Temperature and Temperament: Evidence from Twitter

Patrick Baylis*

May 22, 2018

Can social media reveal preferences for environmental goods? I demonstrate a new approach to nonmarket valuation, combining more than a billion Twitter updates with natural language processing algorithms to construct a rich panel dataset of expressed sentiment. To begin, I document statistically significant declines in contemporaneous expressed sentiment below 12 C (54 F) and above 30 C (86 F) across four distinct measures of expressed sentiment, as well as corresponding increases in profanity use. Next, I use the richness of the dataset to generate insights into temperature preferences; I find evidence of heterogeneity in both regional and seasonal responses, and that the lagged impact of warmer temperatures actually reverses the sign of the contemporaneous impact. I conclude with a valuation exercise that combines the contemporaneous results with estimates of the sentiment impact of an unexpected financial shock to derive willingness-to-pay estimates for temperature.

*University of British Columbia, Vancouver School of Economics; Iona 113, 6000 Iona Dr., Vancouver, BC V6T 2E8; pbaylis@mail.ubc.edu. I am grateful to Maximilian Auffhammer, Severin Borenstein, and Solomon Hsiang for their invaluable suggestions, as well as to Michael Anderson, Hunt Alcott, 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 two anonymous referees. This paper was previously circulated under the title "Temperature and Temperament: Evidence from a billion tweets."

As the possibility of substantial changes in Earth's climate becomes more certain, economists have become increasingly interested in calculating the full scope of benefits and costs resulting from these changes. Acute environmental stressors like typhoons, hurricanes, and other marked changes in the external environment can have dramatic and immediate impacts on economic well-being (Hsiang and Jina 2014) while more gradual environmental changes, such as temperature increases, have more subtle but perhaps costlier long-run economic impacts (Burke, Hsiang, and Miguel 2015b). Other work has estimated the cost of changes in income, health, agriculture, civil conflict, natural disasters, and other economic outcomes (Carleton and Hsiang 2016).

A subset of the literature has examined whether increased ambient temperature will induce a significant change in the amenity value of the climate itself. Because outdoor meteorological conditions are non-rival and non-excludable, there are no direct markets from which researchers might infer preferences for outdoor temperature. This is a classic problem in the environmental valuation literature (Pearce 2002), and in the case of temperature most prior work has relied on hedonic valuation approaches to estimate individuals' willingness-to-pay for different climate characteristics. 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 (Cragg and Kahn 1997; Sinha and Cropper 2015; Albouy, Graf, Kellogg, and Wolff 2016). However, because the climate to date has varied relatively little across time, these values are identified using cross-sectional differences in climates.

This paper demonstrates a new, cost-effective method to estimate preferences over public 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. First, I sketch a conceptual model to suggest why sentiment might provide useful insights into the sign and possibly the strength of preferences for non-market goods. Next, I build the dataset using 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 unit of observation, although in the appendix I also estimate a model with individual tweets as the 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. I validate these measures by demonstrating that they correlate with one another, that state-level annual averages correlate with polling data on subjective well-being, and that the four measures all show similar within-week patterns of fluctuation.

The analysis uses 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 and accelerating decline in sentiment away from moderate temperatures, with peak sentiment occurring roughly around 22.7 C (72.7 F). The point estimate of the difference in expressed sentiment between 21-24 C and above 39 C is a statistically significant and between 0.1 and 0.2 SD, depending on measure used. This is comparable to roughly 75% of the average difference in sentiment observed between Sundays and Mondays. 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. As a means of understanding the mechanism by which sentiment responds to temperature, I also estimate the relationship between online profanity and temperature and find a U-shaped relationship, 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, I extend the baseline results to explore seasonal and regional heterogeneity in the response of sentiment to temperature. I find notable differences in both. 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. Regionally, areas that are colder tend to have stronger response to warm temperatures, and vice-versa. Finally, I document temporal features in the response using a cumulative dynamic lag model, which I estimate using deviations from the contemporaneously preferred temperature observed above. While there is no statistically significant lagged impact of colder-than-preferred temperatures on expressed sentiment, the lagged impacts for warmer-than-preferred temperatures are positive, and in fact negate and then reverse the contemporaneous impact after a few days.

I conclude the empirical exercise with a demonstration of a preliminary method to value contemporaneous shifts in sentiment using a subsample of users who received and tweeted about parking or speeding tickets. I document the cumulative sentiment change, including lagged effects, induced by the receipt of the ticket and combine it with the median ticket value in my sample to value changes in expressed sentiment: on average, the exercise implies individuals are willing to pay between \$0 and \$3 to experience a different ambient daily temperature. Finally, I discuss the implications of these findings for the existing literature on the value of climate amenities and assess the potential for social media data to serve as a complementary measure of nonmarket valuation.

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" approach, 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 also 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 (Interagency Working Group on Social Cost of Carbon 2013).[1] As of July 2014, the 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).

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

[1] Three IAMs are used to derive the social cost of carbon: DICE, FUND, and PAGE. Additional details and comparisons of models are described in Diaz (2014).

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 (Burke and Lobell 2012). But 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 (Deschnes and Greenstone 2011), and other outcomes. This paper focuses instead on the welfare cost of changes in amenity values resulting from rising outdoor temperatures.

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 environmental economics literature (Pearce 2002).

Climate can be viewed as a public amenity[2]: 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 and Cropper

[2] Economists typically refer to public goods and bads rather than amenities; I use the term 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.

(2015) and Albouy, Graf, Kellogg, and Wolff (2016) identify implicit values for different climates using observed choices about household decisions on where to live. These approaches are particularly appealing because it is straightforward to back out monetary valuations of different climates from the model estimates. However, since historical climate changes thus far have been modest at best, 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 models. 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.

Finally, Dennisenn, Butalid, Penke, and Van Aken (2008) uses an online survey to find that weather impacts are variable across individuals, but that those variations do not correspond to observable characteristics.

A small literature attempts to value environmental goods using self-reported happiness data (Welsch and Khling 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 attempts to assess these preferences and the challenges they face in controlling for unobservable sources of cross-sectional variation. The method I demonstrate 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 lagged responses and geographic variation in the response of expressed sentiment.

One view of expressed sentiment is that it captures a form of "experienced utility". The concept of experienced utility predates the modern neoclassical definition of utility, which for clarity and followingKahneman 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). To fix ideas, consider a characterization of expressed sentiment as the sum of recently experienced utility. Formally,

$$S(c) = \sum_{t=0}^{\infty}(1-\delta)^t[u(c_t) - u(c_{t-1})] \tag{1}$$

where S is sentiment, c is a bundle of goods, δ is the discount rate, and u is an experienced utility function. This formulation holds that expressed sentiment is the sum of recent changes in utility: the implication is that individuals who have experienced losses more recently will reflect those losses more strongly in their expressed sentiment (Kahneman and Sugden 2005). This model of sentiment abstracts from the underlying causes of changes in emotions (Russell 1980) but captures the intuition that emotional responses tend to be more attuned to recent events than to temporally distant ones. This conceptualization of sentiment aligns with previous work on emotions and affect: first, that temporary changes in external conditions, such as the weather, can affect statements about overall life satisfaction (Schwarz and Clore 1982), and second, that mood can influence decision making (Loewenstein and Lerner 2003).

The formulation of the model provides a convenient demonstration of the assumptions underlying the sentiment approach to preference analysis and valuation. Suppose that some shock at time t, e.g., a change in temperature, changes an individual's allocation of good c_t. Assuming that experienced utility in all periods prior to t is not affected by future changes in c, it must be the case that $\text{sgn}(\frac{\partial S}{\partial c_t}) = \text{sgn}(\frac{\partial u}{\partial c_t})$, or that the sign of changes in sentiment reflect the most recent changes in instantaneous utility.

2. Data

While it would be prohibitively expensive to estimate daily sentiment across the United States using, for example, a survey, publicly available updates on social media provide a low-cost alternative. By combining a large set of geo-located 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. Some 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 (Twitter 2015).

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.[3] I collect the vast majority of geolocated tweets produced within my sample period, which ends in October 2016.

Geo-located tweets are those for which the user has consented to have his or her location information shared. 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 location information entered by the user. In principle, Twitter limits the total number of tweets delivered through the Streaming API to 1% (Morstatter, Pfeffer,

[3] There are two substantial gaps, 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.

Liu, and Carley 2013) of the total tweets created. Since I request only geolocated tweets from 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. The top 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 in a quantitative measures is known as "natural language processing", or NLP. Within NLP, the set of techniques designed specifically to quantify expressed sentiment is known as "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, it is my view that analyses using expressed sentiment should demonstrate reasonable consistency across multiple measures and justify the choice of measures used. In this paper, I translate tweet content into four measures of expressed sentiment derived from prior work: Hedonometer (Dodds and Danforth 2010), LIWC (Pennebaker et al. 2007), AFINN (Nielsen 2011), and VADER (Gilbert 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 simply the unweighted average for 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 appendix 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. The online appendix 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. Table 2 shows the correlations between the five measures at the CBSA-day level. All of the measures are strongly positively correlated with each other, except the measure of profanity which is negatively correlated with all measures.

[Table 2]

Next, Table 3 documents the correlation between annual state-level averages for each measure and compares them to the annual state-level Gallup index of subjective well-being (Gallup, Inc 2014, 2015, 2016). In aggregate, the four sentiment measures are highly correlated. The last row compares the measures to state-level measures of well-being estimated by Gallup in 2016. These correlations are positive (except for the profanity measure, as expected), and range between 0.2 and 0.4, indicating that expressed sentiment does not perfectly capture well-being.

[Table 3]

I do not view this as a serious defect in these measures: expressed sentiment is a different object than subjective well-being, and cross-sectional differences in language and nuance could drive some of the differential. Nevertheless, this finding does suggest the importance of geographic fixed effects to account non-uniformity in language and tone across space. The measures capture notable geographic heterogeneity: the bottom panel of Figure 1 documents average sentiment (as measured by VADER) by CBSA across the country. Two patterns are visually evident: 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.

To better understand the measures, I conduct a validation exercise that examines how sentiment changes over the course of the days of the week. First, Figure 2 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 the results later in the paper, it is useful to note that the average difference in sentiment measure between Sunday and Monday is approximately 0.1σ across measures.

[Figure 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 measure of minimum temperature, maximum temperature, and precipitation at roughly 4×4 kilometer grid cells for the entire United States. The data are produced using the Parameter-elevation Relationships on Independent Slopes Model, which interpolates measurements from more than 10,000 weather stations and applies a regression correction to account for altitude and other influences on local climate that a distance-based interpolation method might fail to capture (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 (Center for International Earth Science Information Network Columbia University, United Nations Food and Agriculture Programme, and Centro Internacional de Agricultura Tropical 2005).

Prior work suggests that other weather variables besides temperature and precipitation may be drivers determinants of emotional state (Dennisenn, Butalid, Penke, and Van Aken 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.

3. Empirical approach

I identify the causal effect of temperature on 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. After accounting for this variation, I follow previous work in the field in interpreting the estimated coefficients as representing the causal effect of temperature on expressed sentiment (Dell, Jones, and Olken 2014).

Specifically, I estimate the following statistical model:

$$\overline{S}_{cd} = f(T_{cd}) + P_{cd} + \phi_c + \phi_m + \phi_y + \varepsilon_{cd} \tag{2}$$

Let c, d, m, and y index CBSA, day, month, and year. $\overline{S}_{cd}$ 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. $P_c d$ is daily precipitation. ϕ_c and ϕ_m, and ϕ_y represent CBSA, month of year, and year fixed effects. ε_{cd} is the idiosyncratic error term, which I cluster by CBSA. I estimate the model using weighted least squares, where the weights are counts of total scored tweets in a given CBSA.[4]

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.[5] 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 sizes 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 this choice reflects the prior finding that Americans prefer 65 F (18.3 C) average daily temperature (Albouy, Graf, Kellogg, and Wolff 2016). In my sample, because I use daily maximums rather than averages, this corresponds to the omitted bin that I choose.

The bottom 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 inter-regional variation in 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

[4] Because of the large number of dimensions of the fixed effects in this approach, I estimate most regressions using the R package `lfe` (Gaure 2013). [5] Because three does not multiply evenly into 40, the upper limit for the three degree bin specification is 39 C.

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 ϕ_m. 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. Finally, year fixed effects ϕ_y account for potentially correlated trends in both temperature and sentiment that are shared across the sample.

The combination of these fixed effects defines the identification strategy: 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), but possible only in this setting because of the density of the data. The model specified by Equation (2) represents the most defensible tradeoff between minimizing potential bias and maximizing residual variance but I document the sensitivity of the results to other 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.

4. Findings

Estimates of Equation (2) indicate statistically significant declines in sentiment below 12 C and above 30 C. For expositional clarity, I first present the main result for each sentiment measure in Figure 3. I show that the shape of the response functions is remarkably similar across the different measures of sentiment. Second, Figure 4 estimates a splined model for a single measure, VADER. Third, Table 4 tabulates the response of the VADER measure under a range of different choices of fixed effects.

[Figure 3]

Figure 3 documents the temperature response of all four measures of sentiment estimated using Equation (2). 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 largest cold-weather effect and the second-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 largest warm-weather effect and the second-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 the smallest impacts of temperatures on sentiment, with the largest magnitudes slightly less than 0.1 SD. The point estimates have wider confidence intervals and are statistically significant for warm weather temperatures, but less so for cooler temperatures. This difference from the other measures most likely results 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 0.13 SD at maximum. VADER estimates a slightly smaller response than either AFINN or Hedonometer, but a larger response than LIWC.

Each outcome measure in Figure 3 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 0.07 SD to 0.15 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. I next replace the binned $f(T)$ in Equation (2) with a flexible spline model. Specifically, I replace $\sum_{b \neq 20\text{-}25}^{B} \beta_b T_{cd}^{b}$ in Equation (2) 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 4 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 3. 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 22.7 C.

[Figure 4]

Table 4 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. Column (1) reflects Equation (2), the baseline specification, which includes month and year fixed effects. Similar to Figure 3, 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 2) and variation related to holidays, but the estimated effects are virtually unchanged. Column (3) substitutes month of sample fixed effects in place of the year and month fixed effects, again with little impact on the estimated coefficients. Column (4) introduces state-by-month fixed effects to account for regionally distinct seasonal trends. Unlike previous estimates, this does alter the shape of the response function: cooler temperatures are no longer perceived as unpleasant, and the preferred temperature seems to be lower, between 10 and 15 C. Column (5) replaces all of the temporal fixed effects with date fixed effects. This has different impacts on the low and high ends of the temperature response function: it increases the point estimates for cooler days and decreases point estimates for most warmer days, except for the hottest days, which have a still-larger impact.

To summarize, these impacts are largely stable across different selections of fixed effects. With one exception (cooler temperatures in the date fixed effects model), both hot and cold temperatures have a statistically significant negative effects on expressed sentiment. It should be kept in mind that these effects lose some stability when relying on increasingly narrow sources of variation. In appendix Figure A11, I document the residual variation available to the model with each set of fixed effects. The appendix also includes additional specification checks, including user-level estimates (Figure A12), inclusion of additional weather variables (Table A11), and variations on bin width (Figure A10), none of which qualitatively alter the baseline results.

[Table 4]

The negative relationship between temperature and sentiment below 12 C and above 30 C resembles that estimated by Albouy, Graf, Kellogg, and Wolff (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 about 0.1 SD, or roughly half of the difference in sentiment observed on a Sunday relative to a Monday. A large literature has documented the impact of climate on conflict (Burke, Hsiang, and Miguel 2015a); 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 also estimate the relationship between temperature and expressions of profanity. Using a list of more than 300 profanities, I estimate Equation (2) with the percent of tweets in a CBSA-day that contain a profanity as the outcome of interest. Figure 5 plots the results.

[Figure 5]

I find that use of profanity rises in both hot and cold temperatures. Previous work on both conflict (Burke, Hsiang, and Miguel 2015a) 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 2015a). 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.

5. 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.[6] 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.

[6] There is a burgeoning literature on understanding adaptation. For more complete discussions of the subject across a range of areas, see Auffhammer, Hsiang, Schlenker, and Sobel (2013), Houser et al. (2014), Graff Zivin, Hsiang, and Neidell (2015), Auffhammer (2013), and Dell, Jones, and Olken (2014).

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 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 a series of empirical tests designed to suggest whether preference adaptation or sorting is likely to occur in this setting. First, I estimate the degree of heterogeneity in temperature-sentiment responses both across the four quartiles average annual temperatures and across the four seasons of the year, finding important differences in the sentiment response across both of these dimensions. Second, I estimate the extent to which these effects persist over time. I find that medium run cold weather effects are consistent with contemporaneous estimates, while the inclusion of longer-run effects of warm temperatures actually negates and then reverses the contemporaneous effect. While these do not, individually or collectively, serve as sharp tests of adaptation, they do help to inform the extent to which temperature preferences do or do not adapt over time.

5.1. Regional heterogeneity in temperature response

Figure 6 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 return to the splined model of temperature in Figure 4. To represent standard errors, I bootstrap these estimates with 1000 iterations. For each region, the red line indicates the response identified for the entire sample for that season and the lighter gray lines indicates responses estimated for a set of bootstrapped sample. 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. There is one exception: the hottest region has a slightly smaller response to cold temperatures than the second-hottest region, but both responses are more pronounced than that in either the Cold or Coldest regions.

[Figure 6]

I view this as evidence of either preference adaptation, technological adaptation, or sorting, 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.

5.2. Seasonal heterogeneity in temperature response

Willingness-to-pay estimates indicate that individuals value warm winters but cool summers (Albouy, Graf, Kellogg, and Wolff 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 7 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 3: the strength of these within-season preferences is masked by the aggregation across the year.

[Figure 7]

The seasonal responses in Figure 7 suggests important seasonal differences in the effect of temperature on expressed sentiment. The negative impact of cold temperatures in Figure 3 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.

5.3. Cumulative effects of temperature

In this paper I have focused primarily on the contemporaneous, or same-day, effect of temperature on expressed sentiment. This follows the model of Equation (1) is because contemporaneous reactions appear to be the most likely to reflect underlying preferences, such as those that might be derived using an ex ante willingness-to-pay elicitation, since they capture real-time responses of individuals experiencing these various temperatures.

However, it is possible that these effects extend beyond the day in which they are experienced, in which case the contemporaneous estimate could under- or over-estimate the impact of a single day's temperature. To better understand this impact, I estimate a dynamic cumulative lag model designed to elicit longer-run effects of temperature on expressed sentiment (Stock and Watson 2007). This model is a set of regressions that begin with Equation (2) and incorporate a progressively increasing number of lags of $f(T)$. The sum of temperature coefficients over the included lags allows me to estimate increasingly longer-run effects of temperature on expressed sentiment. If the sum of the coefficients is stable at the level of the estimated contemporaneous effect (with zero lags in the regression), then the impact of a single day's temperature increase is only felt on that day, and has no subsequent effect. If the sum of the coefficients changes with the inclusion of more lags, then a change in single day's temperature also has an effect on the days following: an increase in magnitude indicates an increasing impact over time, while a decrease in magnitude indicates that subsequent effects may be "washing out" or negating the contemporaneous impact.

To simplify interpretation, I examine this possibility lagged version of the "degree-days" model used in the climate-agricultural productivity literature (Schlenker and Roberts 2009). I use the splined estimate of the temperature in which peak expressed sentiment is observed, 22.7 C, to create "heating degrees" (HD) and "cooling degrees" (CD) for the recent history of temperatures.[7]. Specifically, I estimate:

$$\begin{aligned}\overline{S}_{cd} = \sum_{j=0}^{M} \beta_j [T_{cd} - 22.7] + \sum_{k=0}^{M} \beta_k [22.7 - T_{cd}] + \\ \phi_c + \phi_m + \phi_y + \varepsilon_{cd} \quad \forall M = 0, 1, 2, ...30\end{aligned} \tag{3}$$

where β_j (β_k) is the effect of each degree above (below) 22.7 j (k) days before the given day, and M is the number of lags included in the model. To estimate the cumulative effect, for each M I plot the sums of the β_j and β_k in the model separately in Figure 8. The two lines in the figure represents the cumulative impact including additional lags in the model for hot (red, labeled CD) and cold (blue, labeled HD) temperatures.

Figure 8 plots the results of this dynamic cumulative lag model. The interpretation of

[7] Note that this is mathematically equivalent to a piecewise linear function in temperature with a break at 22.7 C, but the degree days interpretation is more straightforward.

this model is fairly clear: while both cold (heating degrees) and warm (cooling degrees) contemporaneous impacts are negative, the inclusion of more lags reveals that while the cold temperature effects are stable, the contemporaneous warm temperature effects are washed out by their lagged impact. In other words, while the impact of a hot day today on expressed sentiment is negative, the impact of a hot day today on expressed sentiment tomorrow is positive, as well as the effect on the day after, and the day after that, to the extent that after 3 days the cumulative impact is zero and after 7 days the impact is positive.[8]

[Figure 8]

This finding raises some interpretation questions for this analysis. On one hand, accounting for lagged effects is useful for understanding how temperature affects expressed sentiment. On the other hand, it's not obvious whether lagged reactions to temperature changes still represent the same ex ante preference that are the primary object of interest for this paper. The distinction can be summarized as follows: short-run responses to temperature hold an upside-down U shape, while longer run responses tend to be linearly increasing in temperature until the high 30s. Very hot temperatures remain less preferable.

More broadly, the regional heterogeneity in responses suggests that over long time periods individuals in hotter or colder parts of the U.S. have become partly inured to short-run variations in temperature, although I cannot distinguish whether this has been due to preference adaptation, technological adaptation, or geographic sorting. The seasonal heterogeneity suggests that preference adaptation does not occur seasonally, and that if anything a preference for novelty outweighs adaptation that might be occurring. Finally, the cumulative estimates imply that warm temperatures are contemporaneously dispreferred but have positive effects on sentiment a few days after they are felt. This is

[8] Appendix Table A12 plots a similar model for the binned specifications ands find similar results, although the hottest bin continues to have a negative effect on expressed sentiment. Appendix Figure A13 estimates a monthly, rather than daily, model, and also finds positive impacts of warm temperatures on sentiment, with a similar finding for the hottest bin.

not sufficient evidence to demonstrate short-run adaptation per se, but it does suggest that the interactions between contemporaneous temperature and the history of experienced temperature are likely to be important in understanding the amenity value of climate change.

6. Valuing 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 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, region- and season-specific response, and cumulative effects. These benefits must be weighed against a significant drawback of this approach: the challenge of interpretation. This is a problem also faced by work using measures of life-satisfaction as the outcome of interest: 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, Graf, Kellogg, and Wolff 2016; Sinha and Cropper 2015). By contrast, while I am able to intuitively calibrate the magnitude of the estimates presented here using comparisons to within-week variation in estimated emotional state, backing out estimate of the average willingness-to-pay for changes (or the lack of changes) in temperature is more challenging. 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).

One possible approach to this challenge is to follow Levinson (2012), who converts reported life satisfaction into a dollar value by dividing the response of life satisfaction to pollution levels with the the response of life satisfaction to cross-sectional differences in income. The drawback of this approach is that it is re-introduces the same cross-sectional concerns in the estimation of the denominator. In my data, examination of the bottom panel of Figure 1 reveals that the low relative expressed sentiment levels of southern and more rural areas would be likely to drive a positive correlation between income and sentiment.[9] However, the case for this as a purely causal relationship is weak, since demographic differences and sorting could easily produce a similar pattern.[10]

The ideal experiment would randomly give money to individuals and monitor the resulting change in sentiment. However, the noise in sentiment expression would require a large sample size and, as a result, a large budget. As an alternative, I make use of a different source of exogenous financial shocks: parking and speeding tickets.

I identify more than 8,000 instances where individuals in my sample received parking or speeding tickets. Using only individuals who had at least ten tweets in the 30 days before and after the ticket, I document the sentiment response to receiving a ticket using an event study approach. Specifically, I estimate

$$S_{it} = \sum_{k=-5}^{K=60} \beta_k 1[\text{Date}_t - \text{Ticket date}_i + k] + \text{Trend}_t + \phi_i + \varepsilon_{it} \tag{4}$$

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.

[9] Cross-sectional estimates of the "effect" of temperature on sentiment confirm this conjecture.

[10] There are nearly countless possible candidates for confounding or reverse causality here. A few examples: cultural factors that drive both happiness and economic productivity, weak labor markets that drive down average incomes and decrease happiness due to unemployment, or the well-known relationship between happiness and employability (Mackerron 2012).

The top panel of Figure 9 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 9 plots the cumulative effect of a ticket over time, where each point estimate is $\sum_{k=0}^{K} \beta_k$ for $K \in [0, 60]$. 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 5.77 SD of sentiment by day 60.[11]

[Figure 9]

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.[12] Division of the \$100 by 5.77 SD results in a per-SD value of \$17.32. Table 5 applies this estimate to the contemporaneous estimates I obtain in column (1) of Table 4. This approach resembles that taken by Levinson (2012), who values pollution using the cross-sectional relationship between stated happiness and differences in income, although I argue that my setting provides a more plausibly exogenous shift in income but in doing so relies on a more selected sample: individuals who received speeding or parking tickets and posted about them on Twitter.

[Table 5]

The estimates in column (3) of Table 5 imply, for example, that individuals in this sample would pay \$0.79 to exchange a day with maximum temperature above between 35 and 40 C for a day with maximum temperature between 20 and 25 C. These results should be interpreted with some caution: this estimate 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 (4) serves as the denominator, if this strategy overestimates the

[11] Longer-run impacts are stable, but increasingly noisy, after 60 days. [12] 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.

impact of a parking ticket on sentiment, then the implied valuation of the changes in temperature will be underestimates, and vice versa.

7. Discussion

By using a 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 approaches to non-market valuation. This approach allows me to estimate nonlinear responses of sentiment to temperature and to account for unobserved variation across both space and time. It also allows me to identify spatial and seasonal sources of preference heterogeneity, and to illustrate the difference in short-run and medium-run responses of sentiment to temperature shocks. The exploratory method I demonstrate in Section 6 makes use of plausibly exogenous financial shocks (a parking ticket) to value the changes in sentiment I estimate in the previous section, circling back to the importance of the valuation aspect of nonmarket valuation.

This approach is not without its drawbacks. Of course, the formation of emotional state is far more complex than the model in Section 1 portrays. The physical, biological, and psychological bases for human emotions remain only partly understood (Russell 1980), and the choice of a single dimensional affective scale that is responsive only to changes in levels of utility abstracts away from important nuances regarding the formation of emotion. Moreover, expressed sentiment, particularly from short pieces of text, is likely to be a poor measure of underlying emotional state, and may indeed be biased relative to the emotional state of the underlying populace: the geo-coding Twitter user population is not representative on the United States at large. Finally, the valuation exercise is dependent on the validity of both the temperature and parking ticket estimates.

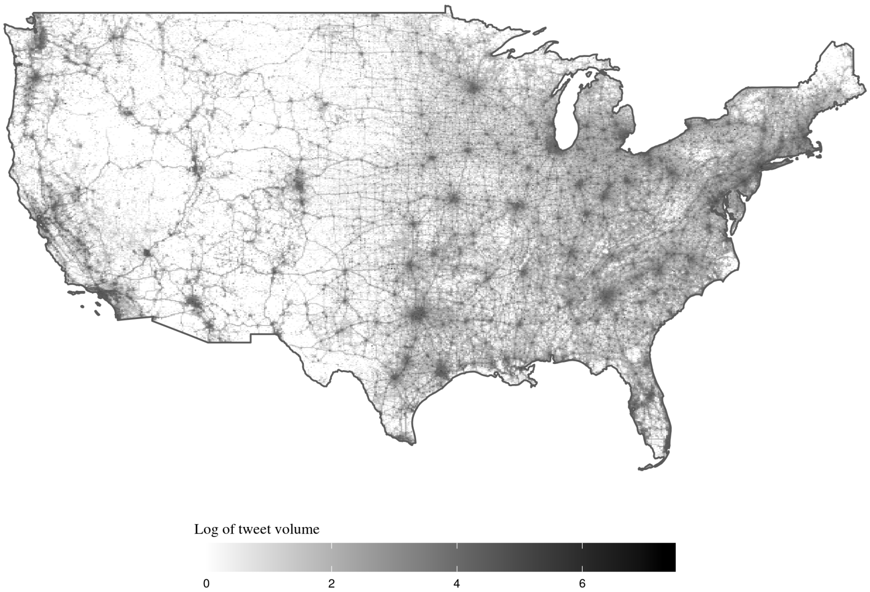
Despite these limitations, this paper makes several contributions to the literature. It introduces a new method and data source to estimate preferences for and valuations of public goods while simultaneously accounting for possible unobservable cross-sectional and seasonal variation. It reveals previously unobservable geographic, seasonal, and temporal dynamics of preferences for temperature and provides suggestive evidence of adaptive capacity in this area. And it demonstrates how NLP and specifically sentiment analysis can enable the econometric analysis of large text-based datasets and suggests a psychological channel through which other impacts of climate change may operate. It is worth considering that these results are obtained for the United States, where air conditioner ownership is among the highest in the world. Speculatively, the relationship between ambient temperature and sentiment could be more pronounced in other countries, although cultural differences in temperature sensitivity could mediate this impact. Broadly, this work provides additional evidence that changes in the amenity value of climate are an important component of the overall costs of climate change.

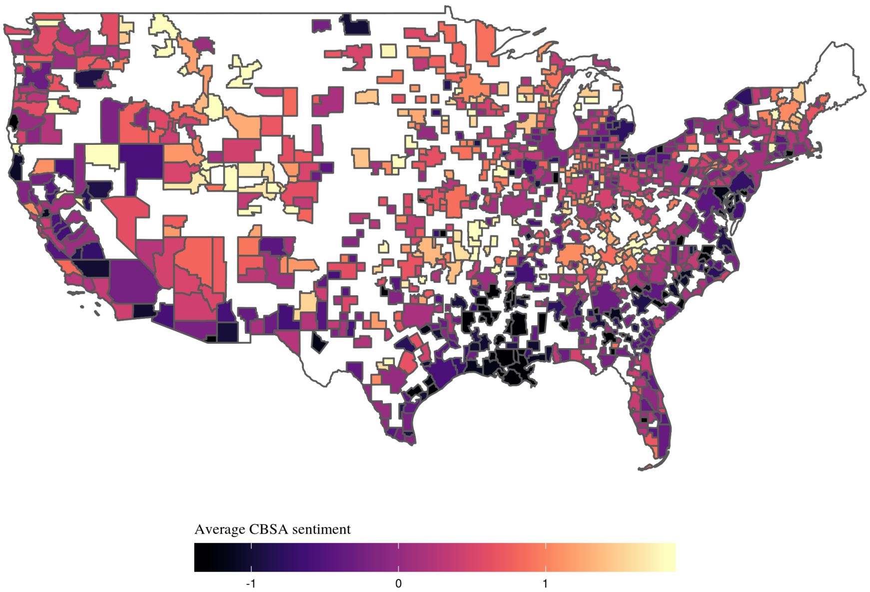
Table 1: Sample characteristics

	Count	Mean	Median	Min	Max
A: Sentiment measures					
AFINN-111	687,252,523	0.5	0.5	-5	4
Hedonometer	1,220,553,949	5.5	5.5	2.8	8.3
LIWC	1,258,416,865	0.3	0.2	-3	4
Vader	1,337,317,534	0.1	0.1	-1	1
B: Weather covariates					
Minimum temperature (C)	1,337,317,534	9.1	10.8	-33.7	31.5
Maximum temperature (C)	1,337,317,534	20.9	23.8	-23.9	47.2
Precipitation (mm)	1,337,317,534	3	0	0	372.7
C: Twitter updates per...					
CBSA	900	1,485,908	190,834	30	92,894,251
User	12,852,098	123	11	1	396,822

Notes: First panel summarizes included measures of expressed sentiment, second panel summarizes weather covariates, and third panel summarizes the number of tweets per CBSA and user.

Figure 1: Tweet density and average sentiment by CBSA





Notes: Top panel: pixel shading represents log (base 10) of count of tweets in sample. Bottom 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	0.78	-0.58
Hedonometer		1.00	0.59	0.73	-0.35
LIWC			1.00	0.77	-0.38
Vader				1.00	-0.40
Profanity					1.00

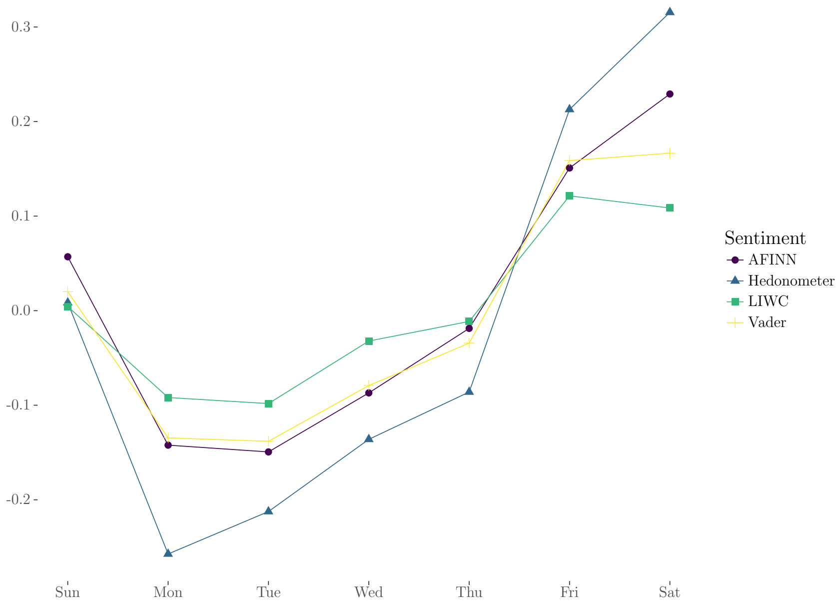
Notes: Pairwise correlations of CBSA-date means of measures of standardized expressed sentiment and profanity measure.

Table 3: Measure correlations: state-year means

	AFINN-111	Hedonometer	LIWC	Vader	Profanity	Gallup
AFINN-111	1.00	0.88	0.96	0.96	-0.94	0.37
Hedonometer		1.00	0.88	0.89	-0.83	0.20
LIWC			1.00	0.97	-0.89	0.32
Vader				1.00	-0.90	0.31
Profanity					1.00	-0.27
Gallup						1.00

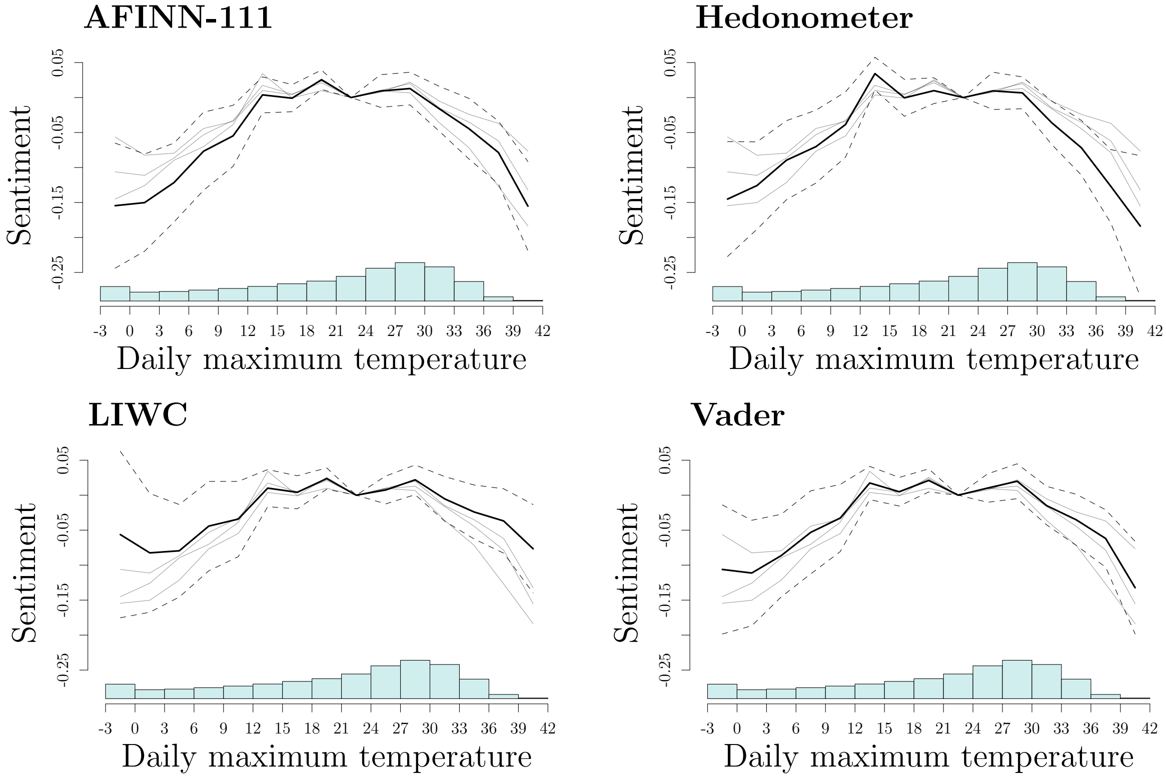
Notes: Pairwise correlations of state annual averages of standardized expressed sentiment and profanity measure and state-level annual Gallup polls of subjective well-being index (Gallup, Inc 2016).

Figure 2: Weekly sentiment



Notes: Lines represent average CBSA-day measures of sentiment by day of week, colored by measure. Excludes holidays.

Figure 3: Effect of temperature on Twitter sentiment



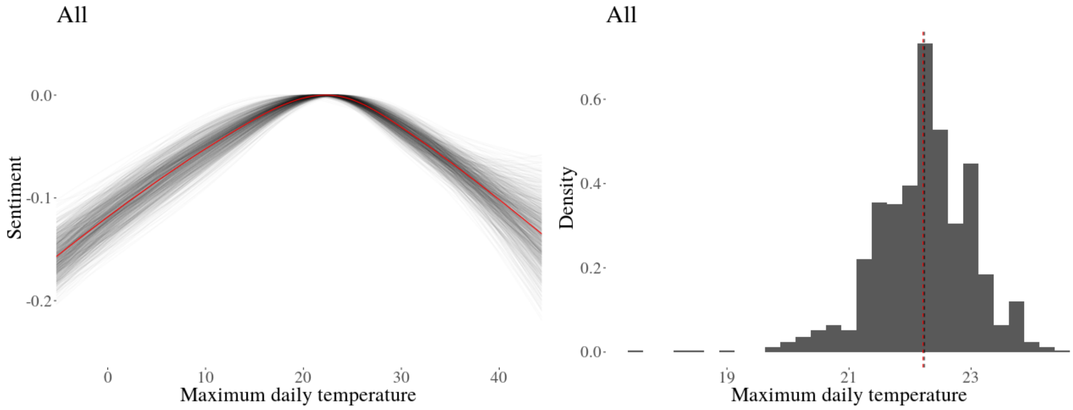
Notes: Panels document the temperature response for each of the standardized four measures of sentiment. Each point estimate represents the difference (measured in standard deviations) in CBSA-day sentiment for the temperature bin T_b relative to 21-24 C, conditional on CBSA, month, and year fixed effects. 95% confidence intervals estimated using cluster robust standard errors by CBSA.

Table 4: Effect of temperature on social media sentiment

	(1)	(2)	(3)	(4)	(5)
Maximum daily temperature T					
$T \leq 0$	-0.10^{**}	-0.11^{**}	-0.15^{***}	-0.02	-0.20^{***}
	(0.04)	(0.05)	(0.04)	(0.03)	(0.05)
$T \in (0, 5]$	-0.10^{***}	-0.11^{***}	-0.15^{***}	-0.04	-0.18^{***}
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
$T \in (5, 10]$	-0.07^{***}	-0.06^{**}	-0.09^{***}	-0.01	-0.10^{***}
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
$T \in (10, 15]$	-0.01	-0.01	-0.02	0.02^{**}	-0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$T \in (15, 20]$	-0.00	0.00	-0.01	0.01^{**}	-0.01^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$T \in (25, 30]$	0.01	0.01	0.01	-0.01	0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$T \in (30, 35]$	-0.03^{**}	-0.03^{*}	-0.02	-0.04^{***}	0.02^{**}
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
$T \in (35, 40]$	-0.06^{**}	-0.06^{***}	-0.05^{**}	-0.07^{**}	0.01
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
$T > 40$	-0.18^{***}	-0.16^{***}	-0.15^{***}	-0.12^{***}	-0.14^{***}
	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)
Fixed effects					
CBSA	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes			
Year	Yes	Yes		Yes	
DOW, Hol		Yes	Yes	Yes	
MOS			Yes		
S×M				Yes	
Date FE					Yes

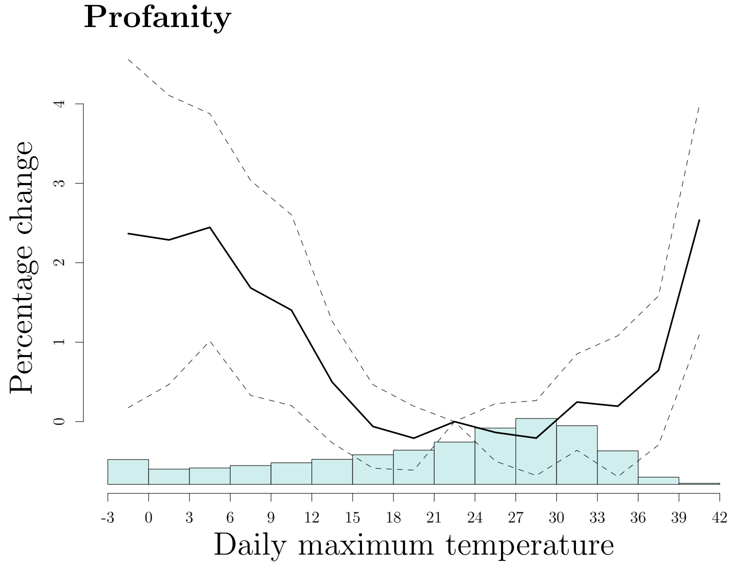
Notes: * p $<$ 0.1, ** p $<$ 0.05, *** p $<$ 0.01. Dependent variable is standardized VADER in a CBSA-day. Coefficients represent the change (in standard deviations) of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. All models include precipitation P, standard errors (in parentheses) clustered by CBSA.

Figure 4: Splined effect



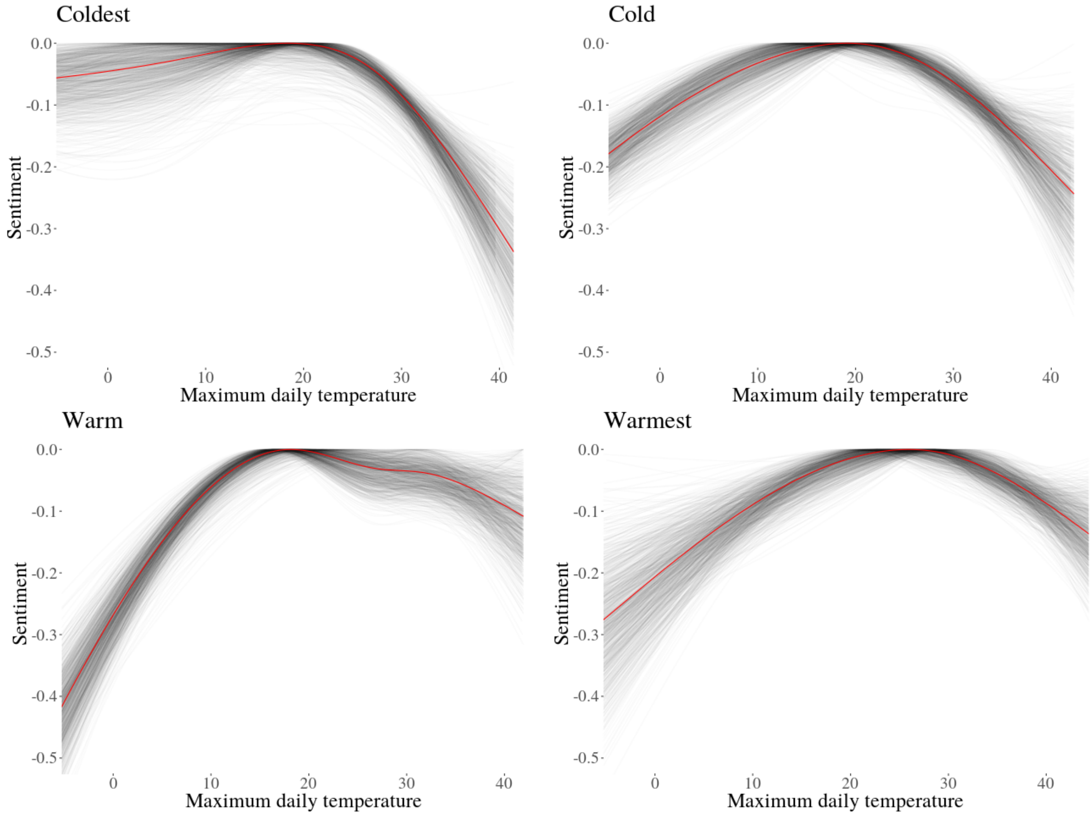
Notes: Left panel: documents the response of the expressed sentiment (measured using VADER and in standard deviations) to temperature using a splined model. Darker red line is the splined response function estimated from the entire dataset, with knots at the 25th, 50th, and 75th percentile of experienced 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.
Right panel: Histogram of estimated preferred maximum daily temperatures for 1,000 bootstrap iterations.

Figure 5: Profanity response to temperature



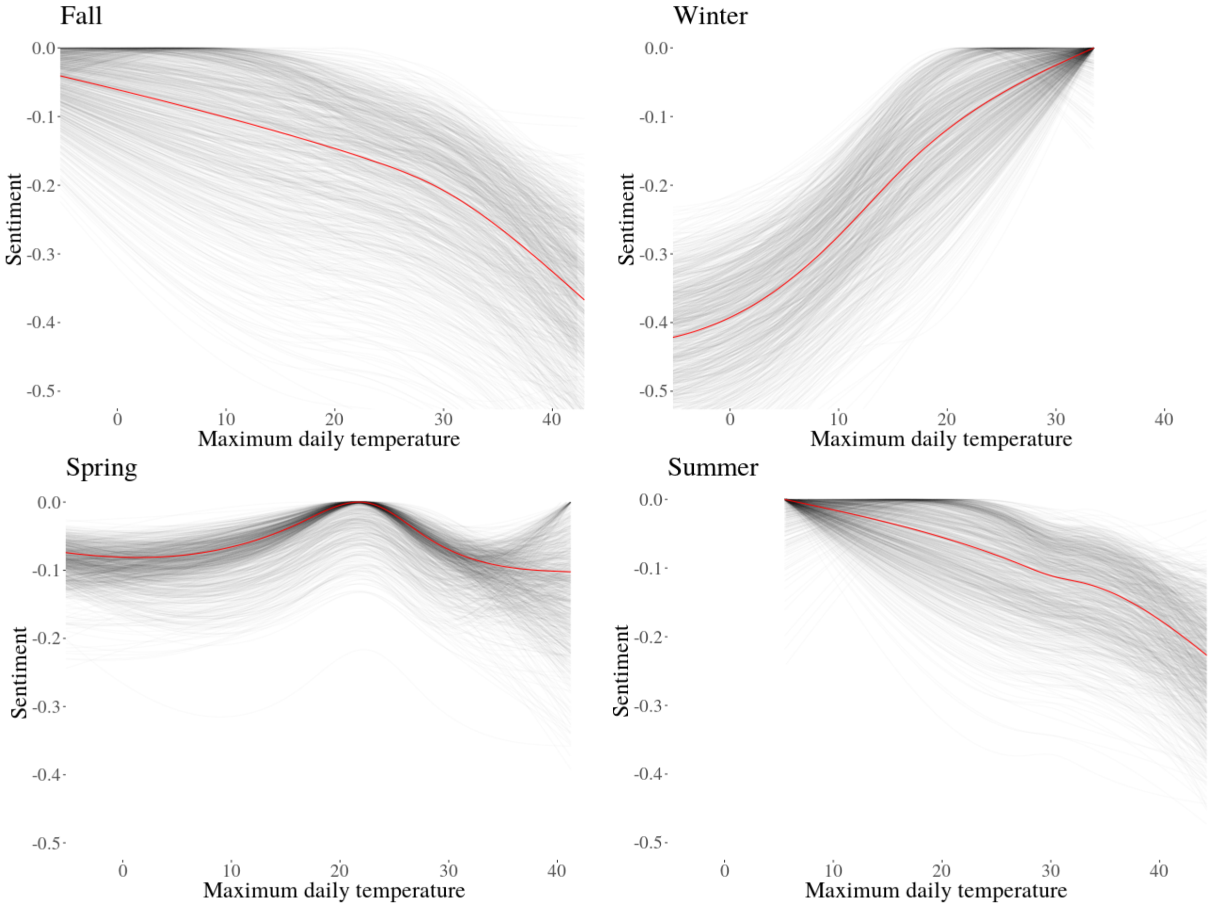
Notes: Figure documents occurrence of profanity in response to temperature, using a list of 300 profanities. Point estimates represent the difference (measured in standard deviations) in CBSA-day profanity occurrence for the temperature bin T_b relative to 21-24 C, conditional on CBSA and state by month of sample fixed effects. 95% confidence intervals estimated using two-way cluster robust standard errors on CBSA and day-of-sample.

Figure 6: Regional responses



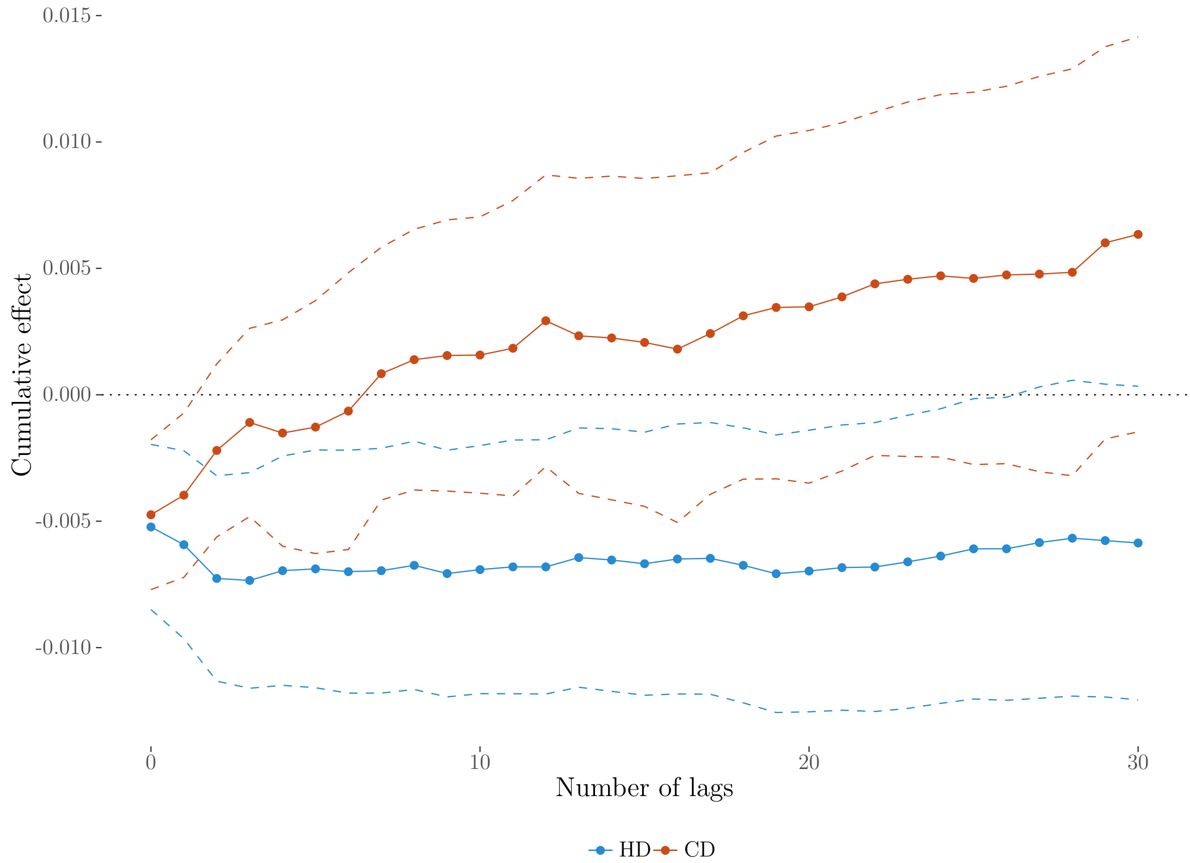
Notes: Panels document the response of the standardized VADER measure to temperature for each of the four regions, where regions are defined by quartiles of average daily maximum temperature during the sampling frame and are labeled, in order of increasing temperature, "Coldest", "Cold", "Warm", and "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.

Figure 7: Seasonal responses



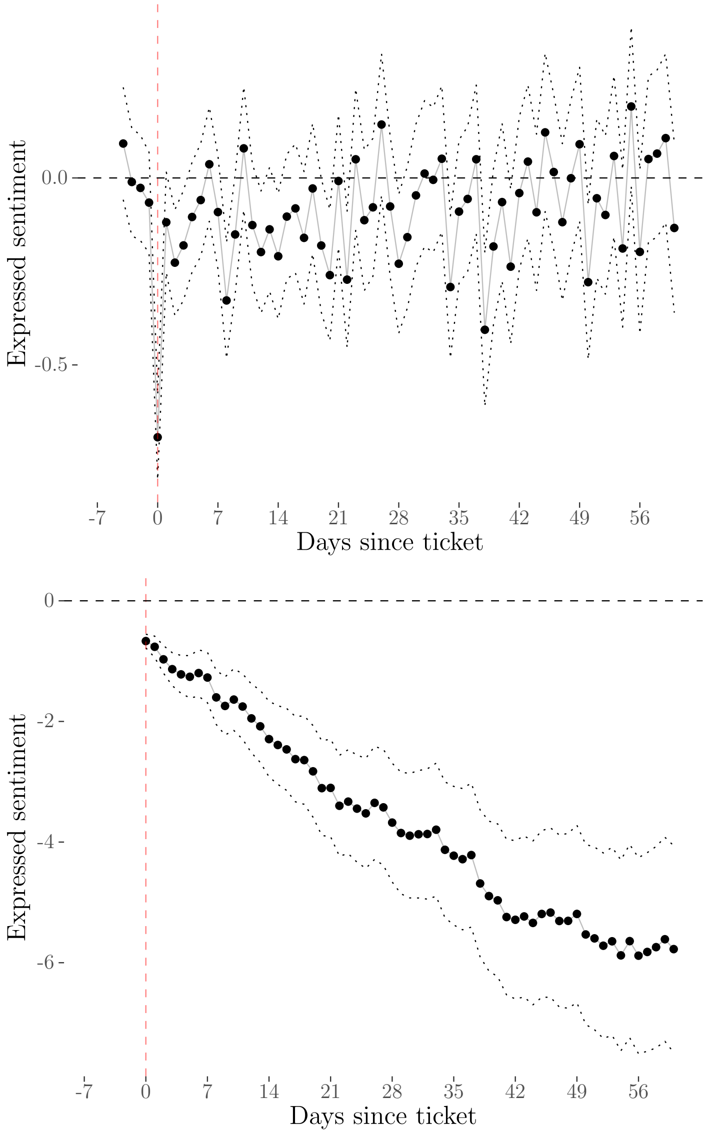
Notes: Panels document the response of the standardized VADER measure to temperature for each of the four regions, where regions are defined by quartiles of average daily maximum temperature during the sampling frame and are labeled, in order of increasing temperature, "Coldest", "Cold", "Warm", and "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.

Figure 8: Cumulative effect



Notes: Dynamic cumulative lagged model estimates, where outcome is standardized VADER sentiment from an increasing number of cold/hot days. Point estimates and standard errors are the sum of coefficients on contemporaneous and lagged measures of degrees below/above 22.7 C, with increasing numbers of lags moving from left to right. 95% confidence intervals estimated using cluster robust standard errors by CBSA.

Figure 9: Ticket 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 30 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.

Table 5: Value of temperature

	(1)	(2)	(3)	(4)	(5)
Maximum daily temperature T					
$T \leq 0$	−1.72	−1.98	−2.58	−0.38	−3.49
$T \in (0, 5]$	−1.80	−1.92	−2.51	−0.65	−3.04
$T \in (5, 10]$	−1.21	−1.10	−1.56	−0.22	−1.80
$T \in (10, 15]$	−0.17	−0.10	−0.27	0.43	−0.81
$T \in (15, 20]$	−0.09	0.01	−0.11	0.25	−0.24
$T \in (25, 30]$	0.14	0.20	0.22	−0.13	0.50
$T \in (30, 35]$	−0.57	−0.45	−0.36	−0.74	0.38
$T \in (35, 40]$	−1.04	−1.04	−0.79	−1.13	0.14
$T > 40$	−3.12	−2.82	−2.62	−2.11	−2.39
Fixed effects					
CBSA	Yes	Yes	Yes	Yes	Yes
Month	Yes	Yes			
Year	Yes	Yes		Yes	
DOW, Hol		Yes	Yes	Yes	
MOS			Yes		
S×M				Yes	
Date FE					Yes

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Reproduction of Table 4, with coefficients multiplied by valuation of a unit SD change in standardized VADER sentiment. Value is obtained by dividing the median observed ticket cost ($100) by the total cumulative sentiment loss from a parking or speeding ticket (5.77 SD) in Figure 9.

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.

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, Solomon M. 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.

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*, edited by J H Burns, H L A Hart, Jeremy Bentham, J H Burns, and H L A Hart. Oxford University Press.

Bradley, Margaret Mm, and Pj Peter J Lang. 1999. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings.* Technical report. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

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

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

Burke, Marshall, Solomon M Hsiang, and Edward Miguel. 2015b. "Global non-linear effect of temperature on economic production". *Nature* 527:235–239.

Burke, Marshall, and David Lobell. 2012. *Climate Change and Food Security.*

Carleton, T. A., and S. M. Hsiang. 2016. "Social and economic impacts of climate". *Science* 353 (6304).

Center for International Earth Science Information Network Columbia University, United Nations Food and Agriculture Programme, and Centro Internacional de Agricultura Tropical. 2005. *Gridded Population of the World.* Palisades, NY.

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

Cragg, Michael, and Matthew Kahn. 1997. "New estimates of climate demand: evidence from location choice". *Journal of Urban Economics* 42 (2): 261–284.

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): 740–798.

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

Deschnes, 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*, number 627.

Gallup, Inc. 2014. "2014 State Well-Being Rankings". http://cdn2.hubspot.net/hub/162029/file-2513997715-pdf/Well-Being_Index/2014_Data/Gallup-Healthways_State_of_American_Well-Being_2014_State_Rankings.pdf.

— . 2015. "2015 State Well-Being Rankings". http://info.healthways.com/hubfs/Well-Being_Index/2015_Data/Gallup-Healthways_State_of_American_Well-Being_2015_State_Rankings.pdf.

— . 2016. "2016 State Well-Being Rankings". http://info.healthways.com/hubfs/Gallup-Healthways%20State%20of%20American%20Well-Being_2016%20State%20Rankings%20vFINAL.pdf.

Gaure, Simen. 2013. “lfe: Linear group fixed effects”. User documentation of the ’lfe’ package, *The R Journal* 5, number 2 (): 104–117.

Gilbert, C J Hutto Eric. 2014. “Vader: A parsimonious rule-based model for sentiment analysis of social media text”. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf.*

Graff Zivin, Joshua, Solomon Hsiang, and Matthew Neidell. 2015. “Temperature and Human Capital in the Short-and Long-Run”. *NBER Working Paper.*

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 *ICWSM.*

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

Houser, Trevor, et al. 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.

Hsiang, Solomon M. 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.

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

Kahn, Jeffrey H., Rene 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.

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.

Levinson, Arik. 2012. "Valuing public goods using happiness data: The case of air quality". *Journal of Public Economics* 96, number 910 (): 869–880.

Livneh, Ben, et al. 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.

Loewenstein, George, and Jennifer S. Lerner. 2003. *The role of affect in decision making.*

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, number 4 (): 1093–1113.

Morstatter, Fred, Jrgen Pfeffer, Huan Liu, and K Carley. 2013. "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose". *arXiv.*

Nielsen, Finn rup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs". *arXiv:1103.2903 [cs].*

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

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, Cindy K Chung, Molly Ireland, Amy Gonzales, and Roger J Booth. 2007. *The Development and Psychometric Properties of LIWC2007.* Austin, TX.

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.

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.

Schwarz, Norbert, and Gerald Clore. 1982. "Mood, misattribution, and judgments of well-being.pdf". *Journal of Personality and Social Psychology* 45 (3): 513–523.

Sinha, Paramita, and Maureen Cropper. 2015. "Household location decisions and the value of climate amenities". *NBER Working Paper Series*, number January: 43.

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

Stock, James H, and Mark W Watson. 2007. *Econometrics.* Addison Wesley.

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.

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

Twitter. 2015. *Company — About.* Visited on 05/01/2015.

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

Welsch, Heinz, and Jan Khling. 2009. "Using happiness data for environmental valuation: Issues and applications". *Journal of Economic Surveys* 23, number 2 (): 385–406.

ONLINE APPENDIX

A. Measures of 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. (2007). 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, Tobin, Massey, and Anderson (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 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".[13] The VADER measure has been validated against a variety of ground-truth data and found to outperform other measures (Gilbert 2014).

The following tables include examples of the word lists used to construct the sentiment scores.

[Tables A6, A7, A8, A9, and A10]

B. Empirical checks

In this section I document a series of checks intended to test the robustness of the result to different sample selection criteria and model specifications.

[13] The VADER Github page is a useful resource for more details on its construction: https://github.com/cjhutto/vaderSentiment.

B.1. 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, Hsiang, Schlenker, and Sobel (2013). Because the weather dataset I use includes precipitation as well, I include both temperature and precipitation in Equation (2) 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, Butalid, Penke, and Van Aken 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 A11]

Table A11 tabulates the regression results from a model that adds diurnal range, relative humidity, wind speed, air pressure, and the percent of the day that was reported as overcast. For comparison, column (1) reports the baseline results from column (4) of Table 4. Column (2) limits the sample to those observations with the additional weather covariates to ensure a fair comparison. This exclusion has a negligible effect on the reported estimates. Finally, column (3) reports the model results when additional weather covariates are accounted for. 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.

B.2. 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 A10. [Figure A10]

B.3. Individual tweet-level analysis

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 3, 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 user fixed effects. To do so, I estimate the following model:

$$E_{id} = \sum_{b \neq 20\text{-}25}^{B} \beta_b T_{cd}^{b} + \phi_i + \phi_m + \phi y + \varepsilon_{id} \tag{5}$$

This model replaces CBSA fixed effects with user fixed effects ϕ_i in equation Equation (2). The model requires the use of the disaggregated sample of tweets in my dataset; for computational reasons, I use a 20% subsample of the users with more than 100 tweets in my sample to estimate this model.

[Figure A12]

To compare the results between the user fixed model and the baseline model, I overlay the estimates from each model in Figure A12. I find qualitatively similar results for the measures, although the estimates for higher temperatures are attenuated in the individual fixed effects model relative to the baseline model. It is possible that this is evidence of some compositional sorting at higher temperatures, but more likely the result of measurement error driven by using a sparser source of variation. The negative response to cold temperature is nearly identical between models, suggesting that the source of the differential is heterogeneous in temperature.

C. Parking and speeding ticket dataset construction

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.

D. Climate projections

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 combine projected changes in climate with the estimates of the response of expressed sentiment to temperature. The nature of the projection exercise can be described mathematically as follows:

$$\int^{T} f(t)\Delta g(t)dt \tag{6}$$

where $f(\cdot)$ represents the damage function (valued in \$), such as the one estimated in Figure 3, and $\Delta g(\cdot)$ is the change in the distribution of climate. By integrating the product of f and g over the range of temperature T I obtain the total damages. Empirically, I estimate the shape of $f(\cdot)$ and combine climate and weather data to obtain $\Delta g(\cdot)$ in order to numerically approximate Equation (6).

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 4 and multiplied by the per-SD valuation from Section 6. The adaptive exercise uses the regionally-specific damage functions estimated in Figure 6 to project changes due to adaptation, also multiplied by the per-SD value. 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 warms an area to the extent that it moves from quartile 3 (the third-warmest region) into quartile 4 (the warmest region), then its damages are estimated using the quartile 4 damage function.

$g(\cdot)$ is estimated using the ensemble average from the output of 20 downscaled climate models[14], 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, Hsiang, Schlenker, and Sobel (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 A14 documents the evolution of per-person annual damages over time, averaged over CBSAs and presented separately for RCP4.5 and RCP8.5, two different climate

[14] 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.

forcing scenarios of intermediate and high warming, respectively (IPCC 2014) and using both the baseline and adaptive methods described above. I estimate annual damages of increasing over time across all scenarios, with damages from RCP8.5 exceeding damages from RCP4.5. I find that the adaptive approach estimates larger damages than the baseline approach. Intuitively, it would seem that adaptation should mitigate the impact of climate change, but this does not occur here. The key element is that the damages from cold weather for the coldest areas of the United States are substantially smaller (see Figure 4) than they are for warmer areas. As a result, warming from climate change does not benefit those areas as much as it does in the baseline scenario. One interpretation is that since colder areas have already adapted to colder temperatures, climate change is less of an improvement than it would have been if those areas had not already adapted. This effect declines in importance as temperatures increase: the adaptive and baseline scenarios converge by end of century for RCP8.5.

[Figure A14]

The top panel in Figure A15 maps end of century damages under RCP8.5 by CBSA under the baseline scenario, documenting clear north-south heterogeneity in the extent of amenity costs due to climate change. The bottom panel does the same using the adaptive scenario.

[Figure A15]

Table A6: 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 A7: 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 A8: 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 A9: Sample of tweets with sentiment scores

Tweet (first 50 characters)	(1)	(2)	(3)	(4)
Yes! Chicago today	1	6.27	0	0.50
What a great way to end 2014 @ SNOW IN TUCSON WHAT	3	5.46	1	0.99
@GottliebShow does he get that money plus whatever	0	5.52	1	-0.70
This was hella worth it	2	5.23	0	-0.25
Happy New Years! Had the best time @WattsBarChurch	3	5.83	3	0.82
@jarlenykillz *Turns straight*	1	5.63	0	0.00
@dandandempsey haven't cut mine in 4	-1	4.76	-1	-0.26
Not a very nice chef in Jcpenny http://t.co/2Gsm7	3	5.72	0	0.44
Go follow my boy. @oLewisss he's a sick GFX artist	-2	5.34	0	-0.96
@love_the_Ks so awesome! If you need any beer rec	3	5.72	1	0.94
happy to be this http://t.co/XMebBnIVGF	3	6	1	0.42
@g_jackson24 I better see you nigggaaaaa	2	6.3	1	0.23
Feeling and love are both over rated they are best	2	5.69	2	-0.01
@wasabisauce lets waste time chasing squirrels	-1	4.68	0	0.47
someone please get me out of my house rn	1	5.78	0	-0.94
As long as my momma told me Happy New Year Idc bou	1	5.48	1	-0.61
Tonight is going to be good	3	5.77	1	-0.71
makes me happy seeing close friends having a good	3	6.36	2	-0.55
All my friends are drunk rn	-2	5.87	0	-0.89
MSU for the win	4	6.11	1	0.77
I want chipotle	1	5.81	0	-0.95
@smarcher_ I have light canceling curtains up but	-3	5.34	-1	-0.90
@nnakedd Lmao dude I got you	4	5.87	1	-0.44
Happy new year	3	6.83	1	0.76
First morning of 2015, I check my phoneI go to sn	0	5.83	2	0.56
I wish you cared.	1	6.48	1	-0.98
i hate pitbull	-3	4.13	-1	-0.96
But not really feeling it	1	4.93	0	-0.88

Notes: Tweets selected at random. Scores displayed reflect output of sentiment analysis algorithms used in paper.

Table A10: 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

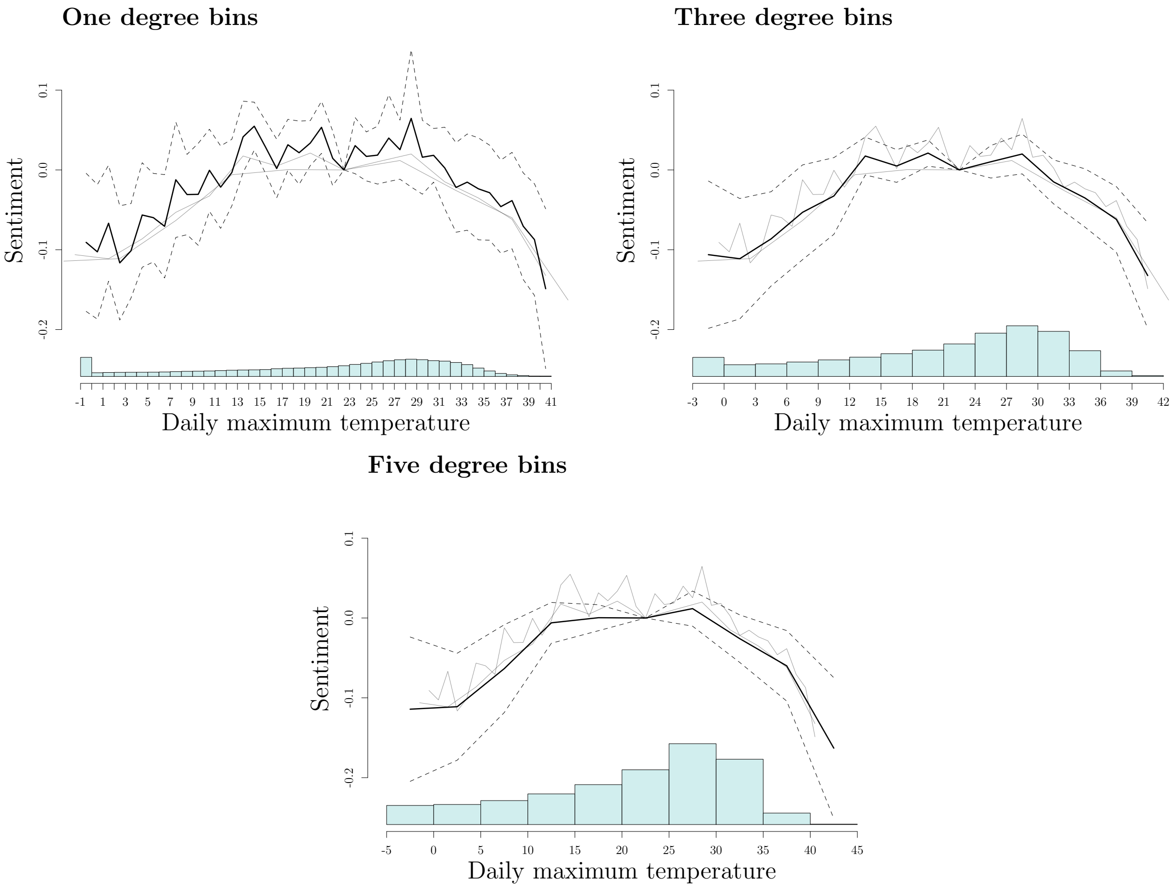
Notes: Author construction.

Table A11: Additional weather variables

	(1)	(2)	(3)	(4)	(5)
Maximum daily temperature T					
$T \leq 0$	-0.15^{***}	-0.13^{**}	-0.01	-0.16^{***}	-0.25^{***}
	(0.04)	(0.05)	(0.03)	(0.05)	(0.05)
$T \in (0, 5]$	-0.14^{***}	-0.11^{***}	-0.01	-0.14^{***}	-0.21^{***}
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
$T \in (5, 10]$	-0.09^{***}	-0.06^{*}	0.01	-0.08^{***}	-0.12^{***}
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
$T \in (10, 15]$	-0.02	-0.00	0.04^{***}	-0.01	-0.06^{***}
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
$T \in (15, 20]$	-0.01	0.00	0.03^{***}	-0.00	-0.02^{**}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$T \in (25, 30]$	0.01	0.00	-0.02	0.01	0.03^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$T \in (30, 35]$	-0.03^{**}	-0.04^{**}	-0.07^{***}	-0.04^{**}	0.02^{**}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
$T \in (35, 40]$	-0.06^{**}	-0.08^{***}	-0.10^{***}	-0.06^{**}	0.02
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
$T > 40$	-0.17^{***}	-0.17^{***}	-0.16^{***}	-0.15^{***}	-0.11^{***}
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Precipitation	-0.00^{***}	-0.00^{***}	-0.00^{***}	-0.00^{***}	-0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Diurnal range	-0.01^{***}	-0.00	0.00^{*}	-0.00	-0.01^{***}
	(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)	(0.00)	(0.00)
Wind Speed	-0.00^{**}	-0.00^{*}	-0.00	-0.00	-0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Station Pressure	0.04^{**}	-0.02	-0.03	-0.00	0.05^{*}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Overcast	-0.11^{***}	-0.11^{***}	-0.09^{***}	-0.13^{***}	-0.12^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Fixed effects					
CBSA	Yes	Yes	Yes	Yes	Yes
Month, Year	Yes	Yes			
MOS				Yes	
DOW, Hol		Yes	Yes	Yes	
S×M, S×Y			Yes		
Date FE					Yes

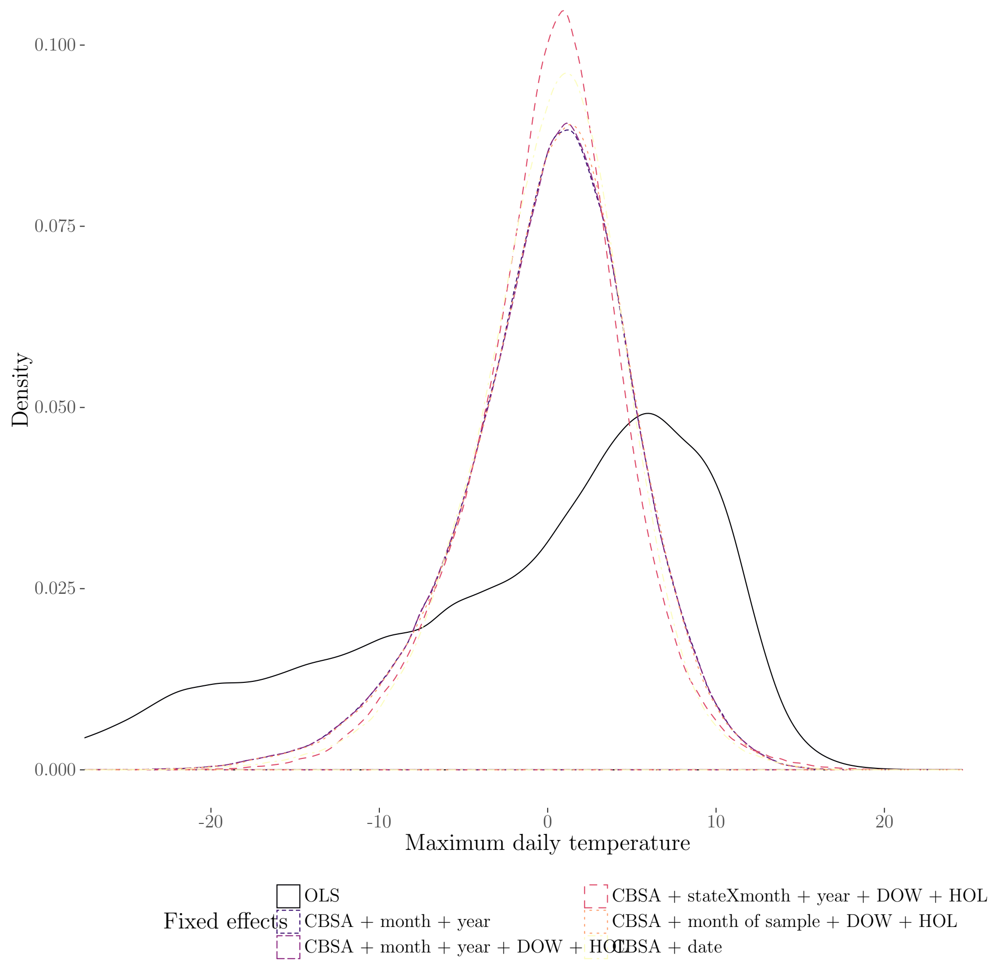
Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Dependent variable is sentiment in a CBSA-day. Coefficients represent the change in standard deviations of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20, 25)$, the omitted category. 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-month of sample and date.

Figure A10: Bin width



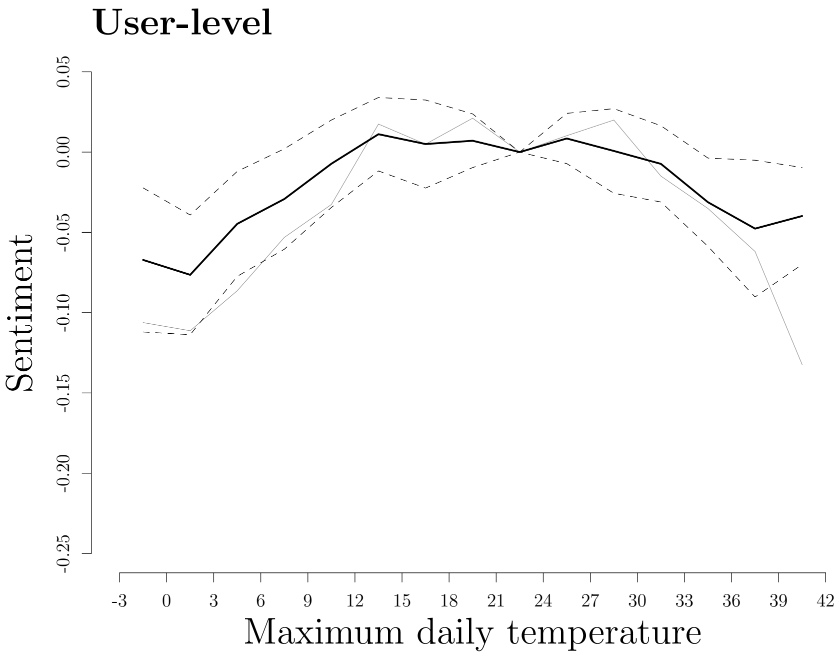
Notes: Comparison of the sentiment response to temperature across bin widths of 1, 3, and 5C. Dark solid and dashed lines indicate model estimates and standard errors for the given bin width, gray lines illustrate estimates for other bin widths for comparison. All models include precipitation P, CBSA,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

Figure A11: Residual variation



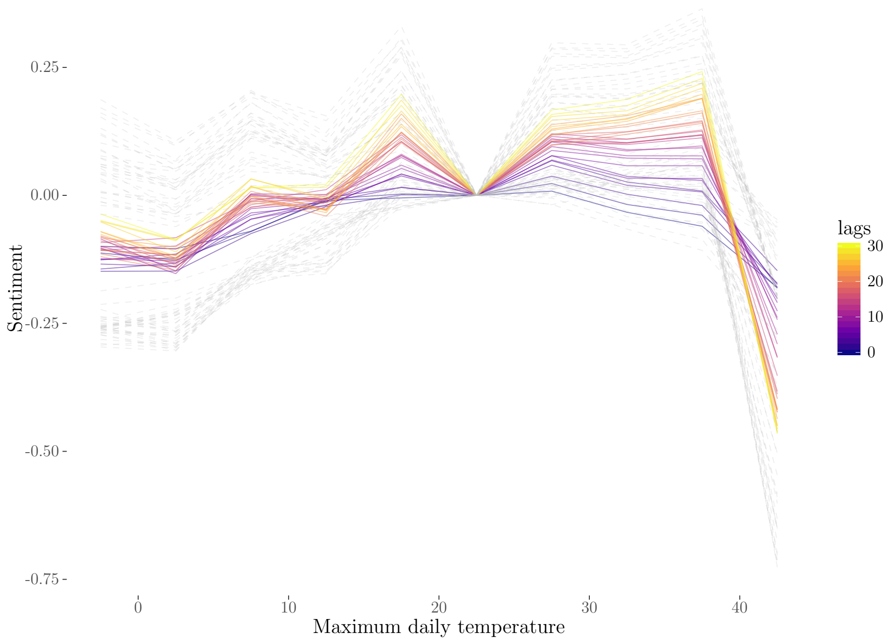
Notes: Density estimates of residual variation for columns of fixed effects in Table 4. Estimates are constructed by regressing maximum daily temperature on the given set of fixed of effects and plotting the density of the residuals from that regression. Each color and linetype represents a different combination of fixed effects.

Figure A12: CBSA-level and user-level model comparison



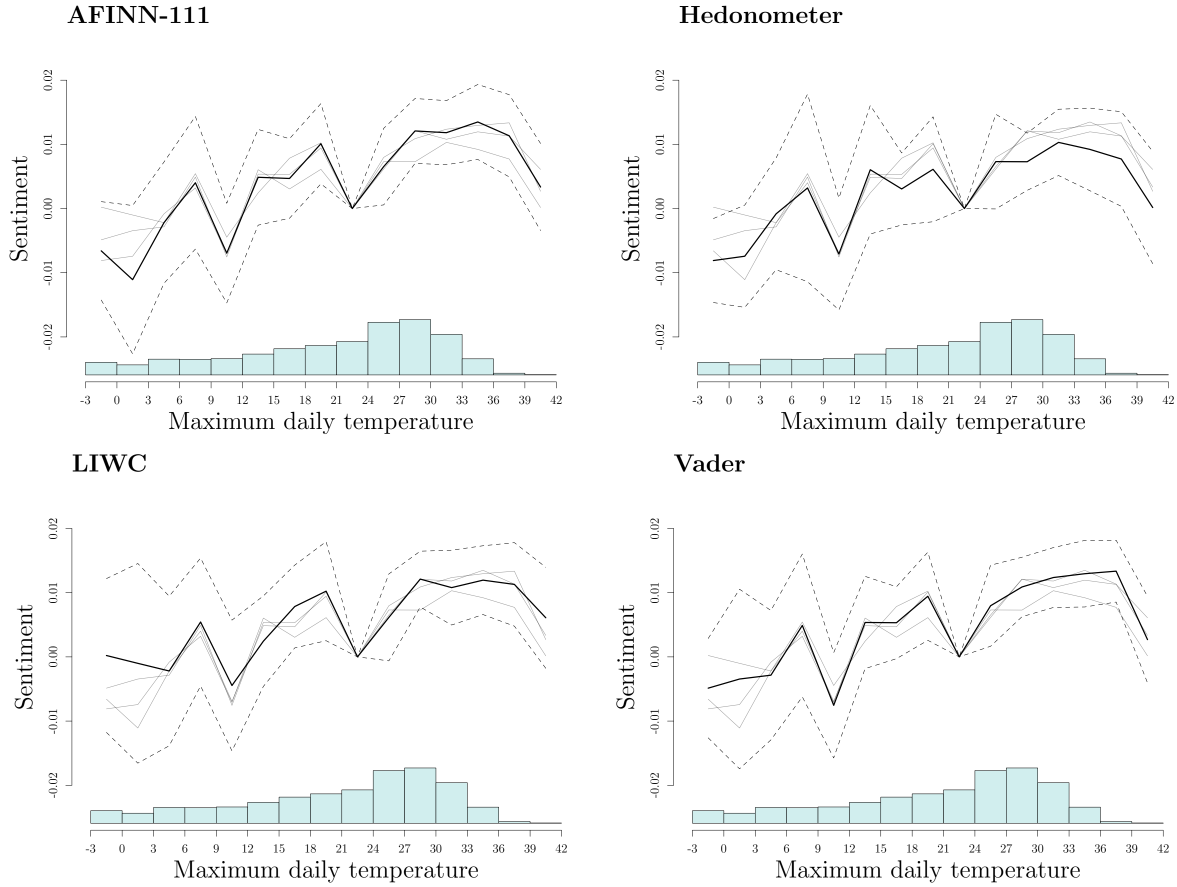
Notes: Plots compare the hedonic response to temperature for models using CBSA and user-level estimates. Dark line indicates response and confidence interval of estimates from binned, user-level model, which replaces CBSA fixed effects in Equation (2) with user fixed effects and is estimate on tweets from a 20% sample of tweets from users with more than 100 overall tweets in sample. The light gray line is the baseline specification using CBSA-day averages as observations, plotted in top left panel of Figure 3. 95% confidence intervals estimate using standard errors clustered by CBSA.

Table A12: Dynamic cumulative lag model (binned model)



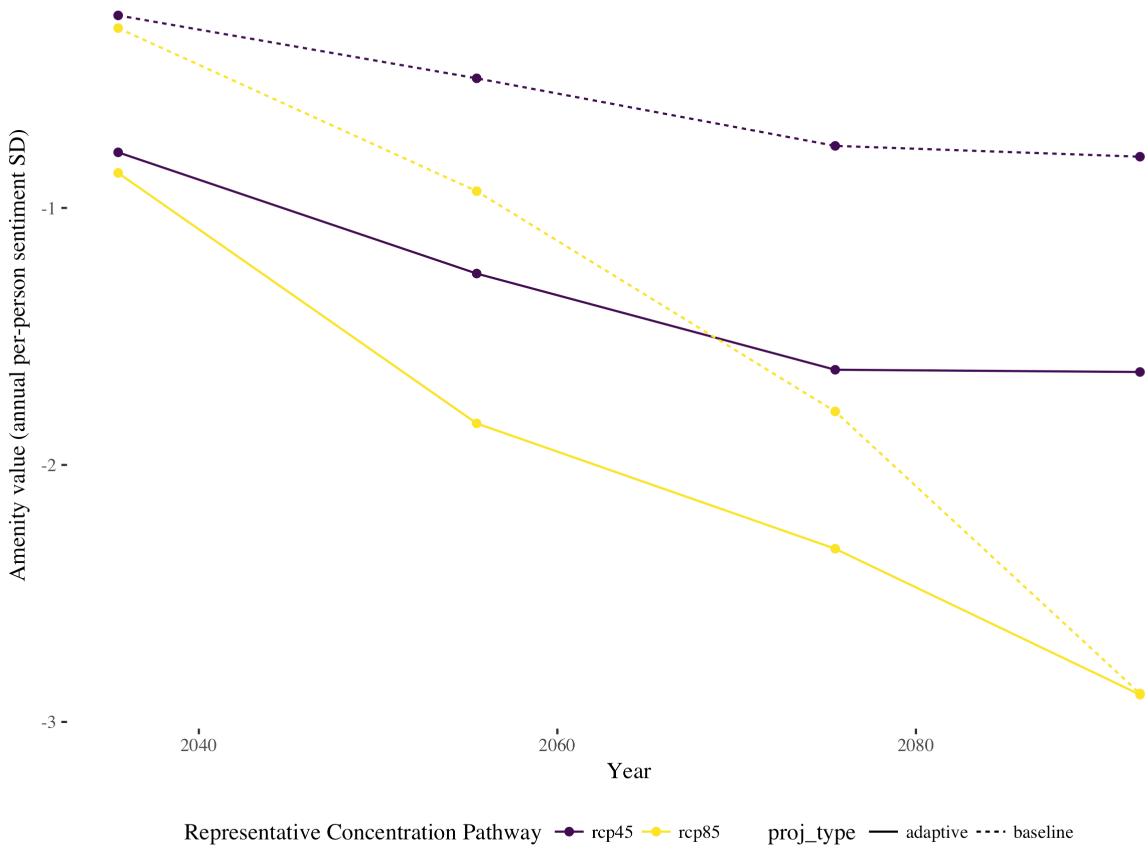
Notes: Dynamic cumulative lag model estimated using five degree C bins. Outcome variable is standardized measures of expressed sentiment. Each line represents one set of summed bin coefficient estimates for M lags, where $M \in 0, ..., 30$ is the number of lags of binned temperature. Increasingly light coloring indicates more lags included. In addition to contemporaneous and lagged daily maximum temperatures, regressions include precipitation, CBSA, month, and year fixed effects. Standard errors clustered by CBSA.

Figure A13: Monthly effect of temperature on Twitter sentiment



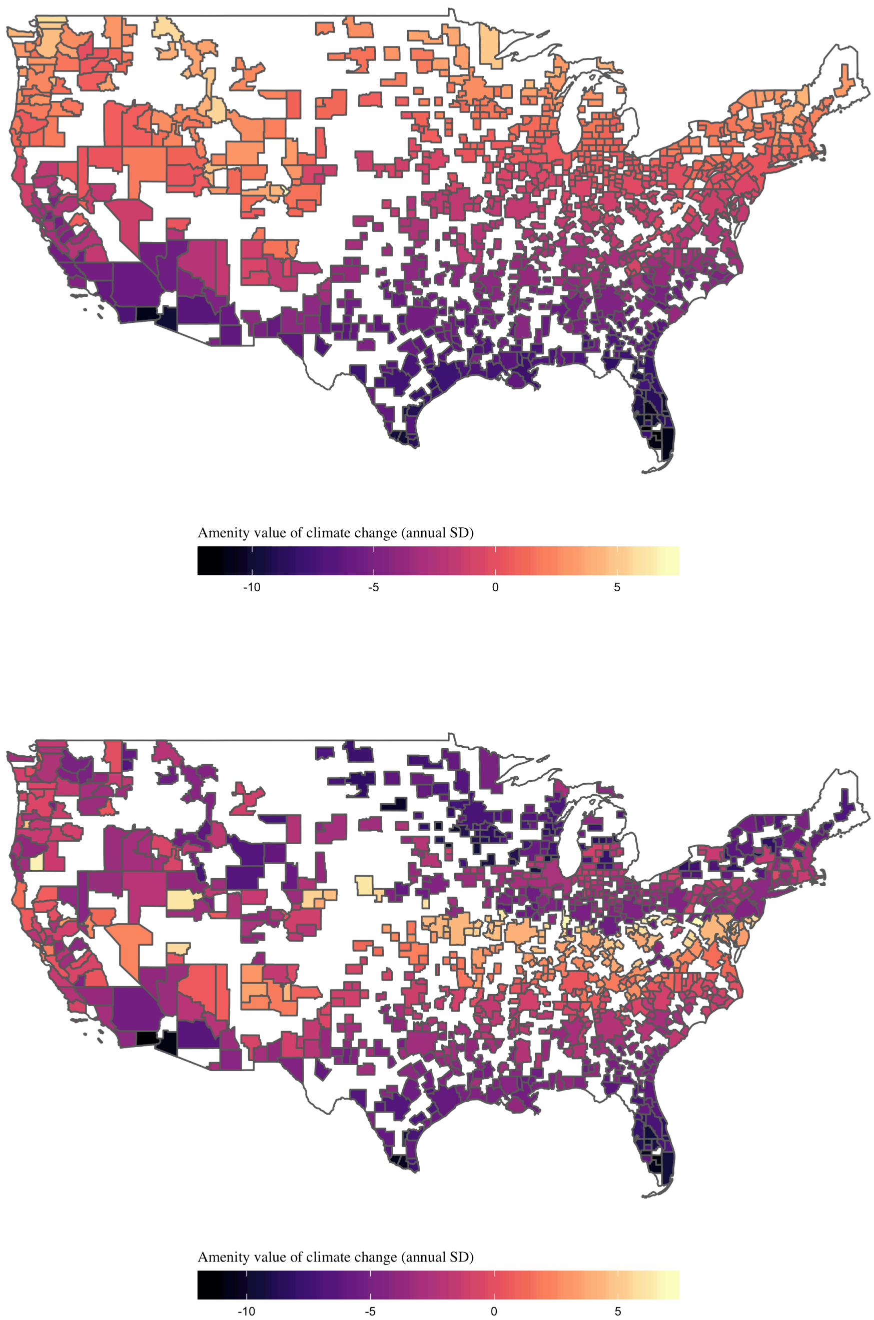
Notes: Monthly estimate of temperature on expressed sentiment. Outcome variable is standardized measures of expressed sentiment. Binned temperature variables indicate the number of days in a month in which the maximum daily temperature fell within the given range. Regressions also include precipitation, CBSA, month, and year fixed effects. Standard errors clustered by CBSA.

Figure A14: Projections of changes in amenity value over time



Notes: Projections of average change in amenity value over time, where average is the weighted average using total number of tweets per CBSA as the weights. 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.

Figure A15: End of century projections of changes in amenity values (RCP8.5)



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