Temperature and temperament

Patrick Baylis*

February 24, 2017

Abstract

Can social media reveal preferences for environmental goods? Without an explicit market for weather, prior work has relied on variation in housing prices or survey data to identify preferences for climate and to value the resulting amenity impact of climate change. This paper demonstrates a novel approach to environmental valuation, combining more than a billion Twitter updates with natural language processing algorithms to construct a rich panel dataset of daily, county-level sentiment. This dataset allows me to estimate a fixed effect model that accounts for unobserved cross-sectional and temporal variation and allows for precise estimates of non-monotonic effects. I document statistically significant declines in expressed sentiment above and below 20-25 C (68-77 F), seasonal variation in climate preferences, and demonstrate a preliminary method to value these changes monetarily. Combining this method with climate projections under the highest emissions scenario, I estimate per-person annual amenity damages of more than $400 by end of century due to temperature rise.

JEL Codes: Q51, Q54

*Stanford University. Center on Food Security and the Environment; Y2E2 349, Stanford, CA 94301; E-mail: baylis@stanford.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, and the Heartland Workshop at Illinois. Errors are entirely my own. This paper was previously circulated under the title "Temperature and Temperament: Evidence from a billion tweets."

1 Introduction

As the possibility of substantial changes in Earth's climate becomes more certain, environmental 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). Similar work estimates the cost of changes in income, health, agriculture, civil conflict, natural disasters, and other economic outcomes (Carleton and Hsiang 2016).

A smaller 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 to elicit the underlying demand curve or preference set 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 choice models to estimate individuals' willingness-to-pay for different climate variables. 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, these costs are identified using cross-sectional differences in climate and are therefore reliant on important assumptions about unobservable variation in climate preferences.

A small literature instead employs subjective well-being surveys to answer this question, comparing the impacts of climate on subjective well-being with the relationship between income and well-being to value climate preferences (Rehdanz and Maddison 2005; Levinson 2012). With enough surveys, it is possible (though usually prohibitively expensive) for this approach to solve the endogeneity problem of the hedonic choice models, but the valuation task relies on the strong assumption that the income is an exogenous driver of subjective well-being (Mackerron 2012).

This paper demonstrates a new, cost-effective method to estimate preferences over public goods that addresses both the identification and power concerns described above: I construct a spatially and temporally rich dataset on daily expressed sentiment, or emotional state, and estimate the relationship between sentiment and outdoor temperature.

Before proceeding, I sketch out a conceptual model that clarifies why sentiment might provide useful insights into the sign and possibly the strength of preferences for non-market goods. The model also helps illuminate some of the concerns with this kind of data and guides the statistical model I choose. I then construct the dataset using a geographically and temporally dense collection of more than a billion geocoded social media updates from the online platform Twitter. I code each tweet using a set of natural language processing (NLP) algorithms designed to extract sentiment, or emotional state, from natural language. The geographic and temporal density of my dataset allows me to resolve identification concerns by accounting for controlling for a wide range of potentially correlated spatial or temporal unobservables using county and state by month of sample fixed effects.

Because of the uncertainty inherent in estimating underlying emotional state from language, I compile four separate measures of sentiment using word lists either directly from previous work in NLP or derived using a machine-learning techniques. I validate these measures by demonstrating that they generally agree with one another, that state-level averages correlate with polling data on subjective well-being, and that the four measures all show similar weekly patterns of fluctuation.

Using geographic information attached to the Twitter updates (known as "tweets"), I match these measures of sentiment to daily weather conditions at the precise location of the user. My identifying assumption is that temperature realizations are as good as random after accounting for spatial and seasonal fixed effects. Allowing temperature to enter the econometric model flexibly, I find strong evidence of a sharp decline in sentiment for both temperature extremes, with the peak sentiment occurring at 20-25 C (68-77 F) on average. The difference in emotional state between 20-25 C and above 40 C (104 F) is statistically significant and comparable in size to half the average difference in sentiment between Sundays and Mondays.

I extend the baseline results to further explore the data and to test for potential sources of bias. First, I demonstrate consistent effects in both direction and standardized magnitude across all measures of sentiment, indicating that the results are not driven by measure design. I document heterogeneity in the effects by season and find patterns consistent with prior work. As a means of understanding the mechanism by which sentiment responds to temperature, I estimate the relationship between lexical aggression and temperature, finding an increase in profanity during hot temperatures.

Next, I demonstrate a preliminary method to value shifts in sentiment by using a subsample of users who received and tweeted about parking or speeding tickets. I document the sentiment change induced by the receipt of the ticket, and then combine it with the median ticket value in my sample to value changes in expressed sentiment. Using this valuation and downscaled climate projection data, I project end-of-century per-person annual amenity costs of climate change to be greater than $400 under the highest emissions scenario, a large but somewhat smaller impact than that estimated by the hedonic choice literature.

In a companion paper, we use data from both Facebook and Twitter to examine the relationship between expressed sentiment and contemporaneous weather (Baylis et al. 2017). Our work in that paper complements and informs the approach I use here: using a different set of sentiment analysis methods, units of analysis, and statistical models, we identify similar response functions between expressed sentiment and temperature. Despite these differences, in both papers we identify a similar central result: that expressed sentiment declines in both hot and cold temperatures. While Baylis et al. (2017) examines the psychological components of this result, in this paper I extend the main result document seasonal shifts in this relationship, to estimate and value preferences for climate, and to project damages based on those estimates.

The paper proceeds as follows: section 2 describes related work and sketches a conceptual framework to motivate the empirical work. Section 3 describes the data and NLP approaches I use, while section 4 lays out the empirical approach and identifying assumptions. Section 5 reports the baseline results, section 6 documents extensions, section 7 values the impacts documented above, section 8 projects future damages, and section 9 concludes.

2 Background

Economists have studied the economic impacts of climate change for more than two decades (Nordhaus 1991; Cline 1992), but the recent availability of panel datasets and advanced econometric techniques 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, scholars have estimated changes in these outcomes resulting from plausibly exogenous historical variation in temperature (Dell, Jones, and Olken 2014); Hsiang (2016) derives the required assumptions for extrapolating effects estimated using variation in weather to those predicted from changes climate.

This scholarship 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

[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).

of CO_2 equivalent had been incorporated into 79 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 hurricanes (Hsiang 2010). Farmers could experience large decreases in yields (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 the 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 dramatic natural disasters or acute economic crises, most of the world will simply experience a steady increase in average temperatures (IPCC 2014). Prior work has projected the impact of these gradual changes on income (Deryugina and Hsiang 2014), crime (Ranson 2014), mortality (Deschênes and Greenstone 2011), and other outcomes. This paper focuses instead on the welfare cost of changes in amenity values resulting from rising outdoor temperatures.

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 non-market 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 understood as a public good: it is non-rival and non-excludable, 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 possible valuation approach: recent work by Sinha and Cropper (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 model estimates. However, since historical climate changes 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.

A related approach 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, a small literature attempts to value environmental goods using self-reported happiness data (Welsch and Kühling 2009). For example, Rehdanz and Maddison (2005) estimate the relationship between climate and self-reported happiness, and include a valuation method based on country-level GDP. Levinson (2012) conducts a similar exercise to estimate WTP to avoid pollution using happiness data, but includes weather as a covariate. However, most of these studies rely on cross-sectional variation in income to value induced changes in self-reported happiness, an approach that may induce bias: income is certainly correlated with happiness, but the causal relationship between the two is not well-established (Mackerron 2012).

The method I propose in this paper mitigates the problem of unobserved correlates over time and space, allows for flexible estimation of non-monotonic effects, and allows for a preliminary method of environmental valuation that makes use of exogenous wealth shocks from parking or speeding tickets. To do so, I employ social-media measured sentiment as a proxy for emotional state, which I argue captures a form of "experienced utility" (Kahneman and Sugden 2005). To better understand this approach, I present a model that characterizes sentiment as the sum of recently experienced utility. Formally,

$$S_t(c) = \sum_{t=t_1}^{\infty} (1-\delta)^{t-t_1} [u(c_t) - u(c_{t-1})]$$

, where S_t is sentiment in time t, c, δ is the discount rate, and u is an instantaneous utility function. This formulation holds that 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 emotional state 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 a good c_t. Assuming that instantaneous 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.

The model suggests two key facts related to this approach. First, because human sentiment has a recency bias, it is only useful for mapping preferences over stimuli that vary rapidly and exogenously over time: slow changes in household income, for example, are unlikely to be strongly reflected in sentiment.

Second, because sentiment is composed of recent changes in instantaneous utility, correlations between c_t and c_{t-1} could bias the estimates, depending on the setting. Third, underlying cross-sectional differences in the levels of the instantaneous utility function u may manifest themselves in level differences in expressed sentiment: if those differences correlate with temperature, this too could bias the estimates. These last two observations suggest the importance of checking for the presence of lagged effects (see appendix) and for the inclusion of a flexible set of controls to account for spurious cross-sectional and temporal correlations (see Table 4).

This paper proceeds under the assumption that the relationship between sentiment and weather provides a useful signal of individuals' underlying preferences for climate, an assumption which is abstracted in the model described above. This paper seeks primarily to identify the shape of these preferences and to calibrate the obtained estimates using intuitive comparisons to, e.g., to the average sentiment expressed on different days of the week or the shift in sentiment induced by receiving a parking ticket.

3 Data

While it would be prohibitively expensive to estimate daily sentiment at the county level using a survey, the advent of social media provides a low-cost substitute. By combining a large set of geo-located tweets with NLP algorithms specifically design to elicit sentiment, I am able to construct a dataset suitable for empirically estimating the sentiment response to meteorological changes. The following section describes the construction of the different sentiment measures and the weather covariates used in this paper. Table 1 summarizes sample characteristics for the variables to be described. 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 counties and by individuals in the data.

Table 1: Sample characteristics

	Count	Mean	Median	Min	Max
A: Sentiment measures					
AFINN	753,604,105	0.4	0.4	-5	5
Hedonometer	1,337,243,400	5.5	5.5	2	8.4
LIWC	459,035,980	0.7	0.7	-9	10.7
Machine-learned	1,463,610,503	-0.1	-0.1	-1	1
Sentiment index	1,443,757,354	0.1	0	-4.7	4.4
B: Weather covariates					
Minimum temperature (C)	1,463,610,503	9.2	11.2	-36.3	32
Maximum temperature (C)	1,463,610,503	21.1	24.1	-26	47.3
Precipitation (mm)	1,463,610,503	3.1	0	0	431.8
C: Twitter updates per...					
County	3,101	642,145	93,415	2,879	68,869,550
User	13,833,325	106	10	1	396,031

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

3.1 Twitter data

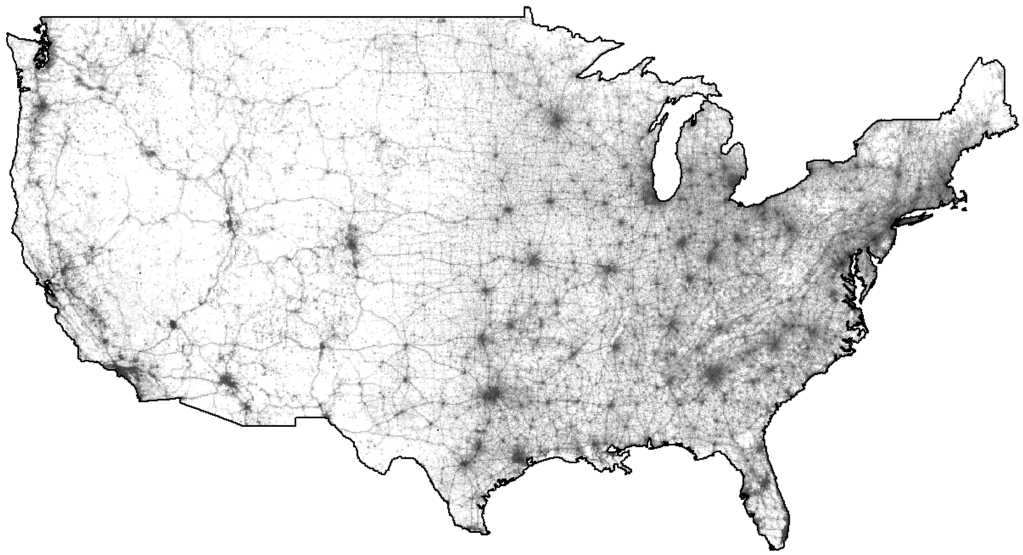
Created in 2006, Twitter is a social networking site built around the public exchange of short ($<$140 characters) Twitter updates. Since its founding, Twitter has become one of the most popular social media platforms worldwide, with 288 million active users sending over 500 million tweets per day (Twitter 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.[2] I collect the vast majority of geolocated tweets produced within my sample period, which ends in October 2016.

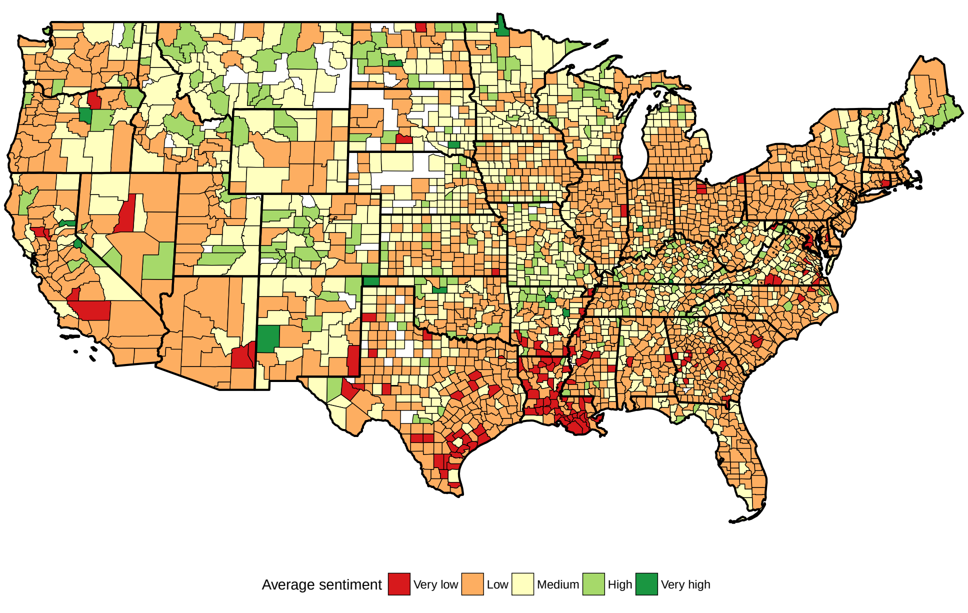
Geo-located tweets are those for which the user has consented to have his or her location information shared. The location information is either produced using the exact latitude and longitude or from a reverse-geocoding algorithm that derives the latitude and longitude from location information entered by the user. In principle, Twitter limits the total number of tweets delivered through the Streaming API to 1% (Morstatter, Pfeffer, 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). Over the course of the sample I collect, the percentage of missed tweets is fewer than 0.01% of the total available. Figure Ia maps Twitter update density, where each pixel

[2]There are two gaps, from June 26th to July 12th, 2014, and from September 18th to October 27th, 2014, corresponding to periods of time when the streaming client was unable to connect to the Streaming API.

is shaded according to the log of the total number of tweets from that area.



(a) Tweet density



(b) Average sentiment by county

I translate tweet content into four distinct measures of expressed sentiment using an NLP approach. Three of these measures use previously established word lists, while the final word list is created using a machine-learning technique. Each of these measures is named for their source material: AFINN, Hedonometer, LIWC, and Machine-learned. Finally, later parts of the analysis use a standardized index of all measures for brevity.

Panel A in Table 1 describes the unstandardized sentiment measures in the sample, although I standardize the measures prior to analysis. 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 of the tweet. Because the independent variable of interest is weather, I also remove tweets that contain any weather-related terms (list of terms available in the appendix) 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 either a pre-existing dictionary (AFINN, Hedonometer, and LIWC) or by training a machine learning algorithm on a labeled set of tweets (Machine-learned).

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 affective state (the psychological equivalent concept to emotional state) of written texts (Bradley and Lang 1999).

The Hedonometer measure is constructed in a similar manner to the AFINN measure, but the dictionary used is that provided by and described in 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.

Finally, 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.

While lexical affinity approaches such as the AFINN, Hedonometer, and LIWC methods are frequently used in the sentiment analysis literature, they can be sensitive to the particular word-sentiment measure mapping chosen by the researcher. To complement these approaches, I construct a measure of hedonic state that classifies tweets as positive or negative using a small set of assumptions and machine learning techniques.

Emoticons are text-based facsimiles of common facial expressions. In general, emoticons can indicate positive moods, e.g. ":)" or ":-)", or negative moods, e.g. ":(" or ":-(". One possible approach would be to limit the sample to tweets that contain either a positive or a negative emoticon. However, since emoticons appear in only about 2% of the sample, this approach substantially limits power. Since most Twitter updates with emoticons contain words as well, researchers in computational linguistics have employed machine learning techniques to leverage the subset of tweets with both emoticons and words to predict the sentiment of the entire set of tweets (Go, Bhayani, and Huang 2009; Kouloumpis, Wilson, and Moore 2011).

I collect a training dataset consisting of all tweets containing either positive or negative emoticons. For this training dataset, I code the hedonic state as binary and assume its polarity (1 if positive, 0 if negative) is indicated by the attached emoticon. For a full list of the emoticons used to collect this dataset, see the appendix. Next, I train an effective, computationally efficient machine learning classifier, Multinomial Naïve Bayes[3], to estimate whether particular words are more likely to be associated with positive or negative emoticons. To score tweets, I take the computed difference between the log-probability that the tweet has a positive polarity and the

[3] I use the scikit-learn implementation of the Multinomial Naive Bayes classification algorithm (Pedregosa et al. 2011).

log-probability that the tweet has a negative probability. Using this difference allows me to make use of the degree of "confidence" that the algorithm has in its classification of each tweet, where the absolute value of the measure indicates more or less confidence in the classification.

In order to summarize patterns across the four separate measures of sentiment that I construct, I use a standardized index of all of the measures. For each tweet, this index, termed the "Sentiment index", is equal to the standardized sum of the available measures. The index is intended to capture general trends across all of the measures and to make the presentation of results more convenient. I will present the validation results (Figure II) and the main result (Figure III) using all of the sentiment measures, but additional tables and figures will focus on the sentiment index.

Table 2 shows the correlations between the five measures at the county-day level. All of the measures are positively correlated with each other, which reflects general agreement between the measures. However, some of the correlations are low, likely a result of both the noise inherent in the estimation of sentiment from text and the considerable differences in the underlying methodologies used to construct each measure. The complexity of measuring emotional state suggests the importance of considering the effects across all measures rather than just one.

Next, Table 3 documents the correlation between state-level averages for each measure and compares them to state-level Gallup estimates of subjective well-being. The four sentiment measures have correlations between 0.6 and 0.9. The last row compares the measures to state-level measures of well-being estimated by Gallup in

Table 2: Measure correlations

	(1)	(2)	(3)	(4)	(5)
(1) Expert	–				
(2) Crowd-sourced	0.58	–			
(3) Lab-based	0.55	0.39	–		
(4) Machine-learned	0.44	0.29	0.33	–	
(5) Sentiment index	0.76	0.7	0.6	0.81	–

Notes: Table displays correlations between the five measures of emotional state, where one observation represents a county-day.

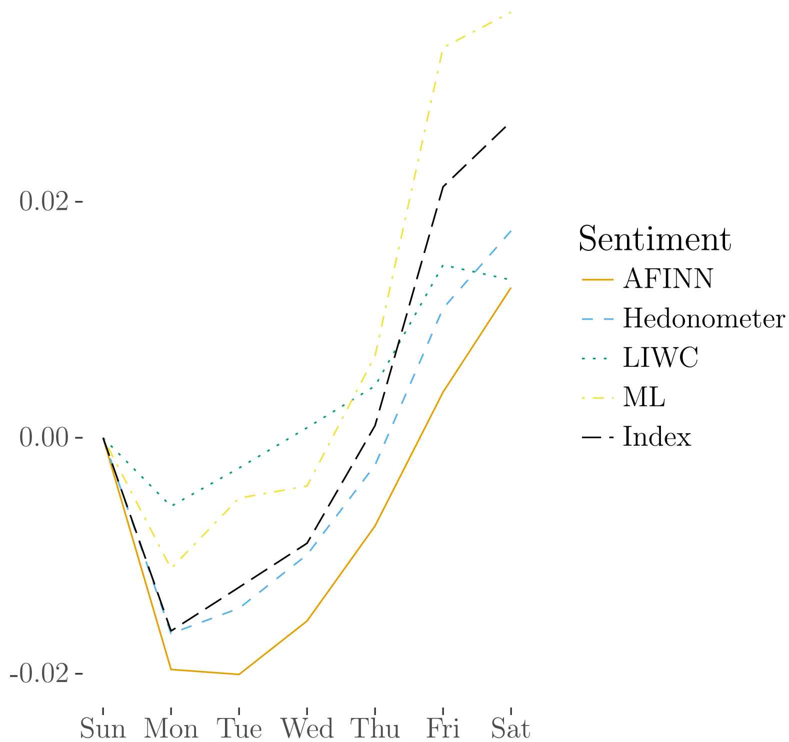
2016. All correlations are positive but low, ranging from between 0.1 and 0.5.

Table 3: State-level pairwise correlations

	(1)	(2)	(3)	(4)	(5)	(6)
(1) Expert	–					
(2) Crowd-sourced	0.89	–				
(3) LIWC	0.91	0.76	–			
(4) ML	0.82	0.6	0.77	–		
(5) Sentiment Index	0.95	0.8	0.88	0.96	–	
(6) Gallup	0.34	0.12	0.32	0.49	0.42	–

Figure II 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 0.01σ across measures.

Figure II: Weekly sentiment



Notes: Lines represent average county-day sentiment by day of week, colored by measure. Measures are standardized to have zero mean and unit standard deviation at tweet level.

3.2 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 then aggregate the gridded data to the county level using population weights (Center for International Earth Science Information Network Columbia University, United Nations Food and Agriculture Programme, and Centro Internacional de Agricultura Tropical 2005) to ensure that the weather covariates reflect the temperature experienced by the average person in the county. Because PRISM days are defined using a day-end naming standard and since most stations underlying the PRISM data report daily maximum temperature for prior 24 hours in the morning (PRISM Climate Group 2015), I use the measurement assigned to the following day as the daily maximum temperature. In other words, the daily maximum temperature indicated for a Tuesday in PRISM is linked to the outcome variable observed on Monday, when the actual observation of maximum temperature typically occurred.

Because prior work suggests that other weather variables besides temperature and precipitation may be important determinants of emotional state (Dennisenn, Butalid, Penke, and Van Aken 2008), I also include a model specification that incorporates 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 (QCLCD).

4 Empirical approach

My empirical approach combines a flexible function of temperature with a panel fixed effects model to identify the causal effect of temperature on sentiment. I justify the choice of a flexible function for temperature in two ways: first, prior work estimating temperature has documented non-linearities across a wide array of responses to temperature (Carleton and Hsiang 2016), and second, because an appropriate flexible functional form should reveal the shape of the underlying response function, linear or otherwise (Hsiang 2016). In order to interpret the estimated coefficients as causal, I require the assumption that temperature is as good as random after accounting for unobserved cross-sectional and seasonal variation (Dell, Jones, and Olken 2014).

The statistical model I estimate is given as follows:

$$\overline{S}_{cd} = \sum_{b \neq 20\text{-}25}^{B} \beta_b T_{cd}^{b} + P_{cd} + \phi_c + \phi_{smy} + \varepsilon_{cd} \tag{1}$$

Let c, s, d, m, y index county, state, day, month, and year, while b indexes five degree C temperature bins. $\overline{S}_{cd}$ is the county-day average of one of the five measures of sentiment described in section 3. T_{gd}^{b} is a indicator variable that captures whether the maximum daily temperature in a county falls within the associated bin b and P_cd is daily precipitation. ϕ_c and ϕ_{smy} represent county and state-by-month of year fixed effects. ε_{cd} is the idiosyncratic error term. I estimate the model using weighted least squares, where the weights are counts of scored tweets in a given county-day.

T_{gd}^{b} specifies five degree bins running from 0 to 40 degrees C, with edge bins for all observations with maximum temperature less than 0 or greater than 40. I use five degree bins in this model to balance flexibility and statistical power, but include analyses with varying bin sizes in the appendix. I select 20-25 C as the omitted category. This choice is aesthetic, since relative differences between conditional means are preserved, but also reflects the prior finding that Americans prefer 65 F (18.3 C) average daily temperature (Albouy, Graf, Kellogg, and Wolff 2016). In my sample, daily maximums exceed daily averages by about 6 C, suggesting that the preferred maximum daily temperature may be in the range of 20-25 C. Section 5 will document the extent to which the response functions I estimate bear this out.

Figure Ib documents cross-sectional variation in sentiment. Although all regions have a mix of high and low-sentiment counties, visual inspection suggests that there is regional variation in sentiment (for example, the southern United States seems to express lower sentiment on average). Additionally, prior evidence suggests that individuals with higher incomes tend to experience higher levels of life satisfaction (Easterlin 2001) and can afford to locate in areas with generally pleasant climate. If this regional variation, which may result from cultural or economic factors, correlates with regional weather differences, a naive estimate of the relationship between weather and sentiment is likely to be biased. To account for this regional variation in sentiment, I include county fixed effects ϕ_c. These fixed effects ensure that the model is estimated on deviations from county averages rather than on the levels of temperature and sentiment themselves. Intuitively, the implication of this modeling choice is that the estimates represent a weighted average of within-county comparisons, e.g., the difference in sentiment in Dane County, WI on a hot day versus a cold day.

A second concern for this identification strategy is the seasonality of both sentiment and temperature. Although the sentiment measures I estimate don't show clear seasonal patterns at the national level, they do suggest that some states may demonstrate within-state seasonal trends that differ from other states. To account for this possibility and to control for any time trends in the data, I include state by month-of-sample fixed effects ϕ_{smy}. 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., Wisconsin in June.

The combination of these two sets of fixed effects define the causal identification strategy I use: I assume that deviations in weather are as good as random after accounting for unobserved variation by county and state-month of sample. This assumption will be familiar to regular readers of the climate impacts literature (Hsiang 2016), and is made possible only by the density of my data. While it would possible to account for additional unobserved sources of variation in my data, e.g., fixed effects by date of sample, each additional fixed effect beyond what is necessary to produce reasonable identification assumption reduces the available variation in the data, potentially attenuating the true effect (Angrist and Pischke 2008). Nevertheless, I investigate other specifications of fixed effects in the appendix. The model specified by equation 1 is the result of this investigation and represents the most defensible tradeoff between minimizing potential bias and maximizing residual variance.

For inference I cluster standard errors two ways, by county-month and by date. 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 20-25 C with a day in temperature bin b.

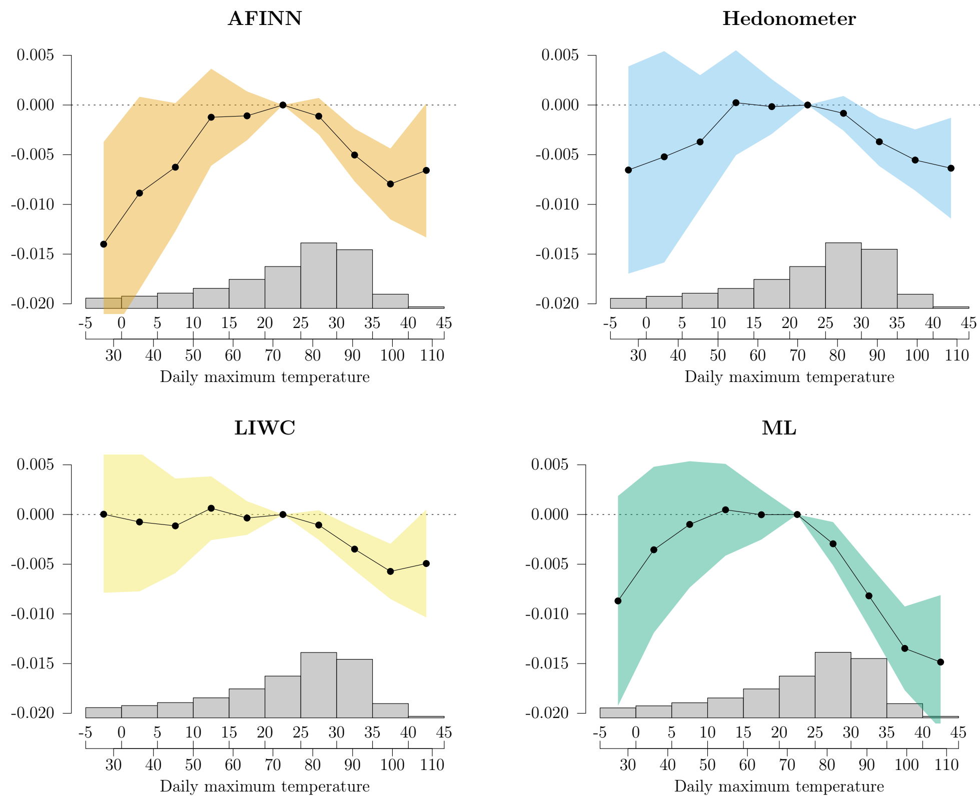
5 Results

Using the preferred econometric specification described by equation 1, I document sharp declines in sentiment above and below 20-25 C, although the point estimates for high temperatures are both more consistent across measures in magnitude and more precisely estimated than for low. For expositional clarity, I first present the main result for each sentiment measure in Figure III. I show that the shape of the response functions is remarkably similar across the different measures of sentiment despite some variation in coefficient magnitudes and the size of confidence intervals across measures. Second, in order to decompose the impact of the included fixed effects, Table 4 tabulates the response function of the Sentiment index under a range of econometric specifications, beginning with ordinary least squares (OLS) and concluding with the preferred specification.

Figure III documents the temperature response of all four measures of sentiment using the preferred specification. 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 emotional state, measured in standard deviations, expected as a result replacing a day with a high of 20-25 C with a day with the temperature maximum lying in bin b. I include a histogram underneath each plot to demonstrate the support of the temperature distribution.

The upper-left subfigure documents a sharp decline in the AFINN sentiment

Figure III: Effect of temperature on Twitter sentiment



Notes: Subfigures document the temperature response for each of the four measures of sentiment described in 3. Each point estimate represents the difference (measured in standard deviations) in county-day sentiment for the temperature bin T_b relative to 20-25 C, conditional on county and state by month of sample fixed effects. 95% confidence intervals estimated using two-way cluster robust standard errors on county and day-of-sample.

measure on cold days and a more modest decline on hot days. The difference in sentiment between days with the coldest temperatures and more temperate days is around 0.015 SD, roughly 50% higher than the difference in sentiment between hot days and temperature days. Confidence intervals are in general larger for estimates of the effect of lower temperatures than for higher temperatures. The upper-right subfigure shows the response of the Hedonometer sentiment measure. This measure finds a similar upside-down U shape but slightly more modest declines in magnitude for both temperature extremes. Again, colder temperature days tend to have slightly more negative sentiment than hotter temperature days, but there is no statistically significant difference between the estimates from the lowest (around 0.007 SD) and the highest (around 0.006 SD) bins. The confidence intervals are larger for the colder temperatures, and in fact fail to reject zero difference between the conditional mean for any cold temperature and the omitted bin.

The bottom-left subfigure documents the response estimated by the LIWC measure. In contrast to the previous two measures, the point estimates show no effect of cold temperature on sentiment, with confidence intervals that range from -0.005 SD to 0.005 SD. As before, there remains a statistically significant effect of temperature of around 0.005 SD. The final subfigure in the bottom-right captures the response of the Machine learned measure. This measure finds the most pronounced upside-down U shape, again documenting a smooth decline in temperature away from 20-25 C. As before, estimates of the effect of cold temperature on sentiment are noisy and fail to reject zero, though estimates of the effects of hot temperature are more precisely estimated. The point estimates range from -0.01 SD for cold temperatures to 0.015 SD for the hottest temperature bin.

Each outcome measure in Figure III 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 0.005 SD to 0.015 SD for the hottest temperature bin. The relationship between sentiment and cold temperatures is less precisely estimated: three of the four measures fail 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.

Table 4: Effect of temperature on social media sentiment

	(1)	(2)	(3)	(4)
Max temperature T				
$T \leq 0$	-0.012^{*}	-0.016^{***}	-0.022^{***}	-0.010^{*}
	(0.006)	(0.005)	(0.004)	(0.006)
$T \in (0,5]$	-0.009^{*}	-0.010^{**}	-0.017^{***}	-0.006
	(0.005)	(0.004)	(0.003)	(0.005)
$T \in (5, 10]$	-0.003	0.0001	-0.012^{***}	-0.003
	(0.005)	(0.004)	(0.003)	(0.003)
$T \in (10, 15]$	0.006	0.011^{***}	-0.005^{**}	0.0001
	(0.004)	(0.003)	(0.002)	(0.003)
$T \in (15, 20]$	0.012^{***}	0.011^{***}	-0.002^{*}	-0.0005
	(0.003)	(0.002)	(0.001)	(0.001)
$T \in (20, 25]$	0	0	0	0
$T \in (25, 30]$	-0.011^{***}	-0.009^{***}	0.0003	-0.002^{*}
	(0.003)	(0.002)	(0.001)	(0.001)
$T \in (30, 35]$	-0.029^{***}	-0.011^{***}	-0.002	-0.006^{***}
	(0.003)	(0.003)	(0.002)	(0.001)
$T \in (35, 40]$	-0.054^{***}	-0.020^{***}	-0.008^{***}	-0.010^{***}
	(0.004)	(0.004)	(0.002)	(0.002)
$T > 40$	-0.025^{***}	-0.027^{***}	-0.014^{***}	-0.011^{***}
	(0.010)	(0.007)	(0.004)	(0.003)
Observations (m)	2.09	2.09	2.09	2.09
Twitter updates (m)	1,443.76	1,443.76	1,443.76	1,443.76
County FE		Yes	Yes	Yes
M-o-s FE			Yes	
State × M-o-s FE				Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Dependent variable is sentiment in a county-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 and the listed fixed effects, standard errors clustered by county-month of sample and date.

Table 4 estimates the effect of temperature on the sentiment index, a composite of the four sentiment measures. Each columns represents a different regression model, differing only in the set of included fixed effects. Column (1) is the ordinary least squares (OLS) estimate, which finds a large negative effect of high temperatures, estimating that the difference between 20-25 day and a 35-40 day is more than 0.05 SD, or more than twice the difference in sentiment between a Sunday and a Monday. The effect of cold temperatures is smaller, around 0.01 SD, but statistically significant. However, since the model includes only the weather covariates, these estimates may be affected by classical omitted variables bias: endogenous sorting, regional lexical norms, income levels, and seasonal variation in temperature and emotional state may likely correlate with both temperature and sentiment. For example, the northern United States tends to be more affluent and experiences lower average temperatures. If affluence has a positive effect on sentiment, this would introduce a downward bias in the coefficients on high temperatures.

To account for unobservables in space, column (2) adds county-level fixed effects ϕ_c to the model. The inclusion of these fixed effects ensures that the model is identified using within-county fluctuations in temperature. The point estimates for the higher temperature bins shrink to about half the size of those estimated in column (1), although the point estimates for the lower temperature bins remain stable. However, this model continues to find substantial positive effects associated with temperatures between 10 and 20 C, which contrasts with intuition and prior evidence.

To control for seasonal variation, columns (3) and (4) add month of sample ϕ_{my}

and state-by-month of sample ϕ_{smy} fixed effects. The first controls for both seasonal trends and time trends, while the second allows those trends to differ by state. The addition of these controls to the model produces estimates that are more in line with intuition: days with maximum temperature from 20-25 C are have the highest sentiment, while days with low or high maximum temperature have the lowest. In particular, including these seasonal fixed effects attenuates the statistically significant positive point estimate on the 10-15 and 15-20 C bins in columns (1) and (2), suggesting that a seasonal trend in temperature may have driven that result. Column (4) reflects the preferred specification.

The negative relationship between temperature and sentiment both above and below a 20-25 C "bliss point" resembles that estimated by Albouy, Graf, Kellogg, and Wolff (2016), who find that individuals would pay to avoid warm temperatures in summer and cold temperatures in winter. The sentiment index model estimates the magnitude of the difference between a moderate day and an extremely cold or hot day to be about 0.01 SD, or roughly half of the difference in sentiment observed on a Sunday relative to a Monday. The following section explores these results in more detail.

6 Extensions

In this section I extend the results from section 5. First, I include additional weather covariates in the model and document similar qualitative results, although the magnitudes of the negative impact of hot temperatures on sentiment are larger. Second, I estimate how the sentiment response to temperature varies throughout the year. I find suggestive evidence that preferences for cold temperatures are highly seasonally dependent. Third, I document increased aggressiveness in warm temperatures using a measure of profanity usage. In the appendix I include additional specification checks, including user-level estimates, variations on bin width, and the inclusion versus exclusion of tweets containing weather-related words.

6.1 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 model (1) in order to avoid absorbing the effect of precipitation on expressed sentiment in the temperature estimates. However, since prior findings indicate that a variety of weather variables can impact stated mood (Dennisenn, Butalid, Penke, and Van Aken 2008). Here I also estimate a model with additional weather covariates compiled from the QCLCD weather station data described in section 3. To minimize measurement error, I include only counties with a QCLCD weather station present.

Table 5: Additional weather variables

	(1)	(2)	(3)
Max temperature T			
$T \leq 0$	-0.010^{*}	-0.011^{*}	-0.008
	(0.006)	(0.006)	(0.006)
$T \in (0,5]$	-0.006	-0.006	-0.002
	(0.005)	(0.005)	(0.006)
$T \in (5,10]$	-0.003	-0.003	0.0003
	(0.003)	(0.003)	(0.004)
$T \in (10,15]$	0.0001	-0.0001	0.003
	(0.003)	(0.003)	(0.003)
$T \in (15,20]$	-0.0005	-0.001	0.001
	(0.001)	(0.001)	(0.002)
$T \in (20,25]$	0	0	0
$T \in (25,30]$	-0.002^{*}	-0.002	-0.003^{***}
	(0.001)	(0.001)	(0.001)
$T \in (30,35]$	-0.006^{***}	-0.006^{***}	-0.009^{***}
	(0.001)	(0.002)	(0.002)
$T \in (35,40]$	-0.010^{***}	-0.009^{***}	-0.015^{***}
	(0.002)	(0.002)	(0.003)
$T > 40$	-0.011^{***}	-0.010^{***}	-0.017^{***}
	(0.003)	(0.003)	(0.004)
Precipitation (mm)	-0.0003^{***}	-0.0003^{***}	-0.0002^{***}
	(0.00004)	(0.00004)	(0.00004)
Diurnal range (C)			0.0003
			(0.0002)
Relative humidity (%)			-0.0002^{***}
			(0.0001)
Wind speed (mph)			-0.0001
			(0.0002)
Air pressure			0.010
			(0.007)
Overcast			-0.004^{**}
			(0.002)
Observations (m)	2,087,861	979,849	979,849

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

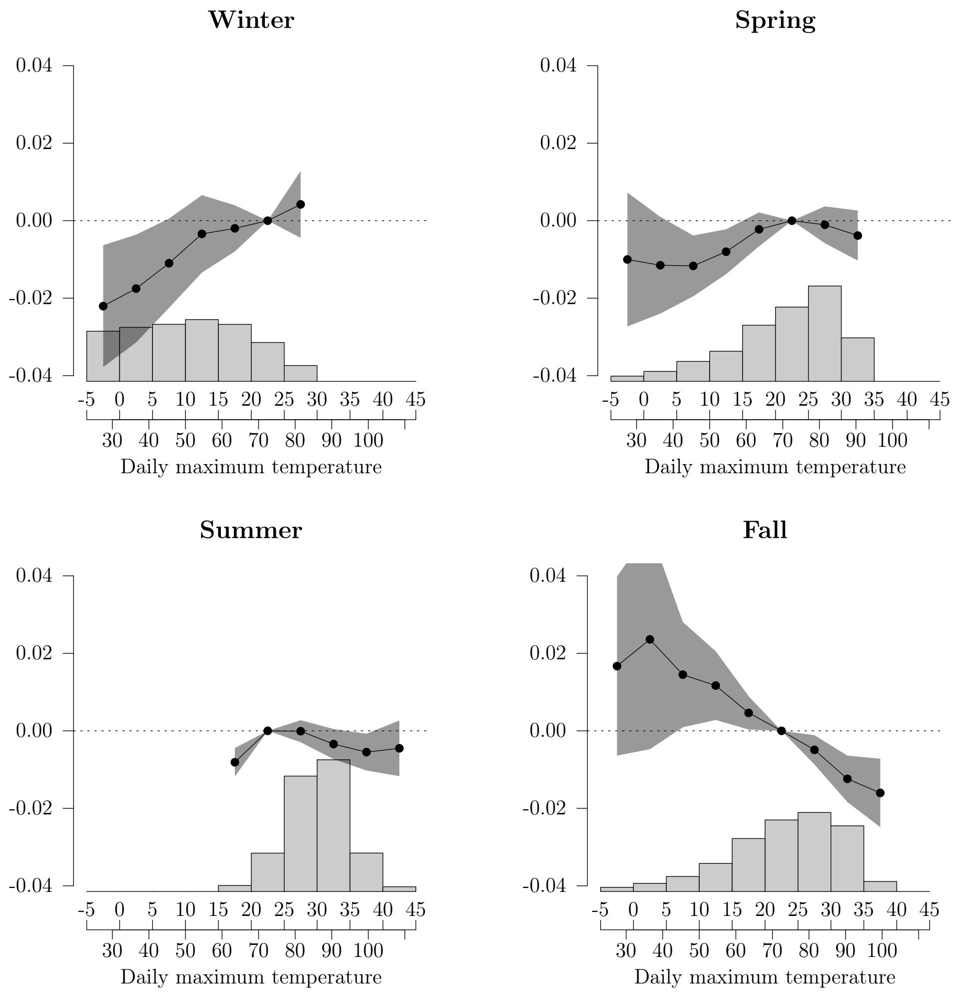
Notes: Dependent variable is sentiment in a county-day. Coefficients represent the change in standard deviations of sentiment between a day with maximum temperature in the associated temperature bin and a day with temperature $T \in [20,25)$, the omitted category. Units of air pressure are inches in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

Figure 5 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.

6.2 Season-specific responses

Willingness-to-pay estimates indicate that individuals value warm winters but cool summers (Albouy, Graf, Kellogg, and Wolff 2016). This could reflect an underlying set of preferences for moderate temperatures that is stable across seasons or seasonal shifting of temperature preferences. To distinguish between these two possibilities, I estimate model (1) separately by season, 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 IV: Seasonal response heterogeneity



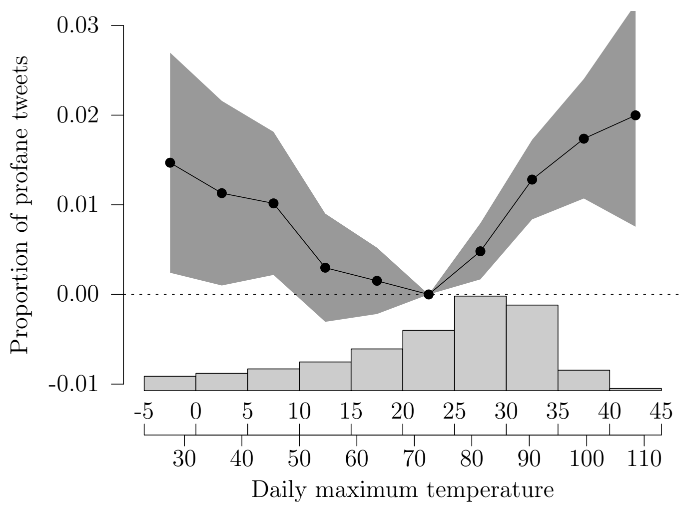
Notes: Subfigures document the response of the Sentiment Index measure to temperature for each of the four seasons. Point estimates represent the difference (measured in standard deviations) in county-day sentiment for the temperature bin T_b relative to 20-25 C, conditional on county and state by month of sample fixed effects. 95% confidence intervals estimated using two-way cluster robust standard errors on county and day-of-sample.

Figure IV documents the response functions by season for the Sentiment index. Note that because each response function is displayed relative to the omitted bin, 20-25 C, the shape of the functions capture within-season preferences for temperature, but the point estimates for individual temperatures are not comparable across subplots. These estimates reveal substantial seasonal heterogeneity in the sentiment response to temperature. In winter, moderate temperatures are preferred to cold temperatures within the observed support of the data. Spring attenuates the relative preference for moderate temperatures, while the response in summer is mostly flat, which a slight decline on hot days. The fall response function demonstrates a dramatic, though noisy, linear response to temperature: cold days are associated with higher sentiment than moderate or warm days. The contrast between spring and fall is notable in part because the temperature distributions are mostly similar across the two seasons. Speculatively, it may be that cold temperatures are preferred in fall as a relief from hot summer temperatures, whereas cold spring temperatures are dispreferred due to the recency of winter.

6.3 Profanity response to temperature

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 model 1 with the percent of tweets in a county-day that contain a profanity as the outcome of interest. Figure V plots the results.

Figure V: Profanity response to temperature



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 aggressive behavior than moderate temperatures. This finding is in contrast to previous work, which did not note an increase in crime or conflict during periods of cooler temperatures. It may be that aggressive behavior responds to temperature discomfort of both kinds, but that cooler temperatures limit opportunities to act on that aggression.

7 Valuing changes in expressed sentiment

By using metrics of emotional state derived from social media as a proxy for instantaneous utility, I am able to avoid many of the econometric biases in the existing literature (documented in Section 2). However, one advantage of the willingness-to-pay based approach is that the derivation of a so-called "money-metric", or a dollar value equivalent, is straightforward. 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 estimate 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).

In order to value the sentiment changes in response to weather, I conduct a separate analysis to estimate the relationship between small exogenous monetary shocks and sentiment. 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 very large budget. Instead, I use a cost-effective alternative, identifying more than 9,000 instances where individuals in my sample received parking tickets. Using only individuals who had at least ten tweets before and after the parking ticket, I document the sentiment response to receiving a ticket.

Figure VI: Parking ticket sentiment

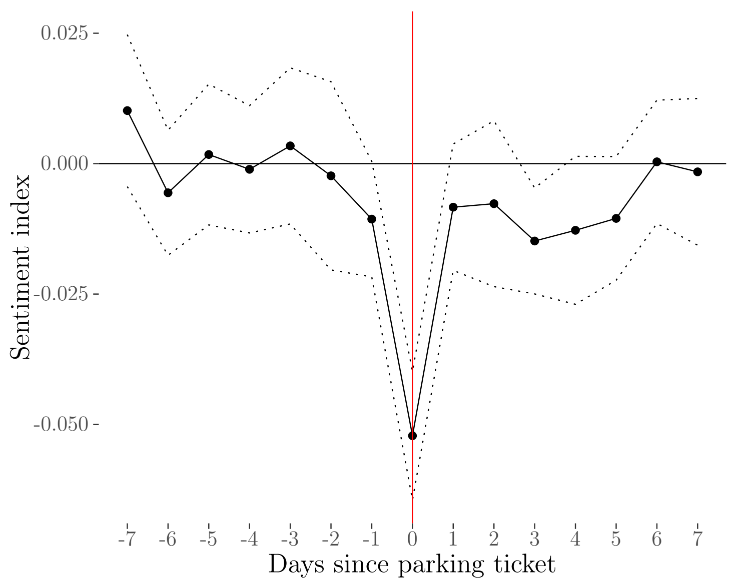


Figure VI estimates average daily sentiment on the days before and after a ticket, relative to the average sentiment in the month prior to the ticket. To value the sentiment impact of the ticket, I divide the sum of the average changes in sentiment on the seven days following the ticket by the median value of the stated ticket, \$100.[4] In total, receiving a parking ticket results in a 0.108 SD reduction in sentiment over the course of the next seven days. This calculation suggests that a 1 SD shift in sentiment is valued at roughly \$926. Table 6 applies this estimate to my main results. This approach resembles that taken by Levinson (2012), who values pollution using

[4]The decrease in sentiment prior to the receipt of the ticket may be statistical noise, or possibly due to individuals reporting their parking tickets after the fact, which I observe in several cases.

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 6: Value of temperature (parking ticket approach)

	Daily value ($)
Max temperature T	
$T \leq 0$	-8.81^{*}
	(5.10)
$T \in (0,5]$	-5.28
	(4.72)
$T \in (5, 10]$	-3.21
	(3.22)
$T \in (10, 15]$	0.07
	(2.40)
$T \in (15, 20]$	-0.43
	(1.24)
$T \in (20, 25]$	0
$T \in (25, 30]$	-1.66^{*}
	(0.95)
$T \in (30, 35]$	-5.90^{***}
	(1.37)
$T \in (35, 40]$	-9.56^{***}
	(1.75)
$T \geq 40$	-10.11^{***}
	(2.87)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Notes: Daily value is computed by multiplying the coefficients from column (4) in Table 4 by $926, the value of a 1 SD change in sentiment. Models include county-by-month of sample and date fixed effects.

The estimates in Table 6 imply, for example, that individuals in this sample would pay $10.11 to exchange a day with maximum temperature above 40 C with a day at 20-25 C. It is worthwhile to consider that 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. These results should therefore be interpreted with due caution, and I include them primarily as a demonstration of the valuation possibilities in this setting. With that consideration in mind, I apply these estimates to projection of future climate changes in the following section.

8 Climate projections

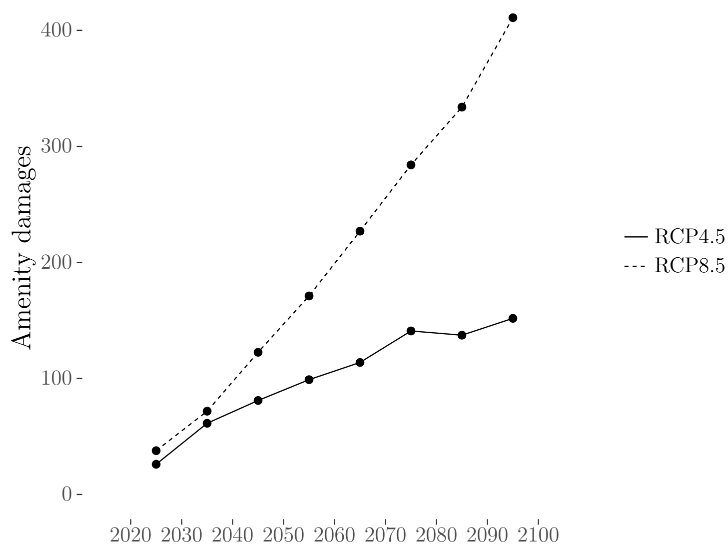
Using the monetary values from Table 6, I project the annual amenity cost of rising temperatures across the United States. Using the ensemble average from the output of 20 downscaled climate models[5], I compile population-weighted average projections for each county for the years 2006-2099. In order 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 2020 and projected monthly averages from 2006-2015, 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 then estimate the

[5]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.

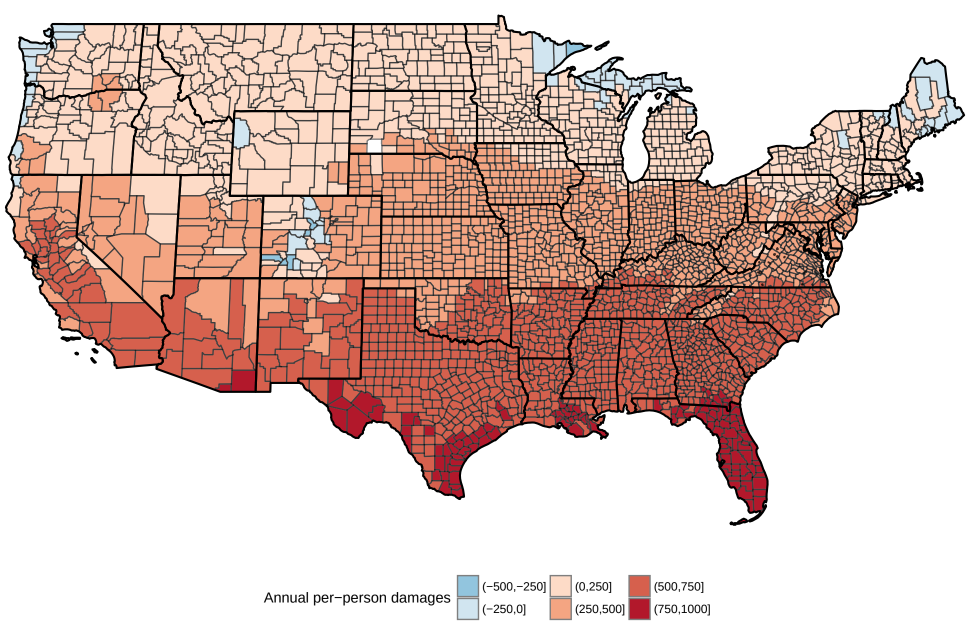
proportional change in time spent in each of my temperature bins T_b between the historical and simulated weather data. Finally, I combine these estimates of per-bin change with the normalized daily value of different temperature bins given by Table 6 and multiple by 365 to obtain a per-person annual cost of projected climate change. I conduct this exercise for both RCP4.5 and RCP8.5, two possible climate futures, where the former represents aggressive emissions cuts and the latter a maintenance of the current emissions path (IPCC 2014).

Figure VIIa documents the evolution of damages over time, averaged over counties and presented separately for RCP4.5 and RCP8.5. I estimate annual damages up to \$400 per person under RCP8.5 and up to \$125 per person under RCP4.5. Figure VIIb maps end of century damages under RCP8.5 by county, documenting clear north-south heterogeneity in the extent of amenity costs due to climate change. In some areas, largely in the southern United States, damages are as high as \$1,000, while in others value of climate amenities actually increases by end-of-century, primarily in the northern United States. By comparison, the central estimates in **??** predict annual welfare losses between 1 and 3% of income by end of century, while my central estimate of \$400 in the RCP8.5 scenario is slightly less than 1% of annual income.

(a) Projected damages over time



(b) End of century projection (RCP8.5)



9 Discussion

In this paper I have sought to add a new chapter to the long history of environmental valuation (Pearce 2002). I interpret sentiment expressed on a social media as a proxy for near-term changes in utility, then compare the magnitude of shifts in expressed sentiment caused by weather to the magnitude of sentiment shifts caused by small but plausibly exogenous shocks to personal income. In principle, this method allows researchers to estimate preferences over public goods while accounting for a wide range of unobservable variation across both space and time by using data on short-term variation in expressed sentiment.

However, this approach is not without its drawbacks. First, the formation of emotional state is far more complex than the model in section 2 portrays. The physical, biological, and psychological bases for human emotions are 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 geocoding Twitter user population is not representative on the United States at large. Finally, this approach is useful only in settings in which the right-hand side variable of interest fluctuations across both time and space.

The valuation section of this paper suffers from an additional weakness, which I describe in more detail in section 7: by valuing shifts in sentiment using an exogenous shock to wealth, the valuation of temperature is now dependent on the validity of both the temperature and parking ticket estimates. In the appendix, I document that models which include date fixed effects and disaggregated models using only the tweet data find attenuated coefficients for warmer temperatures, which would reduce the amenity damages projected in Figure VIIb.

Lastly, 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.

Despite these limitations, this paper makes several contributions to the literature. It introduces a new methodology and data source to estimate preferences for and valuations of public goods while simultaneously accounting for possible unobservable cross-sectional and seasonal variation. It demonstrates how an appropriate use of NLP and machine-learning algorithms can enable the econometric analysis of large text-based datasets and suggests a psychological channel through which other impacts of climate change may operate. Broadly, this work provides supporting evidence that changes in the amenity value of climate are an important component of the overall costs of climate change.

References

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

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

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

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.

Baylis, Patrick, Nick Obradovich, Yury Kryvasheyeu, Haohui Chen, Lorenzo Coviello, Esteban Moro Egido, Manuel Cebrian, and James H. Fowler. 2017. "Weather impacts expressed sentiment". *Manuscript submitted for publication.*

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, 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.*

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

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

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

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

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

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

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

Feddersen, John, Robert Metcalfe, and Mark Wooden. 2012. “Subjective Well-Being: Weather Matters; Climate Doesn’t”. *SSRN Electronic Journal*, number 627.

Go, Alec, Richa Bhayani, and Lei Huang. 2009. “Twitter Sentiment Classification using Distant Supervision”. *Processing* 150 (12): 1–6.

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

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

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

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.

Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. 2011. "Twitter Sentiment Analysis : The Good the Bad and the OMG!" *Artificial Intelligence* 11:538–541.

Levinson, Arik. 2012. "Valuing public goods using happiness data: The case of air quality". *Journal of Public Economics* 96, number 9–10 (): 869–880.

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

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.

Morstatter, Fred, Jürgen 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 Årup. 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.

Pedregosa, Fabian, et al. 2011. “Scikit-learn: Machine Learning in Python”. *The Journal of Machine Learning Research* 12:2825–2830.

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

PRISM Climate Group. 2015. *Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States.* Technical report June 2013.

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.

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.

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

Twitter. 2015. *Company | About.*

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

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