indicates negative emotional state and 5 indicates positive emotional state. The dictionary focuses on words that are indicative of emotional state, and was created by Nielsen (2011) to analyze language typically used in microblogging. The dictionary is refined from an earlier dictionary built by psychologists to assess the sentiment of written texts (Bradley and Lang 1999).

The Hedonometer measure is constructed in a similar manner to the AFINN measure, but instead uses a dictionary constructed by Dodds and Danforth (2010). The authors crowd-source a dictionary of more than 10,000 words using Amazon's Mechanical Turk service, which outsources tasks to users who are paid for their time. Users were asked to rate each word on a scale from 1 to 9, where 1 indicated negative emotional state and

9 indicated positive emotional state, and measures were averaged across users to get a single measure for each word. Unlike the AFINN measure, the Hedonometer measure scores most commonly-used words regardless of whether they are likely to be indicative of underlying emotional state.

The LIWC measure uses the Linguistic Inquiry and Word Count (LIWC) dictionary created by Pennebaker et al. (2015). Like AFINN and Hedonometer, LIWC uses a dictionary-based method to score text. LIWC contains a variety of dictionaries developed using human categorizations of words: I focus on the lists of words that indicate positive and negative emotion, respectively. The strength of LIWC is that the word lists relating to positive and negative emotion have been independently validated by outside researchers. For example, Kahn et al. (2007) conduct a set of experiments that test whether individuals' stated emotional states correspond to the emotional state estimated from their writing samples using LIWC, and find that LIWC is a valid measure of measuring emotional state.

The VADER measure is a "a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains" (Gilbert and Hutto 2014). VADER is licensed as open-source and is a normalized, weighted composite score. The lexicon used by VADER is constructed by aggregating ratings from 10 independent human raters. The list of candidate words for the lexicon is constructed from previously existing measures of sentiment and augmented using lexical features frequently observed in online contexts such as emoticons (e.g, ":)"), acronyms (e.g., "LOL"), and slang (e.g., "nah"). The VADER measure also includes a mechanism that incorporates information about the word order and intensifiers included in the sentence, so that "very good" is measured as having a higher valence than "good". The measure has been validated against a variety of ground-truth data and found to outperform other measures (Gilbert and Hutto 2014). Figure A.2 plots expressed sentiment, measured using VADER, by state.

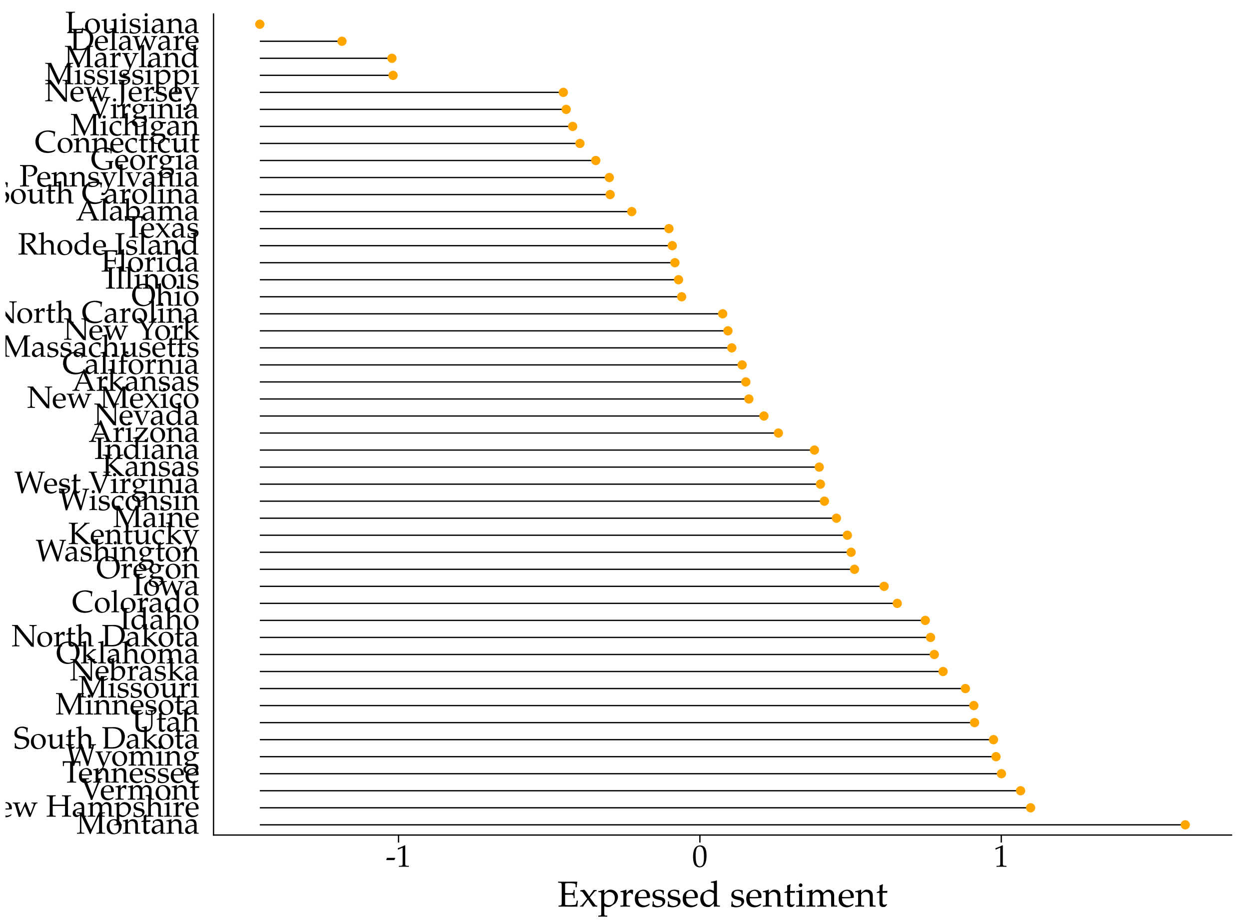
Tables A.1 to A.3 include examples from the word lists used to construct the AFINN, Hedonometer, and LIWC sentiment scores. Table A.4 gives the set of weather words I use to exclude weather-related tweets. I do not provide a list here of the 311 profanities used to construct the list of words used, but these are available upon request.

A.1.3 Construction of crime and mortality word lists

In order to systematically generate a word list relevant to those two topics, I obtain the pre-trained Stanford GloVe vectors, a set of vector representations trained on a set of two billion tweets (Pennington, Socher, and Manning 2014). For each of the crime words given in Ranson (2014) ("assault", "burglary", "larceny", "murder", "manslaughter", "rape", "robbery", and "theft"), I obtain the ten most similar words as measured by cosine similarity. The unique set of 68 words yielded by this exercise is given in Table A.5 is the word list I use to detect crime-related tweets. Beginning with the words "death", "sick", and "hospital", I again use the pre-trained GloVe vector to obtain the ten most similar words for each. This unique set of 40 words forms my mortality-related word list, given in Table A.6.[22]

22. I remove "of" from this word list.

Figure A.2: Expressed sentiment by state



Notes: By-state weighted averages CBSA-date measures of standardized expressed sentiment, computed using VADER. Vertical axis is ordered from lowest to highest, where lowest indicates most negative sentiment.

Table A.1: AFINN word-score examples

Positive Affect		Neutral Affect		Negative Affect	
superb	5	combat	-1	betraying	-3
thrilled	5	apologizes	-1	agonises	-3
hurrah	5	exposing	-1	destroying	-3
outstanding	5	oxymoron	-1	swindle	-3
breathtaking	5	provoked	-1	abhors	-3
roflcopter	4	limited	-1	humiliation	-3
wowow	4	escape	-1	chastises	-3
rejoicing	4	unconfirmed	-1	victimizing	-3
lifesaver	4	passively	-1	bribe	-3
winner	4	blocks	-1	lunatic	-3
miracle	4	poverty	-1	scandal	-3
triumph	4	attacked	-1	outrage	-3
fabulous	4	gun	1	betrayed	-3
roflmao	4	feeling	1	terror	-3
euphoric	4	intrigues	1	abuse	-3
heavenly	4	alive	1	greenwash	-3
fantastic	4	protected	1	falsified	-3
ecstatic	4	unified	1	douche	-3
funnier	4	relieves	1	agonized	-3
winning	4	fit	1	criminals	-3
masterpiece	4	restore	1	defects	-3
masterpieces	4	relieve	1	idiotic	-3
stunning	4	greeting	1	woeful	-3
godsend	4	yeah	1	acrimonious	-3
lmfao	4	cool	1	nuts	-3
lmao	4	vested	1	swindles	-3
rotflmfao	4	clearly	1	lost	-3

Notes: Raw scores shown. Standardized scores used in analysis. Full list includes 2,477 total word-score mappings and can be obtained here: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010

Table A.2: Hedonometer word-score examples

Positive Affect		Neutral Affect		Negative Affect	
laughter	8.5	fui	5.08	suicide	1.3
happiness	8.44	gilbert	5.08	terrorist	1.3
love	8.42	hart	5.08	rape	1.44
happy	8.3	hij	5.08	murder	1.48
laughed	8.26	hun	5.08	terrorism	1.48
laugh	8.22	indonesia	5.08	cancer	1.54
laughing	8.2	jo	5.08	death	1.54
excellent	8.18	john	5.08	died	1.56
laughs	8.18	juan	5.08	kill	1.56
joy	8.16	knee	5.08	killed	1.56
successful	8.16	laws	5.08	torture	1.58
win	8.12	listed	5.08	arrested	1.64
rainbow	8.1	manhasset	5.08	deaths	1.64
smile	8.1	marion	5.08	raped	1.64
won	8.1	martinez	5.08	killing	1.7
pleasure	8.08	medicaid	5.08	die	1.74
smiled	8.08	medicine	5.08	jail	1.76
rainbows	8.06	meyer	5.08	terror	1.76
winning	8.04	might	5.08	kills	1.78
celebration	8.02	morgen	5.08	fatal	1.8
enjoyed	8.02	morris	5.08	killings	1.8
healthy	8.02	nas	5.08	murdered	1.8
music	8.02	necessarily	5.08	war	1.8

Notes: Raw scores shown. Standardized scores used in analysis. Full list includes 10,223 total word-score mappings and can be obtained here: http://hedonometer.org/words.html.

Table A.3: LIWC word examples

Positive emotion	Negative emotion
love	hurt
nice	ugly
sweet	nasty

Notes: LIWC is a commercial product, selected examples are described in Tausczik and Pennebaker (2010). Full list includes 905 total words.

Table A.4: Weather-related stopwords

blizzard	frostbite	precipitation
breeze	frosty	rain
chilly	gail	rainbow
clear	gust	showers
clouds	hail	sleet
cloudy	heat	snowflakes
cold	hot	soggy
damp	humid	sprinkle
dew	hurricane	sunny
downpour	icy	thunder
drizzle	lightning	thunderstorm
drought	misty	typhoon
dry	moist	weather
flurry	monsoon	wet
fog	muddy	wind
freezing	overcast	windstorm
frigid	pouring	windy

Notes: Author construction.

Table A.5: Crime-related words

weapons	burglar	misdemeanor	abuse	fucking
charges	burglars	acquitted	victim	getting
rape	thefts	convicted	assault	gonna
alleged	suspected	conviction	raped	kinda
murder	carjacking	indicted	raping	fraud
charged	embezzlement	pleads	harassment	thief
investigation	manslaughter	sentenced	violence	identity
weapon	cheapness	death	rapes	gta
rifle	perjury	accused	tired	andreas
arrested	dismemberment	killing	feel	hacking
robbery	trickery	murders	really	gtav
break-in	pillage	murdering	ugh	stolen
burglaries	instigating	suspect	bad	
arson	marnier	kill	seriously	

Notes: Table lists crime-related words, generated using similarity indices from Pennington, Socher, and Manning (2014). Details of word selection construction given in Appendix A.1.3.

Table A.6: Mortality-related words

dead	killed	hotel	fucking	shit
after	hosp	icu	getting	being
murder	doctor	tired	gonna	damn
died	medical	feel	kinda	crap
killing	doctors	really	hate	sucks
dies	médico	ugh	stupid	sad
child	nurse	bad	again	cold
kill	center	seriously	awful	pissed

Notes: Table lists mortality-related words, generated using similarity indices from Pennington, Socher, and Manning (2014). Details of word selection given in Appendix A.1.3.

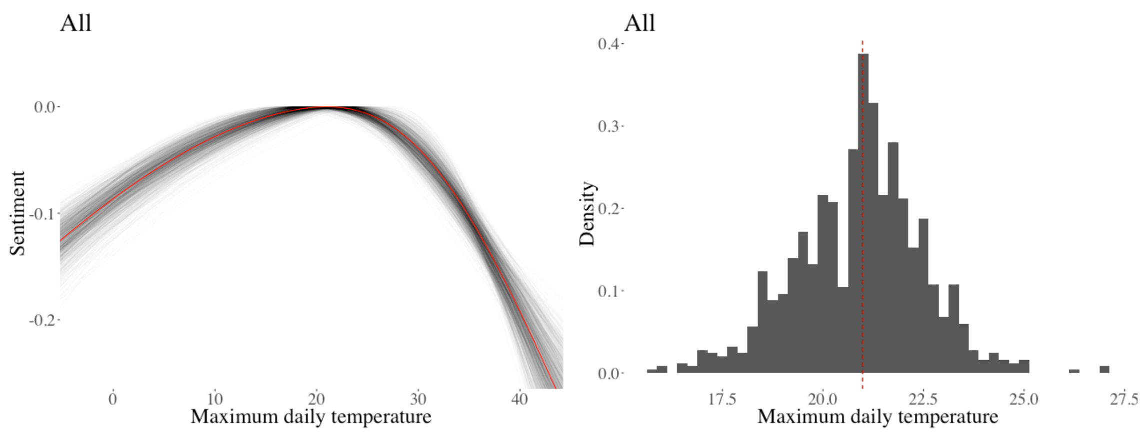
A.2 Sensitivity checks

In this section I document a series of checks intended to test the sensitivity of the estimates to a range of differences in specification and sampling frame.

A.2.1 Splined estimate

I test the sensitivity of the functional form I choose by replacing the binned $f(T)$ in Equation (1) with a flexible spline model. Specifically, I replace $\sum_{b \neq 20\text{-}25}^{B} \beta_b T_{cd}^{b}$ in Equation (1) with a set of basis vectors for a natural spline with knots at the 25th, 50th, and 75th percentile of observed daily maximum temperature in my data. To estimate standard errors, I bootstrap this model 1000 times, which produces the additional benefit of allowing me to estimate the preferred temperature for each run of the model. Figure A.3 documents this result; as expected, the shape of the response function in the left panel is similar to that found for the bottom-right panel in Figure 2. The histogram in the right panel documents the preferred temperature for each run of the model, where the median estimate of preferred temperature is at around 21.0 C.

Figure A.3: Effect of temperature on expressed sentiment (splines)



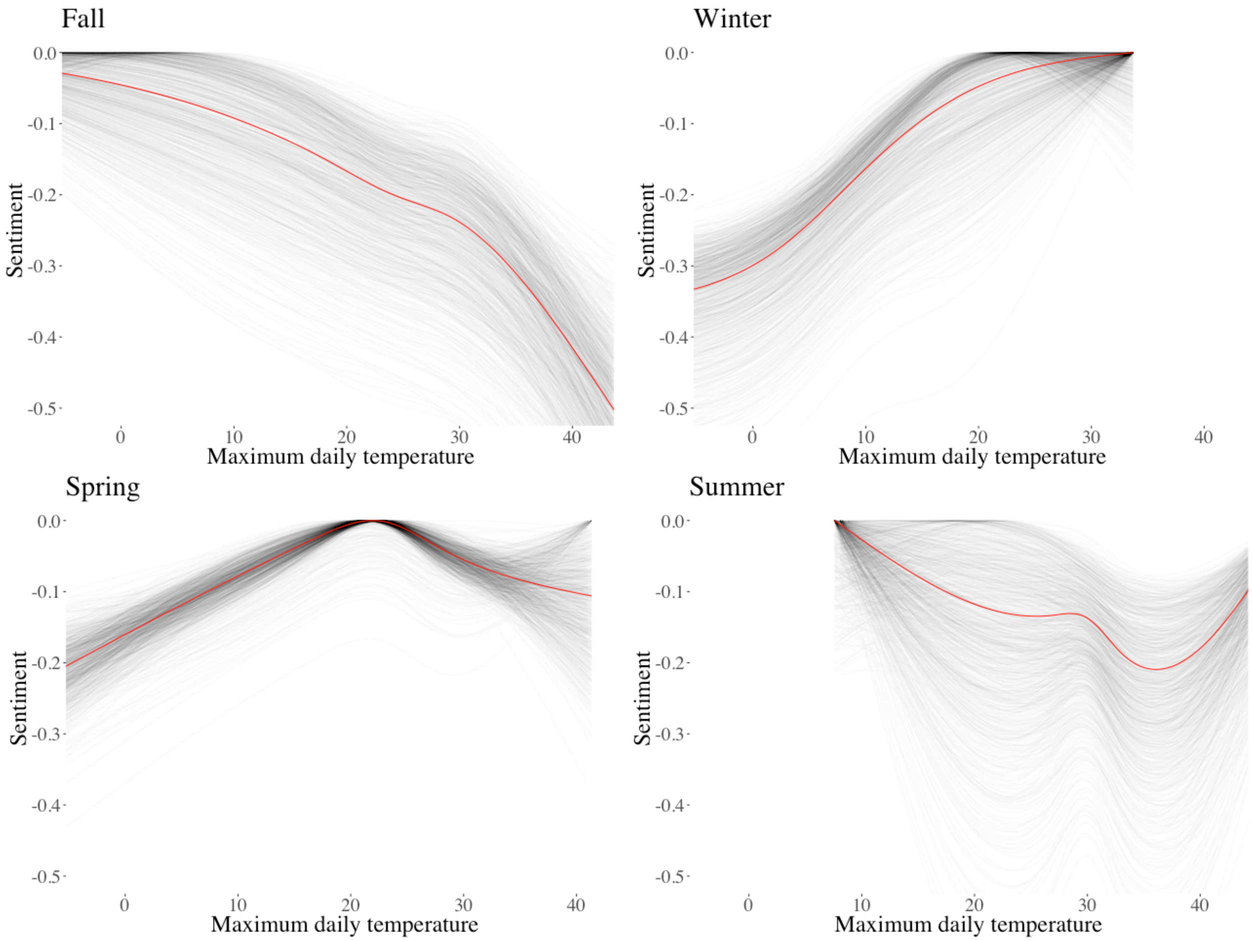
Notes: Left panel: documents the response of the expressed sentiment (measured using VADER and in standard deviations) to temperature using a splined model. Darker red line is the splined response function estimated from the entire dataset, with knots at the 25th, 50th, and 75th percentile of experienced daily maximum temperatures. Lighter gray lines are splines estimated similarly using 1,000 bootstrapped samples. All lines are normalized s.t. the highest point of the spline has y = 0. Regressions include CBSA, month, and year fixed effects. 95% confidence intervals clustered by CBSA and date. Right panel: Histogram of estimated preferred maximum daily temperatures for 1,000 bootstrap iterations.

A.2.2 Seasonal heterogeneity in temperature response

Willingness-to-pay estimates indicate that individuals value warm winters but cool summers (Albouy et al. 2016). This could reflect a stable underlying set of preferences or seasonal shifting of temperature preferences, where the latter would further suggest that adaptive concerns may be in order in this setting. To distinguish between these two possibilities, I estimate the model given in the previous section by season rather than region, where winter is defined as December to February, spring is March to May, summer is June to August, and fall is September to November. Figure A.4 documents the response functions by season. For the purposes of interpretation, the height of each line at a given point should be interpreted as the sentiment response to that temperature relative to the preferred temperature for that sample. I find evidence of seasonality in the response: in the summer and fall, individuals have downward sloping preferences for temperatures: cool temperatures are preferred to warm temperatures. In winter, by contrast, the opposite is the case: preferences slope upwards, indicating that warm winter days are preferred. Strength of preferences in spring are more muted: the "bliss point" centers around 21 C, but the declines expressed sentiment in response to either warm or cold temperatures are more modest. Readers should note that the scale of the outcome variable here is larger than that in Figure 2: the strength of these within-season preferences is masked by the aggregation across the year.

The seasonal responses in Figure A.4 suggests important seasonal differences in the effect of temperature on expressed sentiment. The negative impact of cold temperatures in Figure 2 seems to be due to the combination of both winter and spring responses, since in neither summer nor fall do I observe a negative response to cooler temperatures. This finding suggests that, if anything, individuals become more sensitive to cold (heat) during typically colder (warmer) seasons, and that adaptation does not seem to occur on a seasonal time frame.

Figure A.4: Sentiment responses to temperature differ by season



Notes: Panels document the response of the standardized VADER measure to temperature for each of the four regions, where regions are defined by quartiles of average daily maximum temperature during the sampling frame and are labeled, in order of increasing temperature, "Coldest", "Cold", "Warm", and "Warmest". Dark red line is the splined response function estimated from the entire dataset, with knots at the 25th, 50th, and 75th percentile of experienced daily maximum temperatures. Lighter gray lines are splines estimated similarly using 1,000 bootstrapped samples. All lines are normalized s.t. the highest point of the spline has y = 0. Regressions include CBSA, month, and year fixed effects. 95% confidence intervals estimated using cluster robust standard errors by CBSA and date.

A.2.3 Additional weather covariates

Because different aspects of weather are frequently correlated, models that omit a key meteorological driver of a given outcome may induce a bias in the estimates of the included weather covariates (Auffhammer et al. 2013). Because the weather dataset I use includes precipitation as well, I include both temperature and precipitation in Equation (1) in order to avoid absorbing the effect of precipitation on expressed sentiment in the temperature estimates. However, since prior findings indicate that a variety of weather variables can impact stated mood (Dennisenn et al. 2008), I estimate a model with additional weather covariates compiled from the QCLCD weather station data described in Section 2. To minimize measurement error, I include only CBSAs with a QCLCD weather station present.

Table A.7 tabulates the regression results from a model that adds relative humidity, wind speed, air pressure, and the percent of the day that was reported as overcast to Equation (1). The results are qualitatively similar, but document a more dramatic decline in mood in higher temperatures. Relative humidity and % overcast both negatively affect expressed sentiment, but their effects are small relative to the reported change in sentiment resulting from temperature.

Table A.7: Additional weather variables

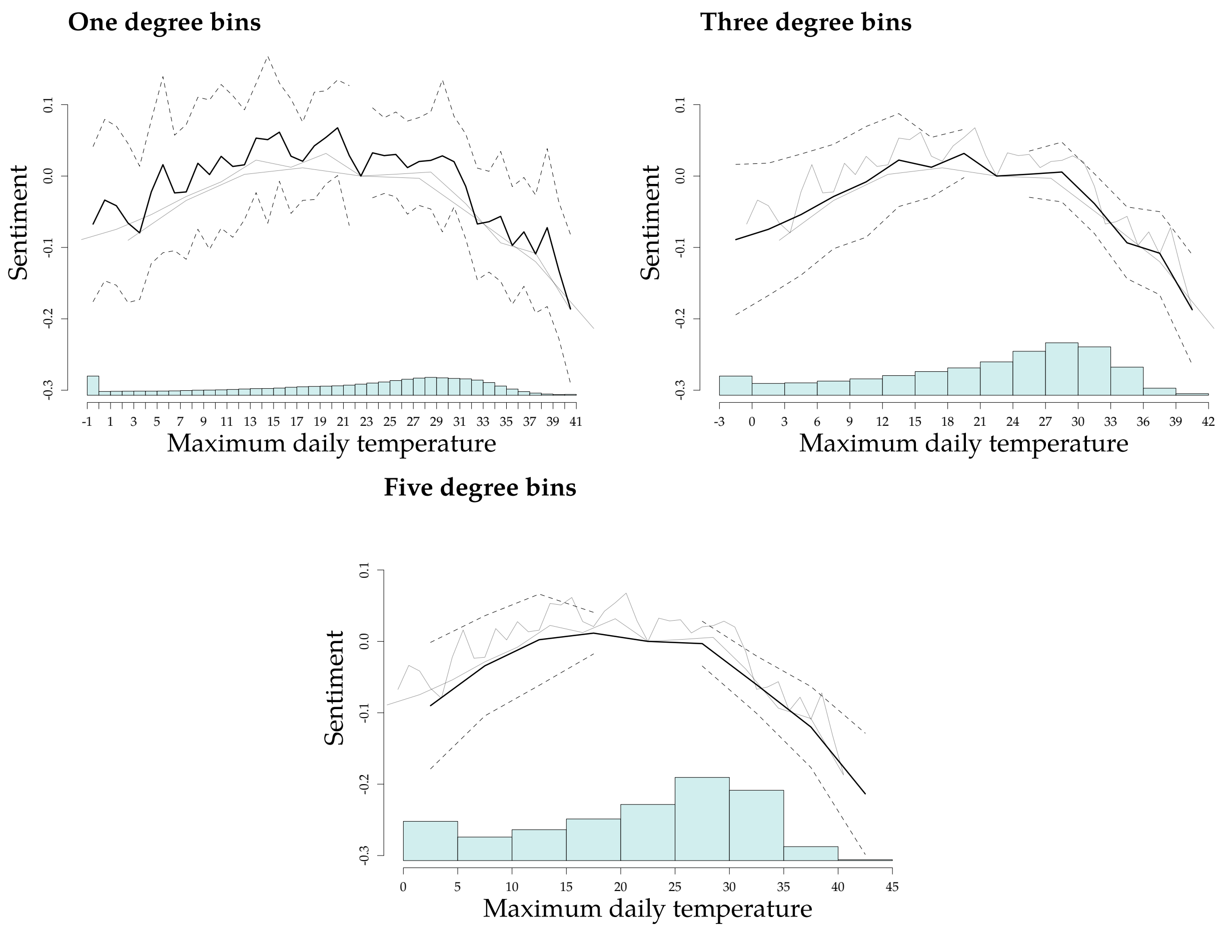
	(1)	(2)	(3)	(4)
Maximum daily temperature T				
$T \leq 5$	−0.11	−0.08	0.03	−0.11
	(0.06)	(0.05)	(0.07)	(0.04)
$T \in (5, 10]$	−0.06	−0.01	0.06	−0.04
	(0.05)	(0.04)	(0.05)	(0.04)
$T \in (10, 15]$	−0.01	0.02	0.06	0.00
	(0.04)	(0.03)	(0.04)	(0.03)
$T \in (15, 20]$	−0.00	0.02	0.05	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
$T \in (25, 30]$	−0.02	−0.01	−0.05	−0.01
	(0.02)	(0.02)	(0.02)	(0.02)
$T \in (30, 35]$	−0.08	−0.09	−0.12	−0.08
	(0.02)	(0.02)	(0.02)	(0.02)
$T \in (35, 40]$	−0.15	−0.17	−0.19	−0.15
	(0.04)	(0.03)	(0.04)	(0.03)
$T > 40$	−0.27	−0.26	−0.21	−0.23
	(0.05)	(0.05)	(0.06)	(0.05)
Precipitation	−0.00	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Relative Humidity	−0.00	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Wind Speed	−0.00	−0.00	0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Station Pressure	0.04	−0.03	−0.04	−0.01
	(0.09)	(0.06)	(0.06)	(0.06)
Overcast	−0.08	−0.09	−0.09	−0.11
	(0.03)	(0.02)	(0.02)	(0.02)
Fixed effects				
CBSA	Yes	Yes	Yes	Yes
Month	Yes	Yes		Yes
Year	Yes	Yes	Yes	Yes
DOW, Hol		Yes	Yes	Yes
S×M			Yes	
MOS				Yes
State trends	Yes	Yes	Yes	Yes

Notes: Dependent variable is sentiment in a CBSA-day. Coefficients represent the change in standard deviations of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. Units of air pressure are inches in hundreds, overcast is a variable from zero to one capturing proportion of daytime with overcast sky. Standard errors clustered by CBSA and date.

A.2.4 Bin widths

Because statistical models employing bin specifications can sometimes be affected by the selection of bin width, I estimate models with 1, 3, and 5 C bin widths in Figure A.5.

Figure A.5: Bin width



Notes: Comparison of the sentiment response to temperature across bin widths of 1, 3, and 5C. Dark solid and dashed lines indicate model estimates and standard errors for the given bin width, gray lines illustrate estimates for other bin widths for comparison. All models include precipitation P, state time trends, and CBSA, month, year, day of week, and holiday fixed effects. Standard errors clustered by CBSA and date.

A.2.5 Choice of sampling frame

Table A.8 documents the sensitivity of the results to the choice of sampling frame. I begin with the preferred specification and choice of sampling frame from Table 3, which excludes weather-related tweets (see Table A.4 for stopwords used) and tweets from users who post more than 25 times a day on average. The second column, "All", includes all users but continues to exclude weather-related tweets. The third column, "≥ 100 also excludes infrequent users from the baseline sample, measured as those who tweet less than 100 times in the sample. One possible explanation is the the impact of sentiment on temperature is a solely a reflection of the documented impact of temperature on criminal behavior (Ranson 2014) or on mortality (Deschênes and Greenstone 2011). If this were the case, these estimates would not reflect direct tastes for climate, but distaste for crime. The fourth column, "No crime/death", eappsec:comparegeneralxcludes tweets related to either crime or death. Appendix A.1.3 describes the construction of these word lists in more detail. The final column excludes tweets directed at other users, i.e., "@mentions", which tend to serve as direct communication between users. Note that the final three columns separately impose their restrictions relative to the Baseline sampling frame. For example, the "No mention" column does not exclude crime-related tweets.

In general, I find that the estimates are remarkably consistent across sampling frame: I observe negative impacts of both temperature extremes on sentiment that tend to be more pronounced for hot temperatures. The sampling restriction that removes users with fewer than 100 tweets from the sampling frame reduces the absolute value of the point estimates relative to the baseline, but the qualitative finding remains the same.

Table A.8: Sensitivity to sampling frame

	Baseline	All	No crime/death	No mention
Maximum daily temperature T				
$T \leq 5$	−0.09	−0.07	−0.09	−0.13
	(0.05)	(0.04)	(0.04)	(0.04)
$T \in (5, 10]$	−0.03	−0.04	−0.03	−0.06
	(0.04)	(0.03)	(0.04)	(0.03)
$T \in (10, 15]$	0.002	0.003	0.01	0.0002
	(0.03)	(0.03)	(0.03)	(0.03)
$T \in (15, 20]$	0.01	0.01	0.02	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)
$T \in (25, 30]$	−0.003	0.01	0.002	−0.004
	(0.02)	(0.02)	(0.02)	(0.02)
$T \in (30, 35]$	−0.06	−0.04	−0.05	−0.07
	(0.02)	(0.02)	(0.02)	(0.02)
$T \in (35, 40]$	−0.12	−0.08	−0.11	−0.13
	(0.03)	(0.03)	(0.03)	(0.03)
$T > 40$	−0.21	−0.14	−0.20	−0.20
	(0.04)	(0.04)	(0.04)	(0.04)
N (millions)	0.7	0.7	0.7	0.7
Tweets (millions)	1160.6	1374.3	1013.7	735.1

Notes: Table shows estimates of weather on expressed sentiment using different sampling frames. Coefficients (in cells) represent the difference (in standard deviations) of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. All sampling frames exclude posts directly related to weather. "Baseline" replicates the estimates from Table 3 column (5), which excludes posts from users with 25 or more tweets a day on average. "All" includes all posts. The last three columns impose separate restrictions relative to the Baseline sampling frame. "≥ 100" additionally excludes users with fewer than 100 tweets overall. "No crime/death" additionally excludes tweets that contain crime or mortality-related words. "No mention" additionally excludes tweets directed at other users (i.e., "@mentions"). All specifications include CBSA, month, year, weekday, and holiday fixed effects, as well as state trends and precipitation. N (millions) is the number of observations (CBSA-dates). Tweets (millions) is the number of posts represented by each regression. Standard errors (in parentheses) clustered by CBSA and date.

A.2.6 Choice of weighting variable

Table A.9 documents the sensitivity of the results to the choice of weights. I begin with the same set of weights used in Table 3, the average number of posts in a CBSA over the sample. The second column uses population weights, and the third is an unweighted regression. I find limited differences across the three specifications. If anything, I find larger magnitudes using the alternative weighting choices.

Table A.9: Sensitivity to weights

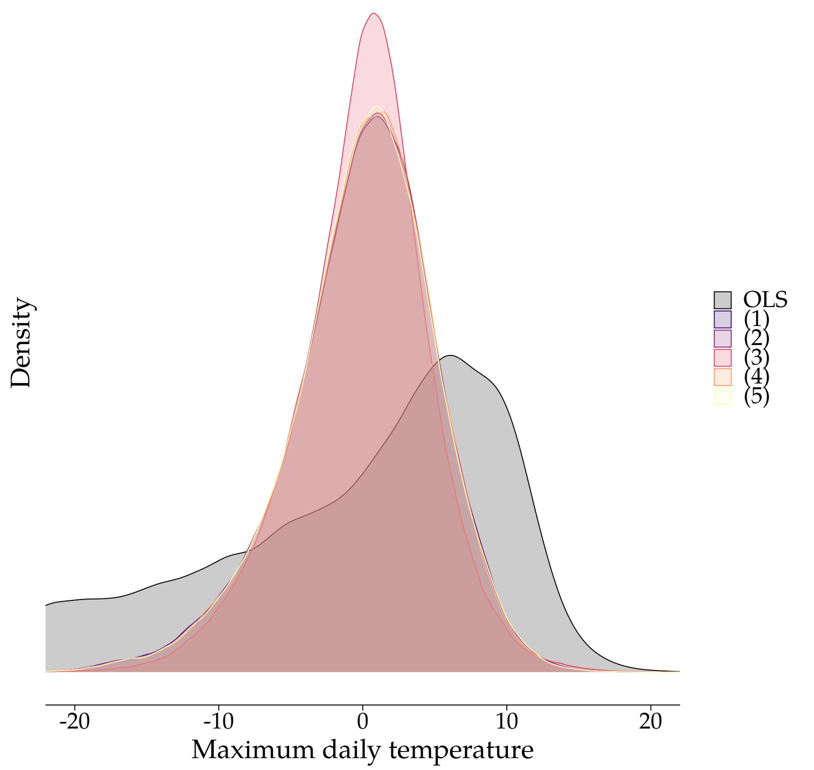
	Mean tweets	Population	Unweighted
Maximum daily temperature T			
$T \leq 5$	−0.09	−0.11	−0.09
	(0.05)	(0.04)	(0.05)
$T \in (5, 10]$	−0.03	−0.05	−0.07
	(0.04)	(0.04)	(0.04)
$T \in (10, 15]$	0.002	−0.01	−0.02
	(0.03)	(0.03)	(0.03)
$T \in (15, 20]$	0.01	0.004	−0.02
	(0.01)	(0.02)	(0.02)
$T \in (25, 30]$	−0.003	−0.002	−0.03
	(0.02)	(0.02)	(0.02)
$T \in (30, 35]$	−0.06	−0.07	−0.11
	(0.02)	(0.02)	(0.02)
$T \in (35, 40]$	−0.12	−0.12	−0.16
	(0.03)	(0.03)	(0.03)
$T > 40$	−0.21	−0.23	−0.16
	(0.04)	(0.04)	(0.10)
N (millions)	0.7	0.7	0.7
Tweets (millions)	1160.6	1160.6	1160.6

Notes: Table shows estimates of weather on expressed sentiment using different weights. Coefficients (in cells) represent the difference (in standard deviations) of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. "Mean tweets" replicates the estimates from Table 3 column (5), which are weighted by the average number of tweets in the CBSA during the county. "Population" takes as its weights the 2015 population for each CBSA, and "Unweighted" runs an unweighted regression. All specifications include CBSA, month, year, weekday, and holiday fixed effects, as well as state trends and precipitation. N (millions) is the number of observations (CBSA-dates). Tweets (millions) is the number of posts represented by each regression. Standard errors (in parentheses) clustered by CBSA and date.

A.2.7 Residual variation

Figure A.6 shows density estimates for the residuals of a regression of the level of daily maximum temperature on an intercept and the five specifications in Table 3.

Figure A.6: Residual variation



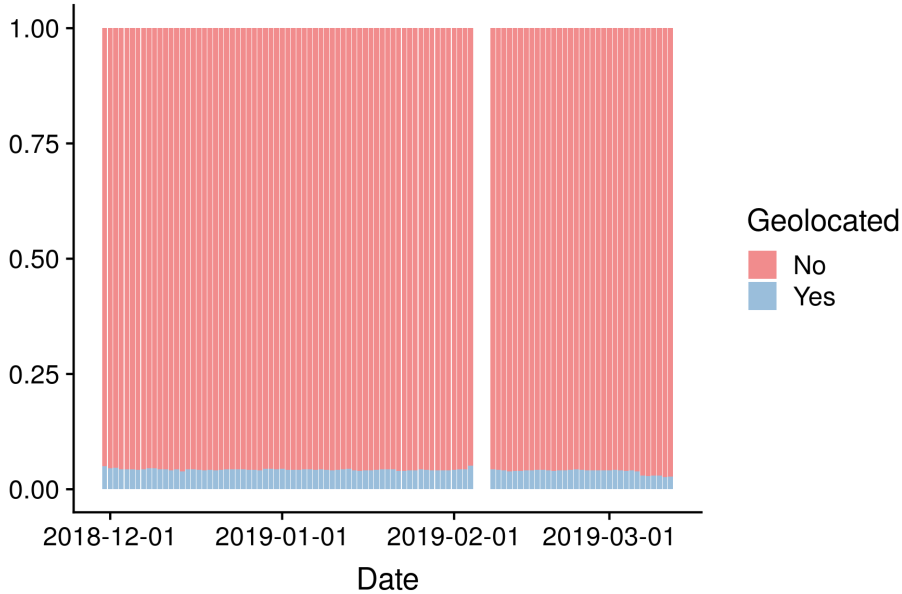
Notes: Density estimates of residual variation for columns of fixed effects in Table 3. Estimates are constructed by regressing maximum daily temperature on the set of controls in each column in Table 3 and plotting the density of the residuals from that regression.

A.3 Representativeness of Twitter sample

This paper identifies preferences for temperature using a sample of Twitter users who choose to geolocate their tweets. While this sample represents more than 11 million users overall, it is nevertheless a selected sample in two senses: first, these are users are active on Twitter, and second, these tweets must have geolocated in order for me to identify the weather they face. It could be the case that these users are fundamentally different from either the population at large and/or from the population of Twitter who do not choose to geolocate their tweets.

This section documents a series of tests designed to evaluate the degree to which the

Figure A.7: Proportion of geolocated and non-geolocated tweets by date, 2018-2019 sample



Notes: Figure shows proportion of geolocated (blue) versus non-geolocated (red) tweets by date. Sampling frame is from late 2018 to early 2019, drawn from a 1% sampling of all tweets posted during this time, but excluding non-english tweets and retweets.

estimates in the paper are represented of the full set of Twitter users and the population at large.

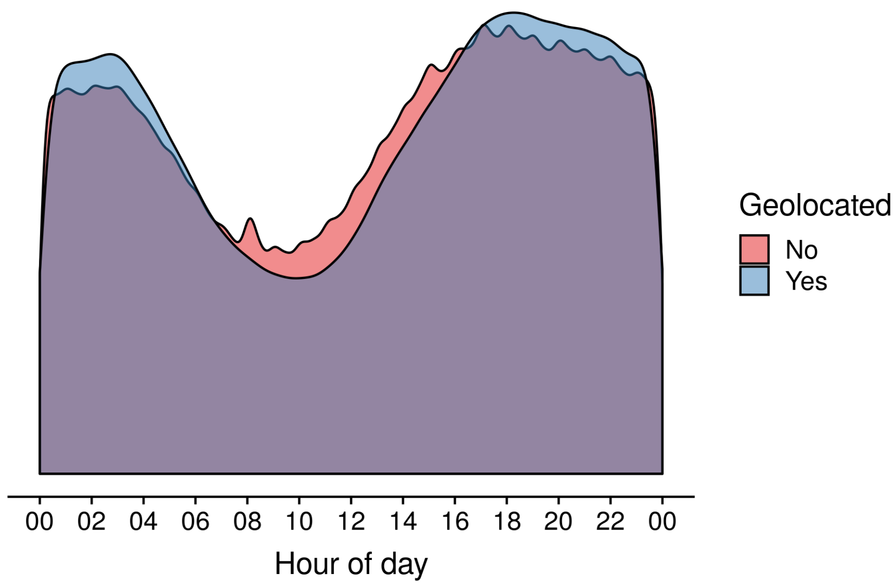
A.3.1 Comparing geolocated tweets to all tweets

Since geolocation is enabled by choice, it may be the case that geolocating users are different from non-geolocating users. Because Twitter does not provide demographic data on its users, identifying these differences is not straightforward.

The following section provides a series of exercises designed to provide suggestive evidence regarding the degree to which the geolocating users differ from non-geolocating users. To collect a comparable set of geolocated and non-geolocated tweets, I download a 1% sample of all tweets for three consecutive months from late 2018 to early 2019. I restrict this sample in two ways: first, I select only tweets written in English (Twitter provides a field that computationally identifies the language in which each post is written) and only non-retweets (to remain consistent with the sample in the paper, which also excludes retweets). I then classify each tweet by whether it was provided with included geolocation information or not. In total, I collect about 41 million tweets during this time period. There are a few days in February during which this streaming client was unable to access Twitter. Figure A.7 gives the proportion of tweets over time that are geolocated — on average roughly 4.5%.

First, I show that these two groups of users tweet at roughly the same times of day.

Figure A.8: Distribution of geolocated and non-geolocated tweets by time of day



Notes: Figure shows densities of geolocated (blue) and non-geolocated (red) tweets by time of day. Sampling frame is from late 2018 to early 2019, drawn from a 1% sampling of all tweets posted during this time, but excluding non-english tweets and retweets.

Second, I demonstrate that these groups use similar vocabularies. Third, I show how the distributions of sentiment are similar across the two categories. Finally, I statistically estimate differences in means between the two groups for average sentiment and the probability of tweeting about weather, crime, and mortality. In general, I find that while there are some differences between the two populations, they are in general quite similar, at least along the dimensions I document here.

This evidence, as well as Dennisenn et al. (2008), who finds that the temperature sensitivity of mood does not respond to observable traits, indicates that the observed preferences for temperature in the geolocated Twitter population are likely to map reasonably well to the Twitter population at large. The following sections provide the relevant comparisons.

Time of day. First, I show that geolocated and non-geolocated tweets are posted at similar times of day. Figure A.8 documents the distribution of tweets over the course of the day, where time is in Coordinated Universal Time (UTC).

There are some slight differences. Geolocated tweets are more frequently produced in the late evening and early morning relative to non-geolocated tweets, although the degree of this difference in small.

Most frequently used words. Table A.10 shows that language use does not vary substantially between these two samples: the top 50 words across each group are very similar, and Kendall's tau rank correlation coefficient for the top 1000 words is 0.76.

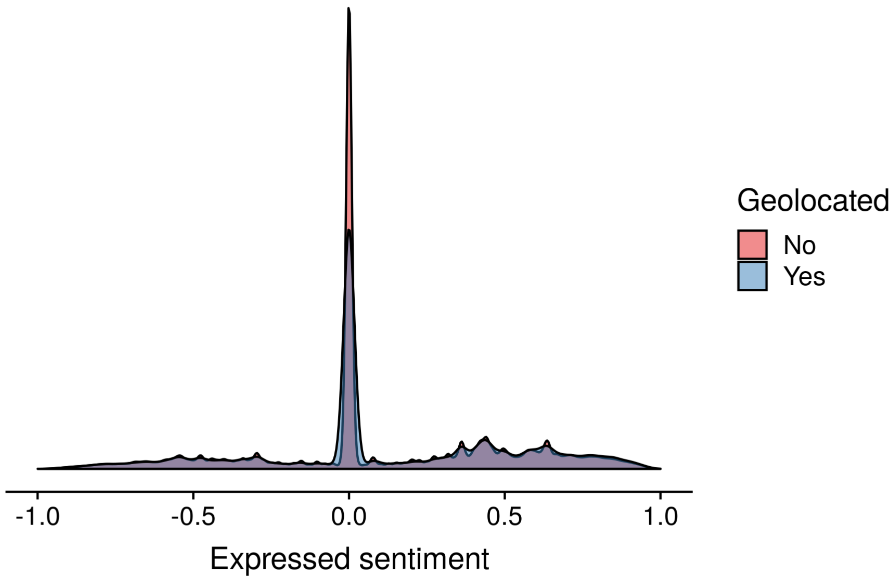
The close correspondence between the two rankings suggests that the population of

Table A.10: Comparison of top word usage across geolocation choice

Term	Rank (non-geolocated)	Rank (geolocated)
im	1	1
just	2	2
like	3	3
one	4	6
dont	5	4
love	6	7
get	7	5
can	8	8
now	9	11
people	10	18
good	11	9
know	12	12
new	13	17
time	14	10
u	15	37
go	16	14
see	17	16
day	18	15
amp	19	13
really	20	21
want	21	22
need	22	25
cant	23	20
think	24	28
got	25	19
thats	26	31
thank	27	29
back	28	24
youre	29	41
much	30	40

Notes: Rank of terms used in non-geolocated versus geolocated tweets. Corpus includes only English tweets that were not retweets delivered through the Streaming API from 2018-11-30 to 2019-12-14.

Figure A.9: Distribution of expressed sentiment across geolocated and non-geolocated tweets



Notes: Figure shows distribution of expressed sentiment (measured in unstandardized VADER scores) for geolocated (blue) and non-geolocated (red) tweets. Sampling frame is from late 2018 to early 2019, drawn from a 1% sampling of all tweets posted during this time, but excluding non-english tweets and retweets.

users who choose to geolocate their tweets are not markedly different from the Twitter population, at least in terms of their word choice while on the platform.

Distribution of sentiment. Third, I compare the distribution of sentiment scores (as measured using VADER, the primary measure of sentiment variable I use) across geolocated and non-geolocated tweets. Figure A.9 documents the density of sentiment. Since I am not comparing this measure to other measures, I do not standardize this measure as I do in the paper.

The distributions of the two samples are quite similar in terms of mean and variance. The non-geolocated sample tends to have a "peakier" distribution. This is likely because VADER scores are not continuously distributed: because of the small number of words in each tweet (and the smaller number of words that VADER counts as indicative of expressed sentiment), certain scores, e.g., zero, are more likely than others. Despite this peakiness, the distributions are closely overlapping.

Difference in means of sentiment and selected topic choice. As a final exercise, I estimate differences in means between expressed sentiment and in the probability of a post relating to weather, crime, or mortality. These latter three follow the word lists given respectively in Tables A.4 to A.6. I do so by running a simple OLS regression of the given outcome on a dummy for whether the tweet was geolocated or not. The results are given in Table A.11.

I note that the standard errors are small here as a result of the very high sample size,

Table A.11: Differences in means of sentiment and topic choice by geolocation

	Sentiment	P(Weather)	P(Crime)	P(Mortality)
Geolocated	0.01	0.004	−0.002	0.01
	(0.0003)	(0.0001)	(0.0001)	(0.0003)
Intercept	0.11	0.01	0.01	0.12
	(0.0001)	(0.0000)	(0.0000)	(0.0001)

Notes: Table documents results of regressing expressed sentiment or a dummy for topic choice (weather, crime or mortality) on an indicator for whether a tweet included geolocation information. "Geolocated" represents the difference in means (with standard error in parentheses) between geolocated and non-geolocated tweets, while "Intercept" is the mean for non-geolocated tweets. Sampling frame is from late 2018 to early 2019, drawn from a 1% sampling of all tweets posted during this time, but excluding non-english tweets and retweets.

since each regression contains 41 million observations, and focus instead on the magnitude of the point estimates. Average sentiment is very slightly higher for geolocated tweets and individuals are slightly more likely to post tweets that include words related to mortality. There are very small differences in the probability that a given tweet is related to weather or crime.

A.3.2 Comparing Twitter users to the population at large

Understanding the degree to which the estimates I obtain apply to the population at large requires a different strategy than the one I use above (in comparing geolocated and non-geolocated tweets). By definition, it is impossible to use Twitter posts for a population that does not use Twitter. Instead, I document a series of assessments designed to provide suggestive evidence of whether or not this sample is likely to be representative of the population at large. First, I report demographic statistics obtained by survey of representative respondents: I find that a large portion of American adults are on Twitter, and that those adults seem to be relatively less selected than most online platforms about which respondents were asked. Second, I discuss a previous paper which examined the degree to which various behavioral risk factors responded to temperature. Third, I present estimates of the cross-sectional correlation between the measures of expressed sentiment I use and three independent surveys of subjective well-being or life satisfaction, and find that, for two out of the three measures, expressed sentiment correlates with reported subjective well-being, although these correlations are fairly weak. Finally, I collect data on CBSA-level wages during the sample to test the degree to which response to temperature correlates with income. Overall, the body of evidence here seems to indicate that these users and their responses to temperature are likely to be representative of those that the population at large would exhibit.

Demographic statistics of Twitter users. Pew Research Center (2019) document the use of a range of online platforms, Twitter included, by different demographic groups (as determined by a nationally representative sample of over 1000 respondents). Twitter is slightly more likely to be used by men than women (24% to 21%) and by those under 30 compared to those over 30-49. 50-64, and 65+ (38% to 26%, 17%, and 7%, respectively). As expected, it is also more heavily used by wealthier, more educated, and more urban populations. Twitter is also somewhat unusual in that it is more likely to be used by the average black or Hispanic person than the average white person, although these differences are quite small (24% and 25% to 21%). Overall, however, the Twitter sample represents both a large amount of the total adults at 22%, though it must be said that the population of users who post regularly is likely to be somewhat smaller. Demographically, Twitter appears to be much less differentially selected than other platforms such as Instagram, Pinterest, or Reddit, and comparably selected to platforms such as Facebook and Youtube (for example, both Facebook and YouTube have a 10 percentage point gap between men and women). To that end, this evidence suggests these users are likely to more mobile and less vulnerable to the impacts, and therefore the estimates I find could be interpreted as a lower bound.

Previous work. Obradovich et al. (2018) investigates the impact of a range of weather and climate variables on reported mental health in the Behavioral Risk Factor Surveillance System (BRFSS), a randomly sampled set of respondents from across the United States. They find that respondents in that sample reported more mental health issues after experiencing warmer temperatures over both a single year and over multiple years (i.e., the effect was not washed away by adaptation), as well as responses to cyclones and to precipitation. These findings, particularly those related to warmer temperatures, are consistent with the impacts that I find in this paper.

Cross-sectional comparison of subjective well-being measures. I next compare the measures of expressed sentiment I use to publicly available surveys on subjective well-being and life satisfaction. I consider three sources in particular: the American Time Use Survey (ATUS), the BRFSS (discussed above), and Gallup polls.

The 2010, 2012 and 2013 versions of the ATUS included a subjective well-being module, although these periods do not overlap with my sample. This module was designed to obtain estimates of sentiment per activity, but in 2012 and 2013 were also required to ask the following question: "Please imagine a ladder with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. If the top step is 10 and the bottom step is 0, on which step of the ladder do you feel you personally stand at the present time?". I collect these data and merge to the Current Population Survey demographic data to obtain the state in which each respondent resides, and take averages by state.

Historically, the Behavioral Risk Factor Surveillance System (BRFSS) included the question "In general, how satisfied are you with life?". In 2011, however, it stopped including it entirely. It re-emerged in 2012 as part of an optional module and was asked in only a handful of states in subsequent years (for example, only Minnesota, Rhode Island, and Louisiana in 2014 included the question). I use data from the 2010 survey, the last year during which the BRFSS included this question for the entire sample. Again, I take averages by state for the purposes of this comparison.

Finally, I collect the annual state-level index of subjective well-being from Gallup, currently known as the "Gallup-Sharecare Well-Being Index" (available here: `https://`

wellbeingindex.sharecare.com) from 2014 through 2016 and take the by-state average.

I compare these measures cross-sectionally to the averages of expressed sentiment I find by state and to the average Gallup index for the same year. As documented in Table A.12, the ATUS measure correlates weakly with the measures of expressed sentiment, but (surprisingly) not well with the Gallup index. The BRFSS has correlations of less than 0.05 with all measures, suggesting that it is either too distant in time or methodologically too different to be comparable to the other measures.

Table A.12: Measure correlations: state averages

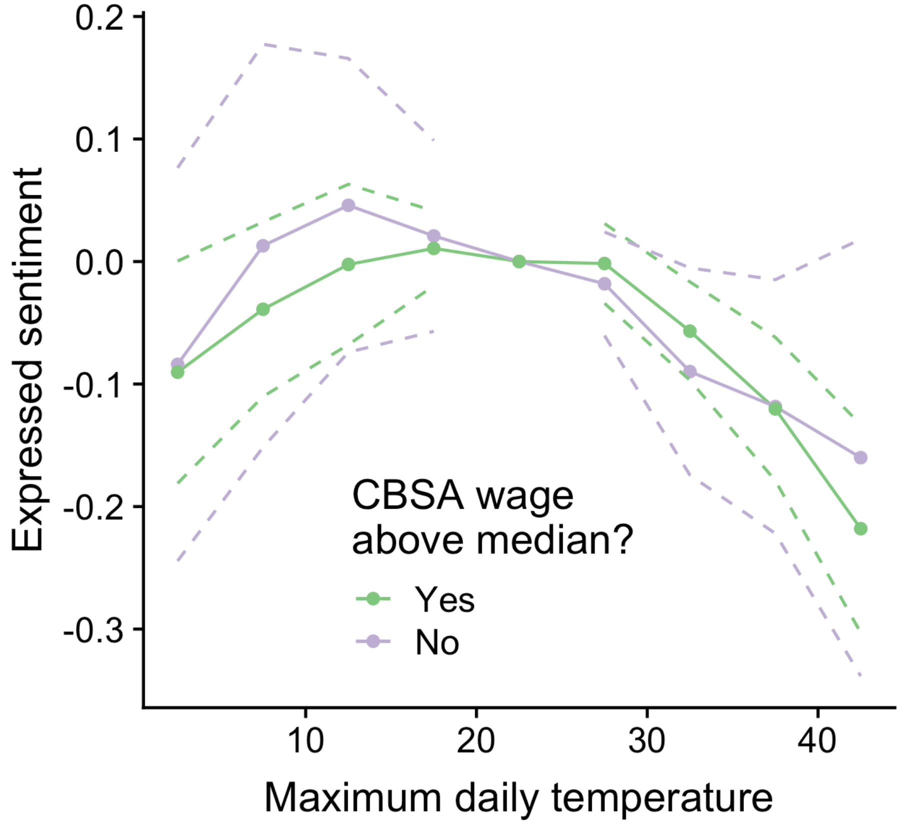
	AFINN	Hedon	LIWC	Vader	Prof.	ATUS	BRFSS	Gallup
AFINN	1.00	0.89	0.98	0.99	-0.97	0.23	-0.00	0.38
Hedon		1.00	0.90	0.91	-0.85	0.40	0.02	0.31
LIWC			1.00	0.98	-0.94	0.29	0.02	0.37
Vader				1.00	-0.96	0.26	-0.01	0.37
Prof.					1.00	-0.23	0.05	-0.31
ATUS						1.00	0.03	0.08
BRFSS							1.00	0.02
Gallup								1.00

Notes: Pairwise cross-sectional correlations of state annual averages of standardized expressed sentiment and profanity measures (2014-2016), ATUS Subjective Well-Being (pools 2010, 2012, 2013), BRFSS Life Satisfaction (2010), and averages of state-level measure of subjective well-being from Gallup (2014-2016).

Impact of temperature does not appear to vary by income. Finally, I collect BLS data on wages by CBSA during my sample to test the degree to which the response to temperature correlates with income, the only demographic measure I found to have sufficient differentiation by CBSA. [23] I split the sample into two groups: CBSAs with above and below average weekly wages during my sampling frame. I estimate the model given in Equation (1) with an additional set of interactions between the binned temperature variables and a dummy variable for wage indicator. Figure A.10 documents the results. I do not find statistically notable differences between CBSAs with below or above median wages, indicating that these marginal effects are not different across areas with higher or lower wages.

23. I also use these data in the main text in Section 4.2 (described in more detail in Appendix A.4.1) to estimate the value of expressed sentiment.

Figure A.10: Response of expressed sentiment does not differ by average wage



Notes: Figure documents the temperature responses for above and below average-wage CBSAs. I define above-wage CBSAs as those with an average weekly wage above the median average weekly wage in my sample. Dotted lines show 95% confidence intervals estimated using cluster robust standard errors by CBSA and date.

A.4 Valuing changes in expressed sentiment

The following sections describe how I obtain a per-SD value of changes in expressed sentiment, as used in Section 4.2.

A.4.1 Valuing changes in expressed sentiment using quarterly wages

To identify the impact of changes in wages on expressed sentiment, I collect data on quarterly wages for each CBSA from the Quarterly Census of Employment and Wages from the Bureau of Labor Statistics during the time period in my sample. I then estimate the following statistical model, where c stands for CBSA, s for state, q for quarter of year, and y for year:

$$\bar{S}_{cqy} = \beta \text{Daily wage}_{cqy} + \phi_q + \phi_{syq} + \varepsilon_{cqy} \tag{5}$$

S_{cqy} is the average expressed sentiment, as measured using VADER, in that CBSA for that quarter and year. Daily wage$_{cqy}$ is the average weekly wage in hundreds of dollars for that quarter, divided by seven. β is the coefficient of interest and represents the impact of increasing daily wages by 100$ for the entire quarter on expressed sentiment. ϕ_q is a CBSA fixed effect and ϕ_{syq} is a state-by-year-quarter fixed effect. In addition to the statistical model given above, I estimate models with a range of different fixed effects.

Table A.13: Impact of daily wages on expressed sentiment

	(1)	(2)	(3)
Daily wage (100s)	0.36	0.23	0.52
	(0.17)	(0.18)	(0.15)
Fixed effects			
CBSA	Yes	Yes	Yes
Quarter	Yes		
Year	Yes		Yes
YQ		Yes	
State-quarter			Yes

Notes: Table documents the impact of wages (in hundreds of dollars) on expressed sentiment on average expressed sentiment. YQ indicates "year-quarter", or quarter of sample. Standard errors (in parentheses) clustered by CBSA.

For all models, I find that increased weekly incomes are correlated (conditional on the fixed effects included) with higher expressed sentiment, although in some cases the impacts are only marginally statistically significant. Identification of β as causal, i.e., that it represents the impact of a policy that increased averages wages by $100, requires that there are no other factors which correlate with both expressed sentiment and weekly wages. This is, undeniably, a strong assumption: for example, economic activity could both increase wages and decrease unemployment. If reduced unemployment increases expressed sentiment, then β would capture the impact of both of those effects.

Nevertheless, interpreting these effects as causal is useful in the sense that it allows us to identify the value of a unit of expressed sentiment, and therefore to value the impact of other drivers of expressed sentiment, such as temperature. Using column (3), I obtain the value of a one SD change in sentiment by taking $\frac{1}{\beta/100}$, which gives \$196.77.

A.4.2 Valuing changes in expressed sentiment using speeding and parking ticket receipt

I demonstrate an alternative approach to valuing expressed sentiment by making use of plausibly exogenous variation in income by focusing on the population of users in my dataset who received parking or speeding tickets and who noted that receipt on Twitter. Using only individuals who had at least ten tweets in the seven days before and after the ticket, I document the sentiment response to receiving a ticket using an event study.[24] Intuitively, this approach compares tweets from users who received a ticket shortly before and shortly after they received it in order to control for cross-sectional variation (since users who receive tickets may be different from other users) and time-series trends (since users who become more likely to receive a ticket may be undergoing confounding behavioral changes). Specifically, I estimate:

$$S_{it} = \sum_{k=-7}^{K=7} \beta_k 1[\text{Date}_t - \text{Ticket date}_i + k] + \text{Trend}_t + \phi_s + \varepsilon_{it} \tag{6}$$

where i identifies the user, and t is the date. The β_k are coefficients reflecting the effect of receipt of the ticket on day k, where k number of days after the ticket was received. The top panel of Figure A.11 plots β_k over the entire period before and after the ticket, where the omitted category is all tweets not including in the 30 day window. The bottom panel of Figure A.11 plots the cumulative effect of a ticket over time, where each point estimate is $\sum_{k=0}^{K} \beta_k$ for $K \in [0, 7]$.

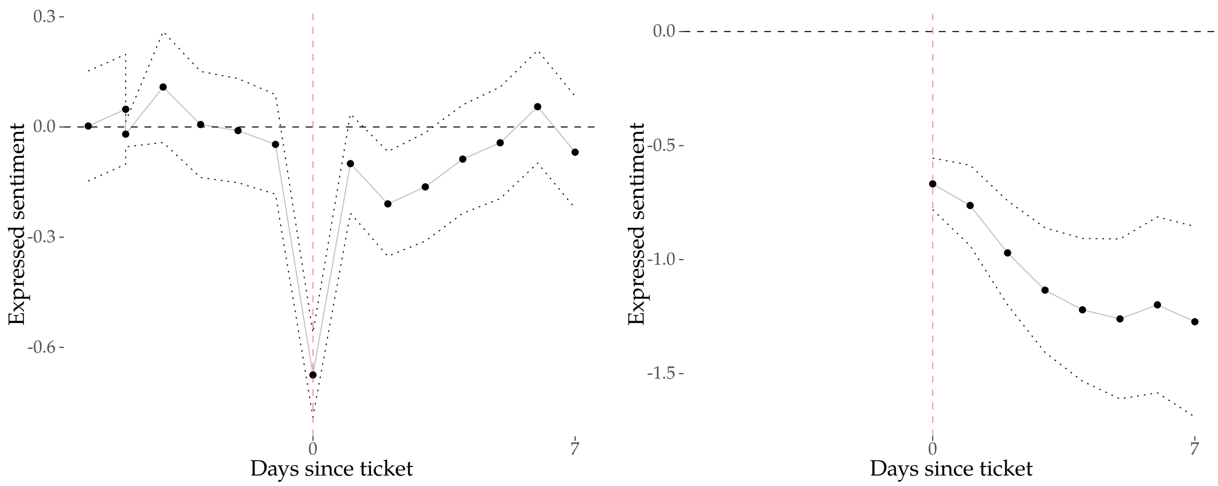
As expected, the receipt of a ticket causes a negative shock in expressed sentiment, which is most sharply experienced on the day of the ticket receipt, but accumulates over time, eventually resulting in a total loss of 1.27 SD of sentiment by day 7.

To value the sentiment impact of receiving a ticket, I divide the sum of the average changes in sentiment on the seven days following the ticket by the median value of the stated ticket, \$100.[25] Division of the \$100 by 1.27 SD results in a per-SD value of \$78.60.

24. It is reasonable to consider whether summing impacts on expressed in the days following ticket receipt is appropriate. Since I observe that users respond in the week following to a ticket received previously, I include the impacts of those days as well in the interest of remaining conservative. Omitting them would lead to a larger per-SD value of expressed sentiment.

25. The mean in this sample is \$164, driven by outliers with unrealistically large stated ticket costs. I use the median to mitigate the impact of these outliers and because it results in a more conservative estimate.

Figure A.11: Impact of parking ticket receipt on expressed sentiment



Notes: Left panel: event study estimate of the effect of the receipt of a parking or speeding ticket on standardized VADER sentiment, where receipt of a ticket is self-reported on Twitter. Sample limited to users who received at least one ticket during the sample period, and who had at least 10 tweets in the seven days before and after the ticket receipt. Right panel: estimates from dynamic cumulative lag model. where outcome is standardized VADER sentiment from an increasing number of days since ticket receipt. Point estimates and standard errors are the sum of coefficients on contemporaneous and lagged measures of dummy variable for ticket receipt, with increasing numbers of lags moving from left to right. 95% confidence intervals estimated using cluster robust standard errors by CBSA and date.

The second column in Table 7 applies this estimate to the contemporaneous estimates I obtain in column (5) of Table 3.

In contrast to the previous section, the source of variation in income here is an unexpected financial shock, but because it relies on a highly selected sample (users who experienced parking or speeding tickets during my sample) and may conflate the emotional distress of receiving a ticket with the financial cost, I emphasize that this, too, is a demonstrative exercise. It is reliant both on the validity of both the empirical strategy estimating the effect of temperature on sentiment described earlier in the paper and the one described in this section. Because the estimate obtained $\sum_{k=0}^{K} \beta_k$ in Equation (6) serves as the denominator, if this strategy overestimates the impact of a parking ticket on sentiment, then the implied valuation of the changes in temperature will be underestimates, and vice-versa.

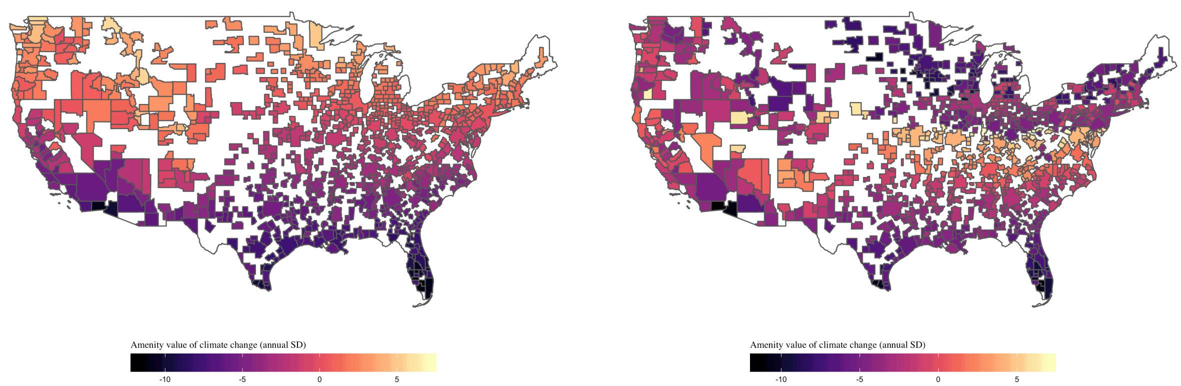
To identify individuals who received parking tickets, I search for tweets containing all of the following words: "got", "a", and "ticket", and at least one of the following words: "parking", "speeding", or "traffic". After a manual review of these tweets, I remove entries which contain any of the following phrases: "out", "she", "mom", "dad", "almost", "got away with", "you", and "ya". This results in about 8,000 ticket-related tweets. Next, I search these tweets for "$" symbols followed by a number to identify the cost of these tickets. Finally, I identify the user responsible for each tweet and retrieve the full range set of their tweets in my sample. I limit the sample to users with at least 10 tweets within

30 days of ticket receipt.

A.4.3 Maps of projected damages under RCP 8.5

Section 4.3 projects the main results into the future using climate projections. Figure A.12 documents regional heterogeneity of damages under RCP 8.5 and for both baseline and adaptive scenarios.

Figure A.12: End of century projections of changes in amenity values (RCP8.5)

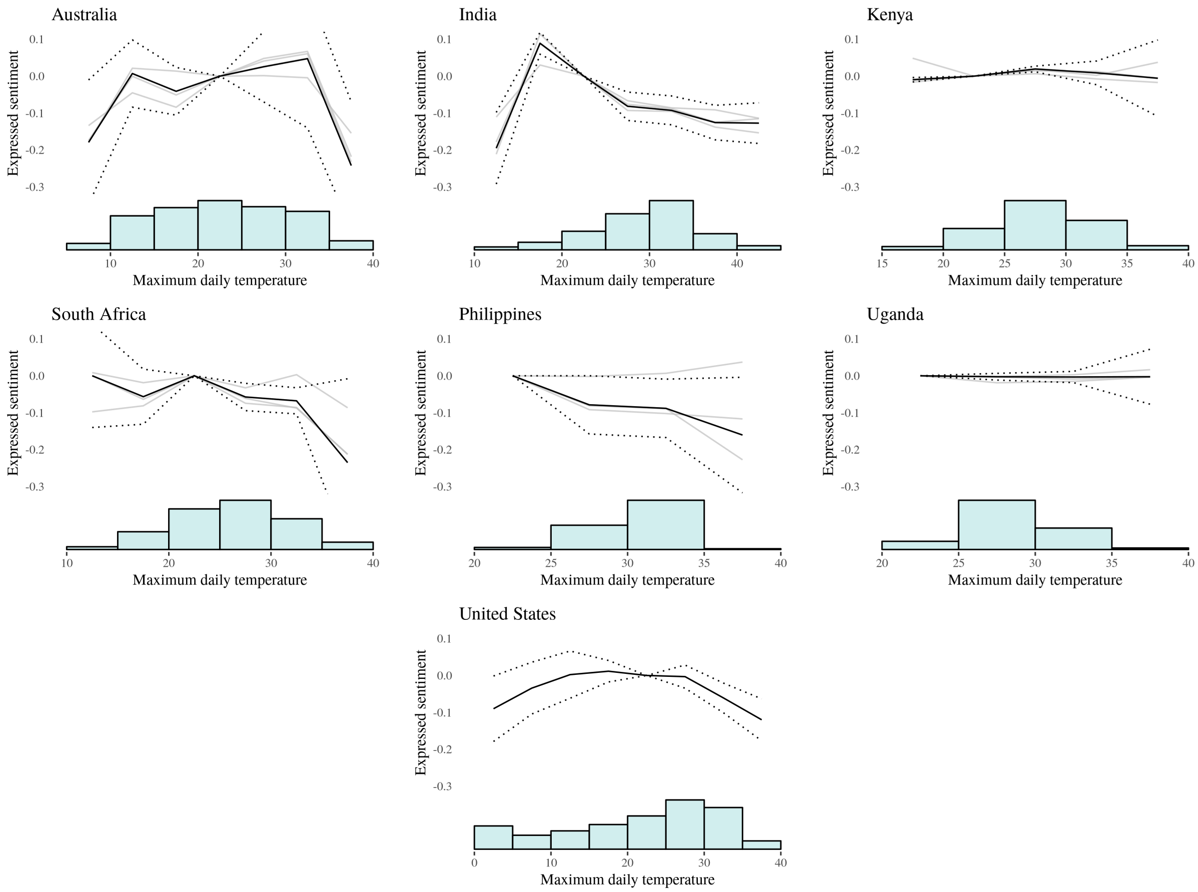


Notes: Left panel: projection of end-of-century climate damages in SD of expressed sentiment by CBSA under RCP8.5 using the baseline projection method. Right panel: projection of end-of-century climate damages in SD of expressed sentiment by CBSA under RCP8.5 using the adaptive projection method.

A.5 Comparison of sentiment response across all countries

Section 5 documents the response of expressed sentiment to temperature in six other English-speaking countries around the world in Figure 7. For space reasons, I do not include a equivalent response of the U.S. in that figure. Figure A.13 replicates Figure 7, but includes the U.S. response as well in order to facilitate comparison.

Figure A.13: Effect of temperature on expressed sentiment around the world (U.S. included)



Notes: Panels document temperature response for seven countries. Solid lines show the regression coefficients on temperature and represent the difference (measured in standard deviations) in CBSA-day sentiment for the temperature bin T_b relative to 20-25 C, controlling for state, month, and year fixed effects. For non-U.S. countries, light and solid lines represent estimates use alternative specifications of fixed effects including: state plus month of sample, state by month of year plus year, and state plus date. The United States response is provided for comparison and plots the estimates in Table 3, column (5). Dotted lines show 95% confidence intervals estimated using cluster robust standard errors by state.

References

Albouy, David, Walter Graf, Ryan Kellogg, and Hendrik Wolff. 2016. "Climate Amenities, Climate Change, and American Quality of Life." *Journal of the Association of Environmental and Resource Economists* 3 (1): 205–246.

Auffhammer, Maximilian, Solomon Hsiang, Wolfram Schlenker, and Adam Sobel. 2013. "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change." *Review of Environmental Economics and Policy* 7 (2): 181–198.

Bradley, Margaret Mm, and Pj Peter J Lang. 1999. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.* Technical report. The Center for Research in Psychophysiology, University of Florida.

Dennisenn, J., Ligaya Butalid, Lars Penke, and Marcel A.G. Van Aken. 2008. "The effects of weather on daily mood: A multilevel approach." *Emotion* 8 (5): 662–667.

Deschênes, Olivier, and Michael Greenstone. 2011. "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US." *American Economic Journal: Applied Economics* 3 (4): 152–185.

Dodds, Peter Sheridan, and Christopher M. Danforth. 2010. "Measuring the happiness of large-scale written expression: Songs, blogs, and presidents." *Journal of Happiness Studies* 11 (4): 441–456.

Gilbert, CJ, and Eric Hutto. 2014. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." In *Eighth International Conference on Weblogs and Social Media.*

Kahn, Jeffrey H., Renée M. Tobin, Audra E. Massey, and Jennifer A. Anderson. 2007. "Measuring emotional expression with the Linguistic Inquiry and Word Count." *American Journal of Psychology* 120 (2): 263–286.

Nielsen, Finn Årup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." *arXiv.*

Obradovich, Nick, Robyn Migliorini, Martin P. Paulus, and Iyad Rahwan. 2018. "Empirical evidence of mental health risks posed by climate change." *Proceedings of the National Academy of Sciences* 115 (43): 10953–10958. eprint: https://www.pnas.org/content/115/43/10953.full.pdf.

Pennebaker, James W, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. *The development and psychometric properties of LIWC2015.* Technical report.

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation." In *Empirical Methods in Natural Language Processing (EMNLP),* 1532–1543.

Pew Research Center. 2019. *Use of different online platforms by demographic groups,* April.

Ranson, Matthew. 2014. "Crime, weather, and climate change." *Journal of Environmental Economics and Management* 67 (3): 274–302.

Tausczik, Y. R., and J. W. Pennebaker. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29 (1): 24–54.