Rapidly-Adjusting Perceptions of Temperature in a Changing Climate

Frances C. Moore, University of California Davis

Nick Obradovich, Massachusetts Institute of Technology

Flavio Lehner, National Center for Atmospheric Research

Abstract

As the global climate changes, people are exposed to weather that is increasingly unusual relative to historical or pre-industrial conditions. However, expectations, memory limitations, and cognitive biases may influence people's subjective experience of the weather. How do people judge today's weather as typical or atypical? And how might that judgement shift in response to gradually-changing climatic conditions? Here we show that experience of weather in recent years, rather than longer historical periods, determines the baseline against which current weather is evaluated, potentially obscuring the signal of anthropogenic climate change as subjectively experienced. We employ variation in decadal trends in temperature at weekly and county resolution over the continental United States, combined with discussion of the weather drawn from over two billion social media posts. These data indicate that the remarkability of particular temperatures, measured as the volume of posts about weather that they generate, changes on relatively short timescales. We develop a learning model from our empirical results and apply it to climate model output to project the perception of temperature anomalies arising from future climate change. The rapidly-shifting baselines we observe have substantial implications for the public perception of anthropogenic warming.

Main Text

Environmental change involves the gradual shifting of system characteristics beyond bounds historically experienced by communities and ecosystems. Though the signal of these changes emerges clearly when examined at long time-scales or at large spatial scales, individual experience of change occurs locally and may be influenced by expectations, memory limitations, and cognitive biases[1,2].

Direct personal experience of environmental change, as an immediate, salient, and highly-trusted information source, may be critical in convincing the public of both the existence of a problem and the need for corrective policies to address it. This has been widely documented in the case of climate change: local weather anomalies alter stated belief in climate change[3–8], and Americans self-report local weather conditions as influencing their opinions on climate change[9]. However, if individuals dynamically adjust their perceptions of 'normal' climate, experience of historically unusual conditions may not provide strong experiential evidence of environmental change over time.

Despite the importance of understanding the perception of both climate change and other forms of gradual environmental change, relatively little work has examined this topic empirically. Climate models have been used to determine the statistical "time of emergence" of the climate change signal[10–12], but the relationship between these metrics and the general public's perception of environmental change is unclear. Other work has applied hypothesized learning models to the problem of inferring the climate state from weather observations but these have not been tested against observed behavior[13–16].

Here we show that the remarkability of particular weekly temperature anomalies adjusts on approximately a five year timescale, suggesting that the sense of "normal weather" shifts relatively rapidly with a changing climate. We measure the remarkability of temperature as the volume of social media posts about weather that it generates and use the substantial spatial and seasonal variation in decadal temperature trends to identify the causal effect of repeated exposure to a given temperature anomaly on the remarkability of contemporaneous temperatures. We show that average climate conditions affect the likelihood that particular temperatures are remarked on (i.e. people respond more to temperatures if they are historically unusual for that location and time-of-year) but that recent trends are important in moderating this effect. Using a dynamic lag model, we find that the effect of temperature anomalies decays with repeated exposure over two to eight years. From these results, we derive a learning model describing how perceptions of normal temperatures might adjust in response to future warming and apply it to climate change projections. The shifting baseline we uncover has large implications for the magnitude of perceived temperature anomalies in a changing climate.

Our social media data consists of all posts on Twitter between March 2014 and November 2016 geolocated within the continental United States, for a total of 2.18 billion tweets (Supplementary Figure 1). Tweets about weather were identified using a simple 'bag of words' approach (Supplementary Methods), and the classification was validated manually for 6,000 selectively-sampled tweets (Supplementary Methods, Supplementary Table 1). Twitter is a medium uniquely suited to examining this phenomenon because its wide geographic scope and high temporal resolution allow us to sample variation in both spatial and seasonal climate trends, and because the low marginal cost of tweeting provides a nearly instantaneous response to weather conditions, unaffected by market distortions or imperfections that might make responses to weather events on other margins more difficult to interpret.

We draw data on daily maximum temperature and total precipitation for the period 1981-2016 from the PRISM data set and aggregated these to the county level from a 0.25 degree grid[17]. We combine the PRISM data with cloud cover and relative humidity data from the NCEP Reanalysis II[18].

Social media and weather data are aggregated to the county (spatial) and weekly (temporal) level. We employ weekly rather than daily resolution as weeks are a plausible period over which people might resolve the seasonal climatology of their area (e.g. “end of March”, “mid-late November”)[19]. For each county-week combination a ten year “reference” period is defined as the average of the county-week’s temperature across the years 1981-1990, a period defined based on the earliest-available daily PRISM data. For comparison, a “recent” period was defined as the most recent five years. As an illustration, Figure 1 shows the spatial and seasonal variation in climate trends (defined as the difference between recent and reference temperatures) across the United States for the third week in each calendar month. It shows substantial variation in exposure to temperature changes, both across space and within the year. This variation is what we use to test whether the response to historically-unusual weather conditions changes with repeated exposure to those conditions.

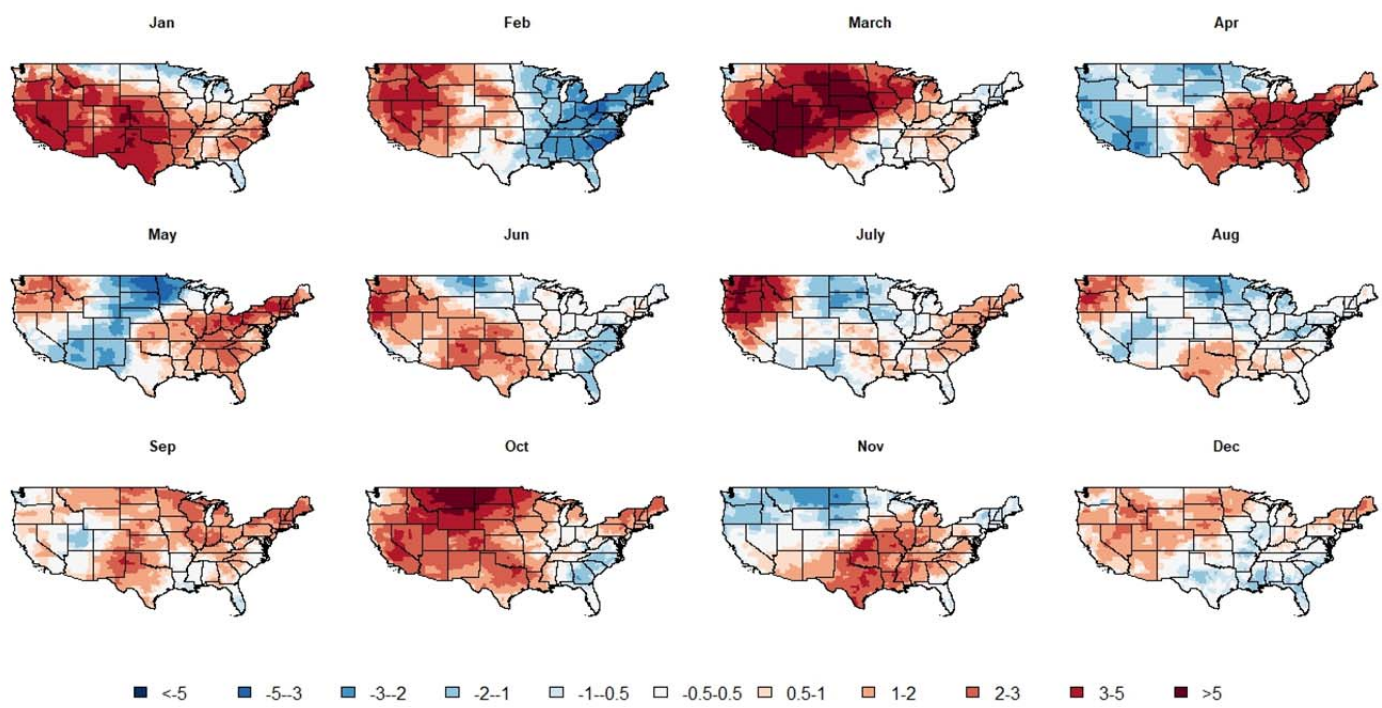


Figure 1: Spatial and seasonal variation in the change in average temperatures between the reference (1981-1990) and recent (2011-2015) time periods (in degrees C). *Values shown are averaged for the third week in each month.*

Our principal empirical model regresses the logarithm of the number of weather tweets in each county-week on functions of reference and recent temperatures. The model includes controls for precipitation, relative humidity, and cloud cover (in order to isolate the effect of temperature) as well as differences in Twitter use in counties and over time using the logarithm of the number of Twitter users. County indicator variables (fixed-effects) control for all time-invariant difference between counties while state by month-of-year indicator variables (e.g. December in California) flexibly control for any regional differences in seasonality. Finally, year fixed-effects control for common time trends across the US over the sample period. The residual variation used to identify the causal effect of temperature fluctuations on social media posts about weather is shown in Supplementary Figure 2. Standard errors are clustered at the state level, allowing for spatial and temporal autocorrelation within a state (more details and the regression equation are given in the Supplementary Methods).

We first look at the importance of the historic reference temperatures (1981-1990) in determining the response to weather conditions. Figure 2a shows that the effect of temperature on social media posts differs depending on the reference temperature for that county and time of year. People are more likely to comment on weather that is unusual for a particular place and time-of-year then on the same weather if it is typical. At the median of the reference temperature distribution (22°C), the quadratic minimum is remarkably close to the reference temperature, meaning people comment least on temperatures close to these reference conditions. At both hotter and colder extremes of the temperature distribution, the response becomes more asymmetric, suggesting that a combination of unexpectedness and consequence might drive the remarkability of particular temperatures. Regression coefficients are given in Supplementary Table 2 and an F-test of the null hypothesis that reference temperatures do not moderate the response to temperature is strongly rejected (F-stat = 19.23, 4 and 48 degrees of freedom, $p<1e\text{-}5$). The curves are also statistically different from each other over much of the temperature range (Supplementary Figure 3). Allowing for a more flexible (quartic) response shows a qualitatively similar effect, but with some evidence for a declining marginal effect at very hot temperatures (Supplementary Figure 4).

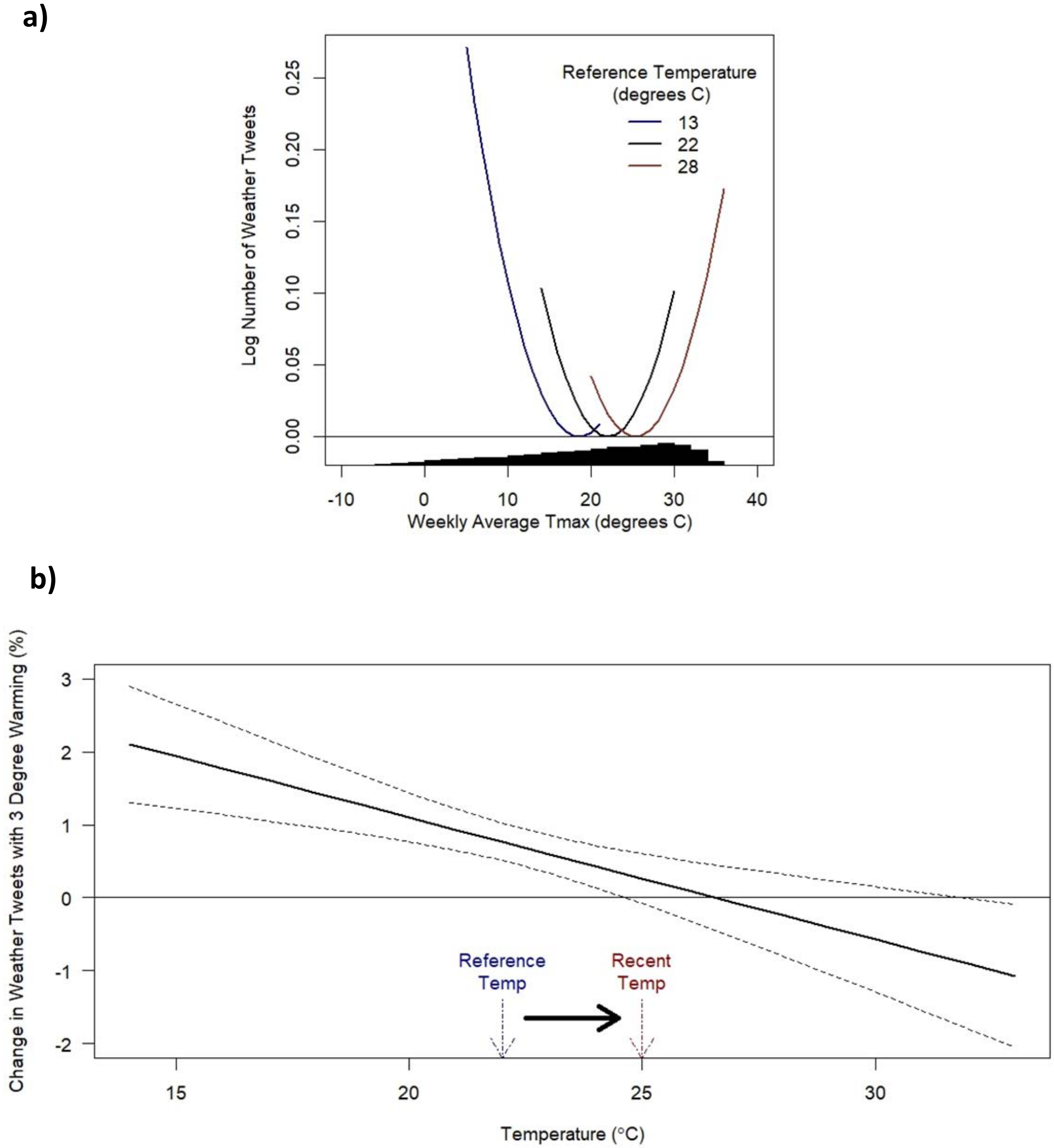


Figure 2: Effect of reference temperatures and recent changes on number of weather tweets. *a) Effect of average weekly daytime temperature on tweeting for three different reference temperatures corresponding to the 25th, 50th, and 75th percentile of the sample. Curves are shown for +/- 8 degrees from reference, which includes >97.5% of the observed weekly average temperature anomalies in our sample. Because the dependent variable is logged, movement along the y axis can be interpreted as % change in the number of weather posts. The histogram shows the distribution of reference temperatures in the sample. b) Percent change in the number of weather tweets in response to contemporaneous temperatures (x-axis), shown for a location that has warmed from 22°C to 25°C (+3°C) between the reference (1981-1990) and recent (last 5 years) time periods. Dashed lines show the 95% confidence interval. Regression coefficients are given in Supplementary Table 2.*

Having established that prior experiences alter reactions to realized temperature, we now investigate whether the historic reference period, more recent experience, or some mixture of the two best explains the volume of comments about the weather. We find that the change in temperature between reference (1981-1990) and recent (prior 5 years) time periods is a highly significant explanatory variable, moderating the response to particular temperatures (Supplementary Table 2). Figure 2b shows the effect of 3°C warming between these time periods in a county-week with an initial reference temperature of 22°C (and therefore a recent temperature of 25°C). The recent warming results in 1% more comments in response to unusually cool temperatures of 20°C than would have occurred without that warming. In other words, 20°C has become more remarkable because the recent warming has made it more unusual relative to the original reference period.

Next, we employ a finite distributed lag model to more precisely estimate the temporal dynamics of subjective baseline adjustment. For each county-week in our sample we use its 15-year history of temperature anomalies, defined relative to the 1981-1990 reference period, to estimate how behavior adjusts in response to repeated exposure to altered temperatures. The model estimates the effect of previous temperature anomalies (experienced between 1 and 15 years ago) on current behavior, allowing for non-linear effects that change over time (for additional details see Supplementary Methods). We split our data and estimate responses separately for the hottest and coldest third of baseline temperatures (greater than 26.3° and less than 16.8° respectively). This is necessary because the asymmetry of the response curves shown in Figure 2a about the reference temperature means that the same temperature anomaly would be expected to have different effects at the hot and cold ends of the temperature distribution, which we allow for by splitting the sample. Model summaries are given in Supplementary Table 3 and additional model results are shown in Supplementary Figure 5.

Figure 3 shows how the effect of hot and cold temperature anomalies on the volume of posts about weather varies as a function of length of exposure to those temperature anomalies, for the coolest third of our sample. We find the same downward-sloping response to instantaneous warmer temperatures as shown in Figure 2a (i.e. the volume of posts increases in response to cold anomalies and decreases in response to warm anomalies). However, this response also decays rapidly with longer-term changes in temperature. After 5-10 years of exposure to historically-unusual conditions, the response either disappears (for warm anomalies) or is even reversed (for cool anomalies). In other words, the kind of temperature considered remarkable changes rapidly in response to repeated exposure. Results for the hot third of the sample are shown in Supplementary Figure 6 but are not statistically significant. Results for the full sample and for alternative specifications of the dynamic lag model are qualitatively similar to those shown in Figure 3 (Supplementary Figures 7 and 8).

Reversal of the instantaneous effect, as seen in the upper panel of Figure 3 between 6 and 12 years of exposure, is robust to alternate model specifications and is consistent with a shifting-baseline model: after continuous exposure to cooler temperatures, reference temperatures (i.e. zero anomaly relative to 1981-1990) are warm relative to recent experience and therefore result in a decreased response, consistent with the instantaneous response to warm anomalies shown in the lower panel. Given the curvature of the response curves for this temperature range (Figure 2a), the same effect for warm anomalies would be expected to be less well-defined, but the uncertainty bounds do not rule out a similar effect.

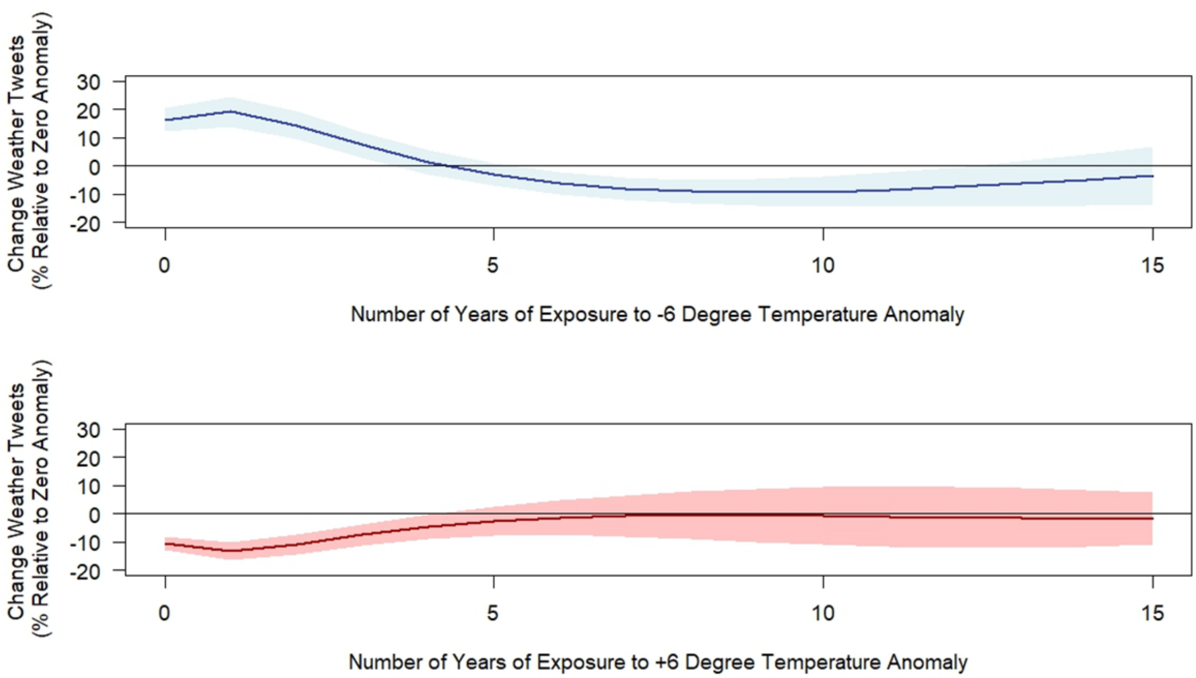


Figure 3: Change in response to temperature anomalies with repeated exposure. *Percent change in the number of weather posts in response to temperatures that are 6°C cooler (upper panel) or warmer (lower panel) than the reference period, as a function of the number of years of exposure to that temperature. A 6°C anomaly represents approximately the central 90% of the sample. Shaded areas give the 95% confidence interval based on standard errors clustered at the state level.*

Using the coefficients of the dynamic-lag model, we derive a learning model that describes how baselines adjust in response to experienced temperatures (Supplementary Methods). Some previous work has hypothesized that recent weather should be particularly important in moderating responses to weather conditions and we find empirical support for this "recency-bias"[6,20]. Temperatures experienced between 2 and 5 years ago appear to be particularly important in defining baselines against which current temperatures are evaluated (Supplementary Figure 9). We combine this learning model with the time-series of population-weighted annual temperatures over the continental United States in order to estimate how perceptions of "normal" temperatures have changed over the historical period. Based on this learning model, warming experienced between 1860 and 2016 has resulted in a shift of over 1°C in aggregate across the US public, suggesting that temperatures that would previously be considered warm are now perceived as normal (Supplementary Figure 10).

Shifting temperature baselines also have implications for the experience of future warming under climate change. Figure 4 shows the population-weighted temperature anomalies under the RCP 8.5 emissions scenario over the continental United States, for 40 realizations of internal variability[21]. Anomalies are defined relative to both a fixed 30-year baseline (1981-2010) and to a shifting baseline defined using our empirically-estimated learning model. While persistent warming over the 21st century results in temperatures that are increasingly unusual relative to a fixed, historically-defined baseline, the rapidly-shifting baseline that we find evidence for here results in perceived temperature anomalies on average only slightly above zero. Thus, although temperatures increase substantially in an absolute sense, because they do so only gradually relative to the rate at which people appear to update their baselines, even the substantial warming generated in a high emissions scenario may not produce perceptions of unusually warm conditions.

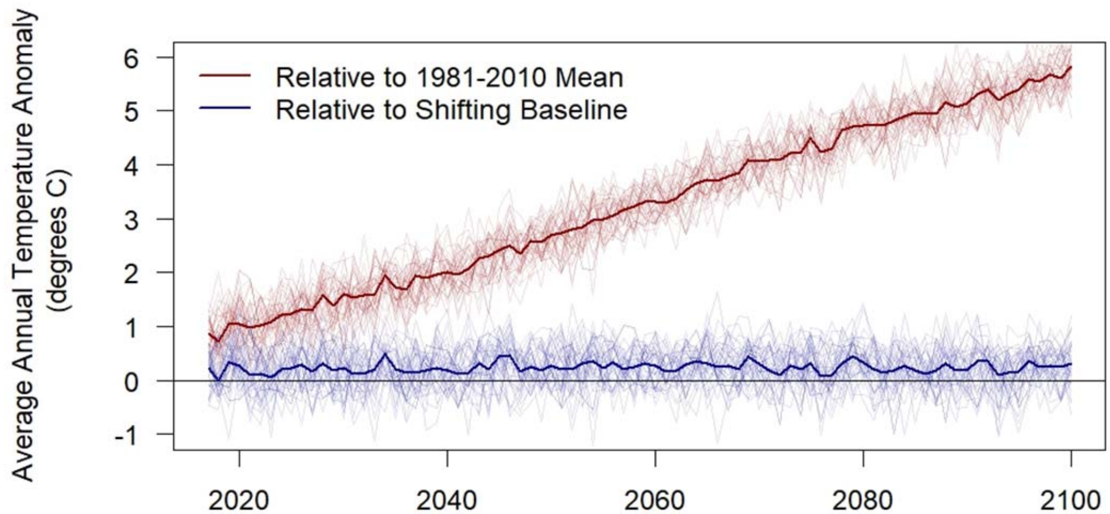


Figure 4: Effect of shifting baselines on the perception of temperature anomalies. *Population-weighted annual average temperature anomalies over the continental U.S. under RCP 8.5 with 40 realizations of internal variability[21]. Anomalies are defined relative to a fixed 30-year period (1981-2010) and relative to a shifting baseline defined using our estimated learning process. Population-weighting uses population density fixed at 2015 values[22].*

Here we show that the remarkability of temperature depends not just on its absolute value, but that it is affected by past experience and expectations. More specifically, the subjective baseline against which temperature is evaluated appears to be dominated by recent experience. Temperatures initially considered remarkable rapidly become unremarkable with repeated exposure over roughly a 5-year timescale. Since this is fast relative to the pace of anthropogenic climate change, this shifting subjective baseline has large implications for the perception of temperature anomalies as climate change progresses.

There are a number of important considerations related to our conclusions. First, it is important to be specific about what we are and are not measuring: our metric of the volume of social media posts about weather measures - in a very literal way - the “remarkability” of temperature. We are not able to determine here precisely what makes a weather event remarkable, though it plausibly involves some combination of surprise and consequentiality. It may be that both are changing in response to repeated exposure, but we do not attempt to resolve their relative contributions.

Secondly, we do not link our conclusions directly to stated belief in anthropogenic climate change or support for mitigation policies. These stated positions appear to be influenced by a range of factors, including cultural worldviews, political affiliations, and the perceived legitimacy of message promotors[23–25]. Moreover, much of the spatial and seasonal variation in temperature trends that we use in our estimate (Figure 1) is likely a result of natural variability rather than anthropogenic climate change, meaning it is unclear how these changes should affect climate change beliefs. It is also possible that warmer temperatures could at once be socially unremarkable and yet still provide evidence for anthropogenic climate change, when processed on a deeper cognitive level than that used in posting on social media (i.e. using System 2 vs System 1 processing)[26,27]. Our conclusion is only that rising temperatures alone will not necessarily provide direct, experiential evidence of anthropogenic climate change if perceptions of normal adjust rapidly, as we find evidence for here. Though many studies have now identified a link between stated belief in global warming and temperature anomalies (typically defined relative to fixed reference periods)[4,5,8,19,28], our results suggest that care should be taken in projecting these findings forward to infer increased public belief in climate change with warmer temperatures.

Finally, we note that our results pertain only to ambient average temperatures. It may well be that more acute extreme events such as storms, droughts, wildfires, or floods may be both more consequential and salient and therefore less prone to normalization[29]. Previous work has found that other variables such as changes in phenology or snowfall might be more strongly attributed to climate change in the public consciousness.[27] It is also possible that particular physical or biological thresholds beyond the range of our data may result in non-linear responses that are not accounted for in this study.

The pre-industrial is often used as a standard reference point in both climate science and policy[30], and unmitigated greenhouse gas emissions over the 21st century will result in large warming relative to this period. Understanding how these historically-unusual temperatures are perceived by people affected, and in particular whether they provide direct sensory evidence for the existence of climate change, requires knowing how weather is socially determined to be “normal” or “unusual”. Here we present evidence that the definition of normal adjusts rapidly in response to changed conditions, resulting in perceived temperature anomalies that are close to zero over the 21st century, even in a high emissions scenario. When coupled with results from the existing literature, our finding suggests it may be unlikely that rising temperatures alone will be sufficient to produce widespread support for mitigation policies.

Acknowledgements

Thank you to Jeffrey Schrader, Patrick Baylis, Angeline Pendergrassie, and seminar participants at University of California Berkeley and University of Nevada Reno for comments on the paper. Thank you to Rudy Huezo for stellar research assistance.

References

1. Tversky, A. & Kahneman, D. Judgment and Uncertainty: Heuristics and Biases. *Science (80-.).* **185,** 1124–1131 (1974).

2. Simon, H. A. A Behavioral Model of Rational Choice. *Q. J. Econ.* **69,** 99–118 (1955).

3. Zaval, L., Keenan, E. A., Johnson, E. J. & Weber, E. U. How warm days increase belief in global warming. *Nat. Clim. Chang.* **4,** 143–147 (2014).

4. Donner, S. D. & McDaniels, J. The Influence of National Temperature Fluctuations on Opinions About Climate Change in the US Since 1990. *Clim. Change* **118,** 537–550 (2013).

5. Egan, P. J. & Mullin, M. Turning Personal Experience into Political Americans ' Perceptions about Global Warming. *J. Polit.* **74,** 796–809 (2012).

6. Kaufmann, R. K., Mann, M. L., Gopal, S., Liederman, J. A., Howe, P. D., Pretis, F., Tang, X. & Gilmore, M. Spatial heterogeneity of climate change as an experiential basis for skepticism. *Proc. Natl. Acad. Sci. U. S. A.* **114,** 67–71 (2017).

7. Myers, T. A., Maibach, E., Roser-Renouf, C., Akerlof, K. & Leiserowitz, A. The Relationship Between Personal Experience and Belief in the Reality of Global Warming. *Nat. Clim. Chang.* **3,** 343–347 (2013).

8. Deryugina, T. How do people update ? The effects of local weather fluctuations on beliefs about global warming. 397–416 (2013). doi:10.1007/s10584-012-0615-1

9. Borick, C. P. & Rabe, B. G. Weather or Not? Examining the Impact of Meteorological Conditions on Public Opinion Regarding Global Warming. *Weather. Clim. Soc.* **6,** 140508110431002 (2014).

10. Hawkins, E. & Sutton, R. Time of Emergence of Climate Signals. *Geophys. Res. Lett.* **39,** L01702 (2012).

11. Diffenbaugh, N. S. & Scherer, M. Observational and Model Evidence of Global Emergence of Permanent, Unprecedented Heat in the 20th and 21st Centuries. *Clim. Change* **107,** 615–624 (2011).

12. Frame, D., Joshi, M., Hawkins, E., Harrington, L. J. & de Roiste, M. Population-based emergence of unfamiliar climates. *Nat. Clim. Chang.* **7,** 407–411 (2017).

13. Kelly, D., Kolstad, C. & Mitchell, G. Adjustment Costs from Environmental Change. *J. Environ. Econ. Manage.* **50,** 468–495 (2005).

14. Moore, F. C. Learning, Adaptation, and Weather in a Changing Climate. *Clim. Chang. Econ.* **8,** 1750010–1 (2017).

15. Lehner, F. & Stocker, T. F. From Local Perception to Global Perspective. *Nat. Clim. Chang.* **5,** 731–

734 (2015).

16. Ricke, K. L. & Caldeira, K. Natural climate variability and future climate policy. *Nat. Clim. Chang.* **4,** 333–338 (2014).

17. PRISM. Oregon State University.

18. Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M. & Potter, G. L. NCEP-DOE AMIP-II Reanalysis (R2). *Bull. Am. Meteorol. Soc.* **83,** 1631–1643 (2002).

19. Howe, P. D. & Leiserowitz, A. Who remembers a hot summer or a cold winter ? The asymmetric effect of beliefs about global warming on perceptions of local climate conditions in the U . S . *Glob. Environ. Chang.* **23,** 1488–1500 (2013).

20. Weber, E. U. & Stern, P. C. Public Understanding of Climate Change in the United States. *Am. Psychol.* **66,** 315–328 (2011).

21. Kay, J. E., Deser, C., Philips, A. S., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., *et al.* The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bull. Am. Meteorol. Soc.* **96,** 1333–1349 (2013).

22. CIESIN, University, C. & CIAT. *Gridded Population of the World, Version 3: Population Density Grid*.

23. Maibach, E. W., Leiserowitz, A. A., Roser-Renouf, C., Myers, T. A., Rosenthal, S. & Feinberg, G. *The Francis Effect: How Pope Francis Changed the Conversation about Global Warming*. (2015).

24. Kahan, D. M., Jenkins-Smith, H. & Braman, D. Cultural cognition of scientific consensus. *J. Risk Res.* **14,** 147–174 (2011).

25. Ruddell, D., Harlan, S. L., Grossman-Clarke, S. & Chowell, G. Scales of perception: public awareness of regional and neighborhood climates. doi:10.1007/s10584-011-0165-y

26. Kahneman, D. *Thinking, Fast and Slow*. (Farrar, Strauss and Giroux, 2011).

27. Akerlof, K., Maibach, E. W., Fitzgerald, D., Cedeno, A. Y. & Neuman, A. Do people '" personally experience "' global warming , and if so how , and does it matter ? *Glob. Environ. Chang.* **23,** 81–91 (2013).

28. Hamilton, L. C. & Stampone, M. D. Blowin' in the Wind: Short-Term Weather and Belief in Anthropogenic Climate Change. *Weather. Clim. Soc.* **5,** 112–119 (2013).

29. Howe, P. D., Boudet, H., Leiserowitz, A. & Maibach, E. W. Mapping the shadow of experience of extreme weather events. *Clim. Change* **127,** 381–389 (2014).

30. Schurer, A. P., Mann, M. E., Hawkins, E., Tett, S. F. B. & Hegerl, G. C. Importance of the pre-industrial baseline for likelihood of exceeding Paris goals. *Nat. Clim. Chang.* **7,** 563–567 (2017).

Supplementary Methods

1. Data Sources and Processing

1.1 Weather Data

Data on daily maximum temperature and total daily precipitation for the period 1981-2016 are from the PRISM data set, gridded at 0.25 degrees resolution[1]. Gridded data are aggregated to the county level (spatial weighting) and then averaged over weeks to give county by week observations. We focus on maximum temperature rather than average or minimum temperatures since they occur during the day and are therefore likely to be most salient to people and most relevant for explaining tweeting behavior, at both the hot and cold ends of the temperature distribution. We also include data on percent cloud cover and relative humidity, also averaged to the county by week level, from the NCEP Reanalysis II[2]. Although the focus of our analysis is the effect of temperature variation, we include these other salient features of the weather that might be correlated with temperature (specifically, rainfall, cloud cover, and relative humidity) as control variables in the regression in order to isolate the effect of temperature itself.

Average annual temperatures from 1850 to 2017 (used for Supplementary Figure 10) are calculated from gridded monthly data provided by Berkeley Earth[3]. Data are averaged over years and then aggregated to the national level using gridded 2015 population data from CIESIN[4]. Temperature projections under RCP 8.5 (used in Figure 4) are 40 realizations of the Community Earth System Model (CESM1) Large Ensemble Project[5]. Annual temperatures are aggregated to the national level using the same weighting by 2015 population.

1.2 Twitter Data

All tweets between March 2014 and the end of November 2016 geolocated within the continental United States were downloaded from the Twitter API (geolocated tweets exclude retweets). Tweets within the continental US were identified using a bounding box filter. Each tweet was allocated to a county using either the 'geo.coordinates' value from the tweet metadata or, if this was missing, the centroid of the 'place' bounding box. This gives a total of 2.18 billion tweets in the sample. The number of geolocated tweets is gradually increasing over this time-period, with the exception of a sharp drop in late 2014, likely associated with a change in the Twitter's opt-in policy for geolocating tweets (Supplementary Figure 1).

Tweets discussing weather were identified using a simple bag-of-words approach. If the tweet contained one of the following words it was classified as a 'weather tweet':

arid, aridity, autumnal, balmy, barometric, blizzard, blizzards, blustering, blustery, blustery, breeze, breezes, breezy, celsius, chill, chilled, chillier, chilliest, chilly, cloud, cloudburst, cloudbursts, cloudier, cloudiest, clouds, cloudy, cold, colder, coldest, cooled, cooling, cools, cumulonimbus, cumulus, cyclone, cyclones, damp, damp, damper, damper, dampest, dampest, deluge, dew, dews, dewy, downdraft, downdrafts, downpour, downpours, drier, driest, drizzle, drizzled, drizzles, drizzly, drought, droughts, dry, dryline, fahrenheit, flood, flooded, flooding, floods, flurries, flurry, fog, fogbow, fogbows, fogged, fogging, foggy, fogs, forecast, forecasted, forecasting, forecasts, freeze, freezes, freezing, frigid, frost, frostier, frostiest, frosts, frosty, froze, frozen, gale, gales, galoshes, gust, gusting, gusts, gusty, haboob, haboobs, hail, hailed, hailing, hails, haze, hazes, hazy, heat, heated, heating, heats, hoarfrost, hot, hotter, hottest, humid, humidity, hurricane, hurricanes, icy, inclement, landspout, landspouts, lightning, lightnings, macroburst, macrobursts, meteorologic, meteorologist, meteorologists, meteorology, microburst, microbursts, microclimate, microclimates, millibar, millibars, mist, misted, mists, misty, moist, moisture, monsoon, monsoons, mugginess, muggy, nor'easter, nor'easters, noreaster, noreasters, overcast, parched, parching, precipitation, rain, rainboots, rainbow, rainbows, raincoat, raincoats, rained, rainfall, rainier, rainiest, raining, rains, rainy, sandstorm, sandstorms, scorcher, scorching, shower, showering, showers, sleet, slicker, slickers, slush, smog, smoggier, smoggiest, smoggy, snow, snowed, snowier, snowiest, snowing, snowmageddon, snowpocalypse, snows, snowy, sprinkle, sprinkling, squall, squalls, squally, storm, stormed, stormier, stormiest, storming, storms, stormy, stratocumulus, stratus, subtropical, summery, sun, sunnier, sunniest, sunny, temperate, temperature, tempest, thaw, thawed, thawing, thaws, thermometer, thunder, thundering, thunderstorm, thunderstorms, tornadic, tornado, tornadoes, tropical, troposphere, tsunami, turbulent, twister, twisters, typhoon, typhoons, umbrella, umbrellas, vane, warm, warmed, warms, weather, wet, wetter, wettest, wind, windchill, windchills, windier, windiest, windspeed, windy, wintery, wintry

A total of 60.1 million weather tweets were identified, representing 2.8% of the sample. Although our empirical analysis focuses on the effect of temperature, we sampled all possible words about weather and then controlled statistically for the effects of other weather variables (specifically precipitation, cloud cover, and relative humidity). This avoids trying to parse the specific subject of each tweet.

We tested our classification using a manual classification of 6,000 tweets. We are particularly concerned with classification accuracy that varies systematically with the variation used to identify parameters in the regression analysis (i.e. the residual variation in temperature after regressing on all control variable and fixed-effects, shown graphically in Supplementary Figure 2). If classification accuracy is systematically different for tweets about unusually hot temperatures compared to unusually cold temperatures, then this could bias our coefficient estimates since correlation between the number of tweets about weather and temperature could be driven by changing classification accuracy not actual changes in the number of weather tweets. In contrast, classification errors that are uniform across the sample, conditional on regression fixed-effects and controls, will add noise but not bias to our estimation.

Therefore, we used a stratified sampling scheme to identify tweets for validation. We first identify county-weeks associated with unusually hot or cold temperatures, conditional on all fixed-effects and controls by identifying county-weeks in the top and bottom 2.5% of the residual distribution from the following regression:

$$T_{cwmys} = \bar{T}_{cwms} + Precip_{cwmys} + Humid_{cwmys} + Cloud_{cwmys} + \log\left(Users_{cwmys}\right) + \theta_y + \vartheta_c + \delta_{ms} + \varepsilon_{cwmys}$$

Where T_{cwmys} is the average maximum temperature in county c in state s in week-of-year w in month-of-year m, in year y, $\bar{T}_{cwms}$ is the average over the reference period (1981-1990) for that county for that week, $Precip_{cwmys}$, $Humid_{cwmys}$, and $Cloud_{cwmys}$ are controls for average precipitation, relative humidity, and cloud cover in that county in that week, $Users_{cwmys}$ is a control for the number of Twitter users in that county week, θ_y is a year fixed-effect, ϑ_c is a county fixed-effect, and δ_{ms} is a state-month fixed-effect. The county-weeks in the tails of this residual distribution are those with largest influence in the estimation of the effect of temperatures and temperature anomalies on posting about the weather (Supplementary Methods, Section 2). Therefore, contrasting the classification accuracy for tweets from county-weeks in the hot and cold tails of this distribution allows us to identify any systematic errors that will bias estimation of the effect we are interested in. This is therefore the focus of our validation exercise.

Using this sample of high leverage county-weeks, we randomly selected 3,000 tweets each from the set of hot and cold county-weeks, evenly divided between those we classified as weather tweets and those we classified as not about weather. This sample was classified manually into weather / not weather tweets using workers on Amazon Mechanical Turk. Each worker classified 150 unique tweets and each tweet was classified by 3 different workers (for a total of 18,000 classifications). The modal of the three classifications was used for the validation analysis.

Supplementary Table 1 shows the results of this validation exercise. Although the fraction of false positives in our automated classification is high (~46%), there is no evidence for systematic differences in the classification accuracy for hot vs cold temperatures. This suggests that classification errors should not strongly bias our results, except in that they introduce measurement error and so may bias results towards zero, meaning we would be reporting under-estimates of the true effect. The false negative fraction is negligible (<0.5%) and is the same at both tails of the distribution.

2. Regression Analysis

The general regression specification used in this paper is as follows:

$$\begin{aligned}\log(W_{cwmys}) &= f\left(T_{cwmys}, \bar{T}_{cwms}\right) + Precip_{cwmys} + Humid_{cwmys} + Cloud_{cwmys} \\ &\quad + \log\left(Users_{cwmys}\right) + \delta_{ms} + \theta_y + \vartheta_c + \varepsilon_{cwmys}\end{aligned}$$

The dependent variable is the log of the number of weather tweets in county c, in week w, in month m, in year y, in state s. (For clarity in subsequent equations, the month and year subscripts are omitted). Using logs requires us to drop any county by week observations that have no weather tweets. In total this is 55,279 county weeks, or 12.9% of the initial sample. The remaining sample size is 373,625 county weeks. The logged dependent variable means the estimated coefficients have a proportional, not absolute effect, on the number of tweets (i.e. the estimated marginal effect is in terms of % change in the number of tweets). This is important given the very different number of tweets in different counties and weeks.

The number of weather tweets is modeled as a function of maximum temperature (T_{cwmys}) and an average of temperature in previous years in that county at that time of year ($\bar{T}_{cwms}$). The exact functional form used varies and is described below. Additional control variables are the average daily precipitation ($Precip_{cwmys}$), the average relative humidity ($Humid_{cwmys}$), and the average % cloud cover ($Cloud_{cwmys}$). We control for the large differences in the number of Twitter users across counties using log of the number of users in each county and week ($Users_{cwmys}$). A set of fixed effects control for unobserved variation: state by month-of-year fixed effects (δ_{ms}) control for any state-specific intra-annual seasonal differences, year fixed-effects (θ_y) control for average differences across years in the sample (2014, 2015, and 2016) related to, for instance, Twitter penetration, and county fixed-effects (ϑ_c) control for all unobserved, time-invariant differences between counties. Supplementary Figure 2 shows graphically how these fixed-effects determine the residual variation in temperature used to identify our model coefficients.

The regressions are estimated using OLS using the lfe R package and residuals (ε_{cwmys}) are clustered at the state level, allowing for both spatial correlation between counties in the same state and for correlation within a state over time. Control variables, fixed-effects and treatment of standard errors are common across all regressions presented in this paper. They are omitted for clarity in the description of specific functional forms of the temperature response below, but are included in all estimations.

2.1 Interactions Model

Our first set of models allow the effect of temperature to differ as a function of reference and recent temperatures using a set of interaction terms in the estimating equation.

We allow the effect of temperatures to vary non-linearly with the reference (1981-1990) climatology, which accounts for the fact that people's response to weather might be mediated by the kinds of conditions that might be expected in that location at that time of year:

$$\log(W_{cwy}) = \beta_1 T_{cwy} + \beta_2 T^2{}_{cwy} + \beta_3 B_{cw} + \beta_4 B^2{}_{cw} + \beta_5 T_{cwy} B_{cw} + \beta_6 T^2{}_{cwy} B_{cw} + \beta_7 T_{cwy} B^2{}_{cw} + \beta_8 T^2{}_{cwy} B^2{}_{cw}$$

Where B_{cw} is the average temperature of county c in week-of-year w in the reference period and other variables are as defined above. This specification fully interacts both the linear and squared terms of the actual temperature (T_{cwy}) and the reference temperature (B_{cw}), allowing both the location of the quadratic minimum (or maximum) as well as its steepness to vary non-linearly with reference temperature. Results are shown in Figure 2a and Supplementary Table 2, column 2.

A robustness check includes higher order temperature terms that allows for a more flexible response. Specifically, we fit a quadratic in observed temperature, and allow this quadratic to vary flexibly with reference temperature:

$$\log(W_{cwy}) = \beta_1 T_{cwy} + \beta_2 T^2{}_{cwy} + \beta_3 T^3{}_{cwy} + \beta_4 T^4{}_{cwy} + \beta_5 B_{cw} + \beta_6 B^2{}_{cw} + \beta_7 B^3{}_{cw} + \beta_8 T_{cwy} B_{cw} + \beta_9 T^2{}_{cwy} B_{cw} + \beta_{10} T^3{}_{cwy} B_{cw} + \beta_{11} T_{cwy} B^2{}_{cw} + \beta_{12} T^2{}_{cwy} B^2{}_{cw} + +\beta_{13} T_{cwy} B^3{}_{cw}$$

Results from this more flexible specification are shown in Supplementary Figure 3. Findings are qualitatively similar to that from the quadratic model in that we recover the same U-shape with minima relatively close to the reference value. The model does show some asymmetry in terms of declining marginal response at very hot temperatures (possibly consistent with air-conditioner penetration at hot temperatures).

To test the effect of decadal climate trends in mediating the effect of observed temperatures, we add the difference between reference (1981-1990) and recent (average of the previous 5 years) periods as an explanatory variable to the quadratic model. This specification uses the exogenous variation shown in Figure 1 to test whether counties that have had recent experience of unusually hot or cold temperatures (relative to the reference period) respond to weather differently than counties that have not. Our specification is:

$$\log(W_{cwy}) = \beta_1 T_{cwy} + \beta_2 T^2{}_{cwy} + \beta_3 B_{cw} + \beta_4 B^2{}_{cw} + \beta_5 T_{cwy} B_{cw} + \beta_6 T^2{}_{cwy} B_{cw} + \beta_7 T_{cwy} B^2{}_{cw} + \beta_8 T^2{}_{cwy} B^2{}_{cw} + \beta_9 (R_{cwy} - B_{cw}) + \beta_{10} (R_{cwy} - B_{cw}) B_{cw} + \beta_{11} (R_{cwy} - B_{cw}) T_{cwy}$$

Where $(R_{cwy} - B_{cw})$ is the difference in the county week temperature between the recent and reference periods.

Results of this model are shown in Supplementary Table 2, column 3. All estimated effects are statistically significant in the expected direction. Recent experience of warming (i.e. positive $(R_{cwy} - B_{cw})$) increases the number of weather tweets at cold temperatures (positive β_9) but decreases it at hot temperatures (negative β_{11}) (i.e. cold temperatures have become more remarkable and hot temperatures less remarkable), with that effect mediated in the expected direction by reference temperatures (positive β_{10}).

2.2 Dynamic Non-Linear Model

A finite-dynamic lag model is used to estimate the timescale on which perceptions of weather events adjust[6,7]. For this specification we focus on anomalies relative to the reference period (i.e. we defined the temperature anomaly as $A_{cwy} = T_{cwy}$-B_{cw}) and allow the response to vary flexibly as a function of the magnitude of the current temperature anomaly and the history of previous anomalies experienced in that county at that time of year. Because the response to particular temperature anomalies differs depending on whether the temperature is cool or warm (i.e. the response is not symmetric about the reference for all temperatures, see Figure 2a), our preferred model splits the data in order to estimate separate responses for the coolest and warmest third of the sample.

The dynamic non-linear model estimates an interaction between two smooth functions – one of the magnitude of the anomaly and one of lagged history of exposure to anomalies:

$$\log(W_{cwy}) = f(A_{c,w,y-k}) * g(k)$$

Where $A_{c,w,y-k}$ is the temperature anomaly experienced in county c in week-of-year w, k years ago. Values of k range between 0 (i.e. current temperature) and 15 (i.e. temperature 15 years ago). Functions f() and g() are smooth, continuous functions and their interaction allows the for the effect of a particular temperature anomaly to vary non-linearly and to vary as a function of how long-ago it was experienced. Our preferred specification uses a cubic polynomial for f() and a cubic spline with two internal knots for g() (knots at 0, 1.3, 4.4 and 15). The former uses 3 degrees of freedom and the latter 4, so the interaction surface estimated in the regression uses 12 degrees of freedom. Decay of the effect of temperature anomalies that we identify in Figure 3, with opposite effects for warm and cold anomalies, is robust to these choices (Supplementary Figure 8). The model is estimated including all controls and fixed-effects described above and standard errors are clustered at the state level.

3. Applying the Learning Model

The dynamic non-linear model shows evidence of a relatively rapid decay in the influence of temperature anomalies as a function of repeated exposure to those anomalies, both for the full sample (Supplementary Figure 7) and for the cooler third of the reference temperatures (Figure 3). This is consistent with people using relatively recent experience of temperature to set their expectations of, and consequently their response to, the current temperature. We define a learning model based on these empirical estimates that describes how previous temperatures appear to be weighted in current expectations, and therefore how expectations adjust dynamically in response to changing temperatures.

We focus on results for the cooler temperatures to parameterize the learning model because they are both more precisely estimated and easier to interpret than results using the full sample. Supplementary Figure 5b shows the estimated lagged effect of hot and cold temperature anomalies (the cumulative sum of these coefficients are shown in Figure 3). We define the "learning period" as the years during which experience of past temperature anomalies reverses the effect of the current anomaly (i.e. during which there is evidence for adjustment of expectations). For both warm and cold anomalies, this is found to be the period between 2 and 8 years ago (Supplementary Figure 5b).[1]

We parameterize our learning model as the weighted sum of temperature anomalies experienced during the learning period, with weights given by the relative magnitude of the estimated lagged coefficients. In other words, the subjectively-defined, moving baseline is given by:

$$\tilde{B}_{cwy} = \sum_{k=2}^{8} w_k T_{c,w,y-k}$$

$$w_k = \frac{\hat{\beta}_k}{\sum_{j=2}^{8} \hat{\beta}_j}$$

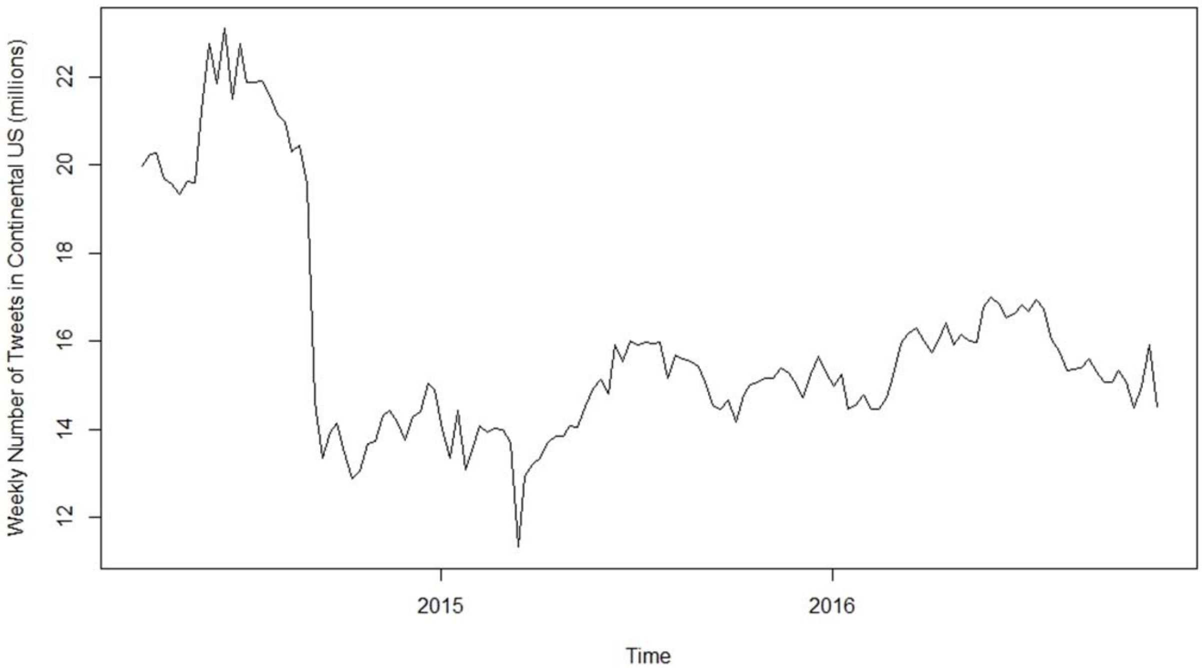
Where $\hat{\beta}_k$ is the estimated effect of the temperature anomaly k years ago (Supplementary Figure 5b). Weights are calculated for both +6 and -6 degree temperature anomalies and are found to be almost identical. Weights based on the +6 degree coefficients are used for the analysis. Since $\tilde{B}_{cwy}$ is a non-linear function of regression coefficients, standard errors are calculated from the estimated variance-covariance matrix using the delta method[8].

The learning model is applied to population-weighted (constant 2015 population distribution) annual average temperatures over the continental United States from 1850 – 2016[4] to give the average change over the industrial period. Because of the eight year lag required for calculating the shifting baseline, this gives perceptual baselines for the period 1858 – 2017 (Supplementary Figure 10).

Temperature anomalies are calculated for the 21st century based on 40 simulations from 1920 to 2100 with the CESM under RCP 8.5[9]. Population-weighted averages are taken over the continental United States (2015 distribution). Rolling perceptual baselines are calculated for the period 1950 – 2100 based on the estimated learning model and then temperature anomalies are calculated on an annual basis relative both to the 1987-2017 average and to the rolling perceptual baseline.

[1] In other words, experiencing a cold (hot) anomaly in just one year results in substantially more (less) weather posts relative to no temperature anomaly in that year. But if the same cold (hot) anomaly was also experienced during the "learning period" (i.e. between 2 and 8 years ago), the response is dampened, consistent with people learning from their experiences in this time period to set their expectations about current temperature.

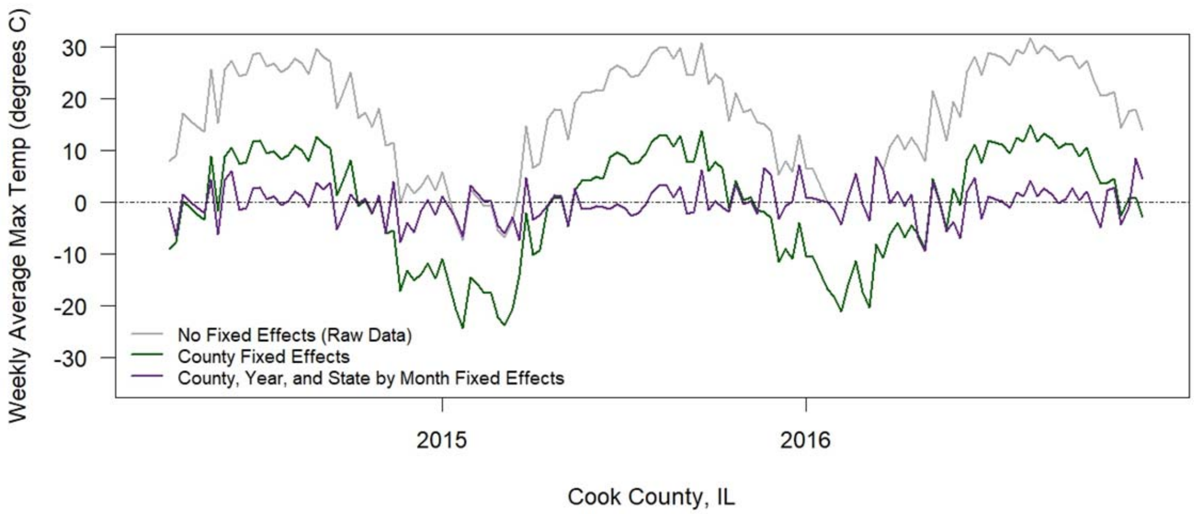
Supplementary Material



Supplementary Figure 1: Number of tweets per week geolocated in the continental United States

Hot Anomalies		Manual Classification	
		Weather	Not Weather
Automated Classification	Weather	806	694
	Not Weather	6	1494
Cold Anomalies		Manual Classification	
		Weather	Not Weather
Automated Classification	Weather	811	689
	Not Weather	5	1495

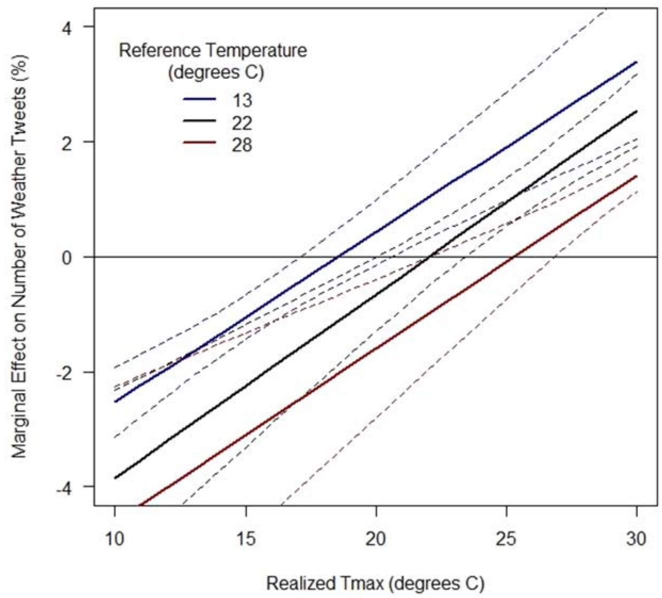
Supplementary Table 1: Results of the manual validation of 6,000 tweet classifications. Tweets were randomly selected from county-weeks with unusually hot and cold temperatures after controlling for all regression controls as well as reference temperature. Each tweet was classified by three different people and the modal classification used in the validation. Additional information in Supplementary Methods.



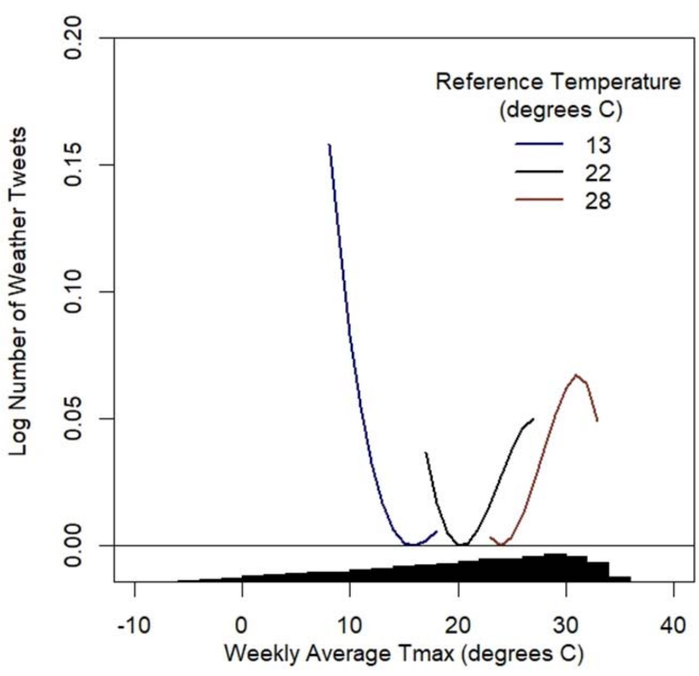
Supplementary Figure 2: Graphical depiction of residual variation in temperature used in the regression model, for Cook County IL. Raw temperature values are shown in grey. County fixed-effects remove the mean for each county over the period of twitter data to center the temperatures around zero (green line). State by month-of-year fixed-effects remove the seasonality for the state. This residual variation (purple line), interacted with average temperatures in the reference and recent time periods is used to identify model coefficients.

	Naïve Model	Informed Model	Change Model
Tmax	-2.952e-2*** (4.000e-3)	-1.800e-2*** (2.692e-3)	-1.808e-2*** (2.676e-3)
$Tmax^2$	6.941e-4*** (7.991e-5)	7.702e-4** (2.431e-4)	8.663e-4*** (2.506e-4)
Reference		-4.707e-3 (4.054e-3)	4.424e-3 (4.930e-3)
$Reference^2$		1.108e-3** (3.989e-4)	1.064e-3* (4.172e-4)
Tmax * Reference		-3.496e-3*** (5.268e-4)	-3.839e-3*** (5.415e-04)
$Tmax^2$ * Reference		7.912e-5*** (1.073e-5)	8.124e-5*** (1.054e-5)
Tmax * $Reference^2$		5.065e-5* (2.138e-5)	5.360e-5* (2.112e-5)
$Tmax^2$ * $Reference^2$		-1.887e-6*** (4.011e-7)	-1.915e-6*** (3.974-07)
Change (Recent – Reference)			1.917e-2*** (3.967e-3)
Tmax * Change			-1.670e-3*** (4.590e-4)
Reference * Change			1.147e-3* (4.977e-4)
Adjusted R^2 (projected model):	0.291	0.294	0.294
F-Test of Informed vs Naïve Model:	19.23*** (4, 48 dof)		
F-Test of Change vs Informed Model:	26.32*** (3, 48 dof)		

Supplementary Table 2: Regression results comparing a naïve model excluding the effect of reference temperatures, an informed model that allows the response to temperature to differ depending on reference temperatures for that county for that time of year, and a change model that adds the change in temperature between reference and recent periods as an explanatory variable. Dependent variable is the logged number of weather tweets. All specifications include controls for mean precipitation, relative humidity, % cloud cover, and the number of Twitter users (logged), as well as county, state by month-of-year, and year fixed effects. Standard errors are clustered at the state level. Significance codes: *p<0.05; **p<0.01; ***p<0.001



Supplementary Figure 3: Marginal effect of 1 degree warmer temperatures on the number of weather tweets for three reference temperatures (approximately the 25th, 50th, and 75th percentiles of reference temperatures in our sample). This the gradient of the curves shown in Figure 2a. Dashed lines show the 95% confidence intervals.

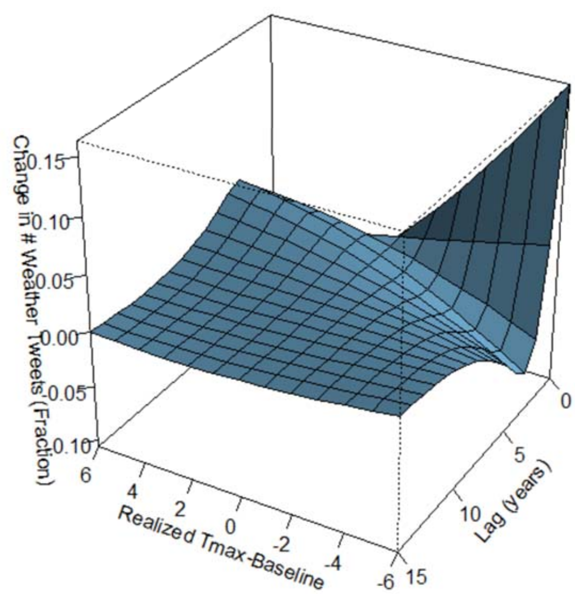


Supplementary Figure 4: As Figure 2a, response curves are from a model allowing for an asymmetric (quartic) response to temperature.

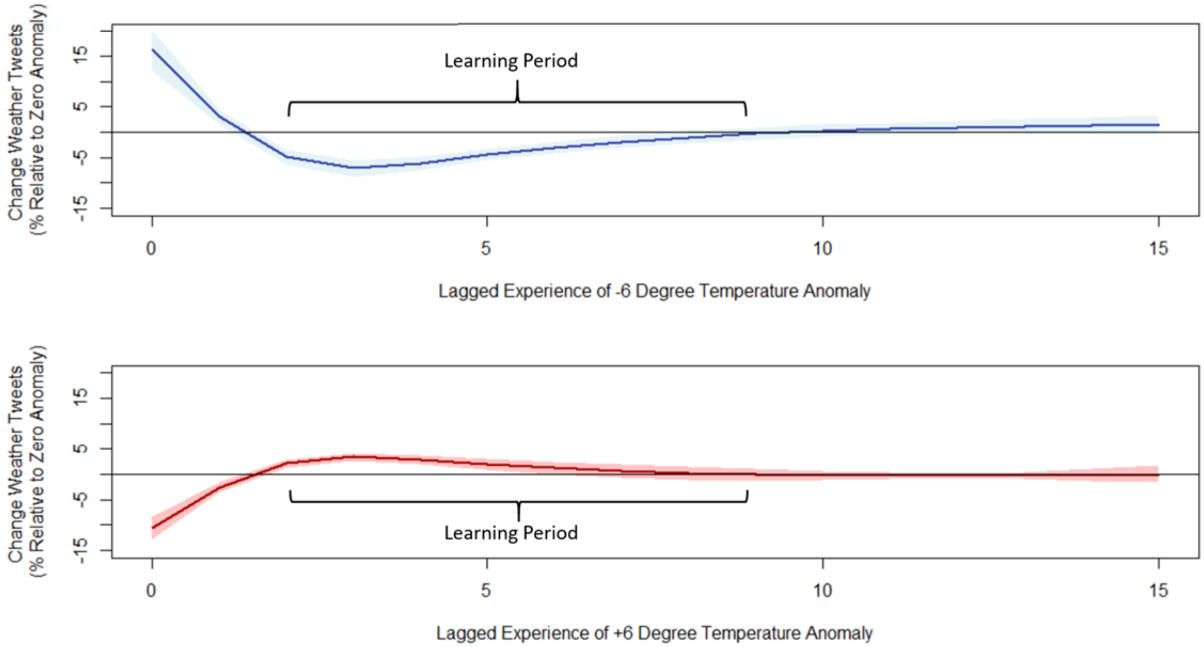
		Full Sample	Coldest Third	Hottest Third
Adjusted R^2 (projected model):		0.289	0.261	0.226
F Test Temperature Anomaly Terms:		19.23 (12, 48 dof, p<1e-12)	49.33 (12, 48 dof, p<1e-12)	6.46 (12, 48 dof, p<1e-10)
Number of Observations:		373,625	124,528	124,568
Spline Basis	Temperature:	Cubic, no internal knots	Cubic, no internal knots	Cubic, no internal knots
	Lags:	Cubic, 2 knots equally spaced in log space	Cubic, 2 knots equally spaced in log space	Cubic, 2 knots equally spaced in log space
Number of Temperature Anomaly Variables in Cross-Basis:		12	12	12
Number of Lags:		15	15	15

Supplementary Table 3: Summary of non-linear dynamic lag models fit using the dlnm package in R[10]. The same model is fit using the full sample and divided into the coolest third of reference temperatures and the hottest third of reference temperatures. Temperature anomalies are defined relative to the reference period (1981-1990). Dependent variable is the logged number of weather tweets. All specifications include controls for mean rainfall, relative humidity, % cloud cover, and the number of Twitter users (logged), as well as county, state by month-of-year, and year fixed effects. Standard errors are clustered at the state level.

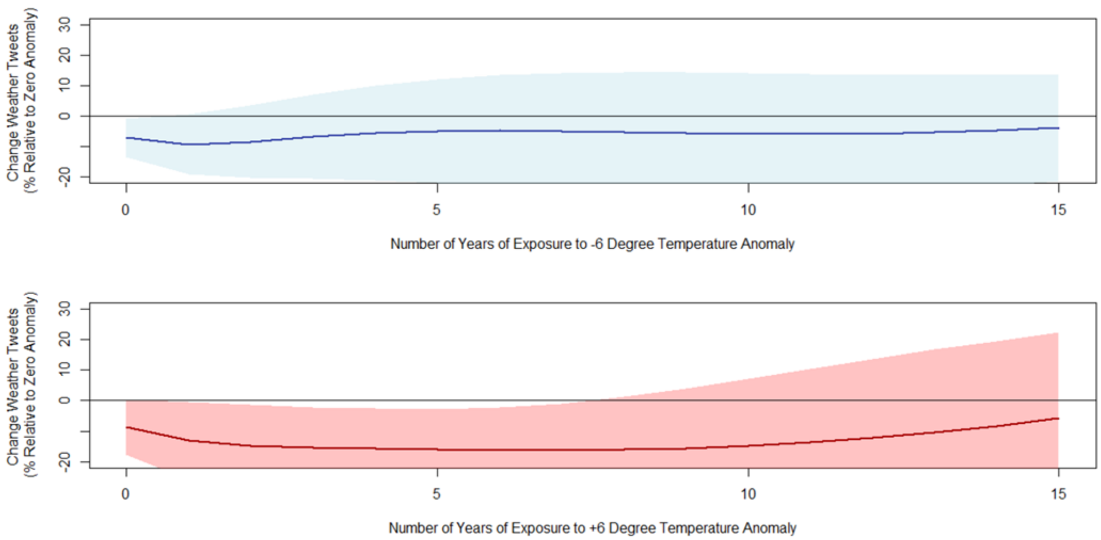
a)



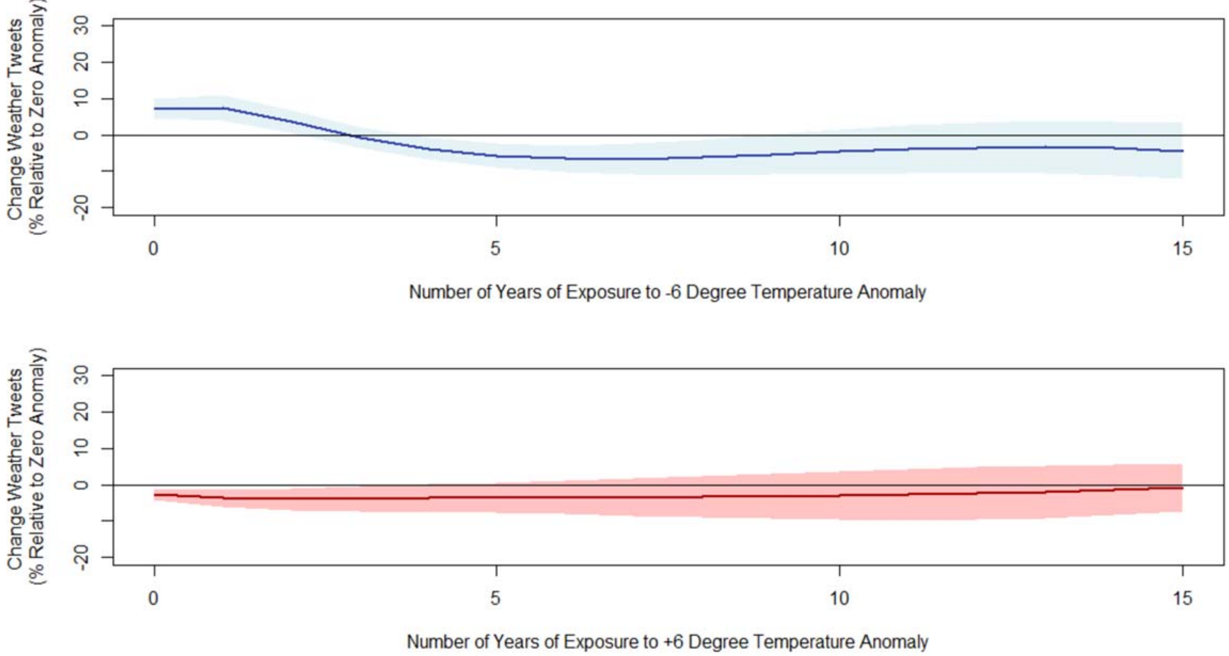
b)



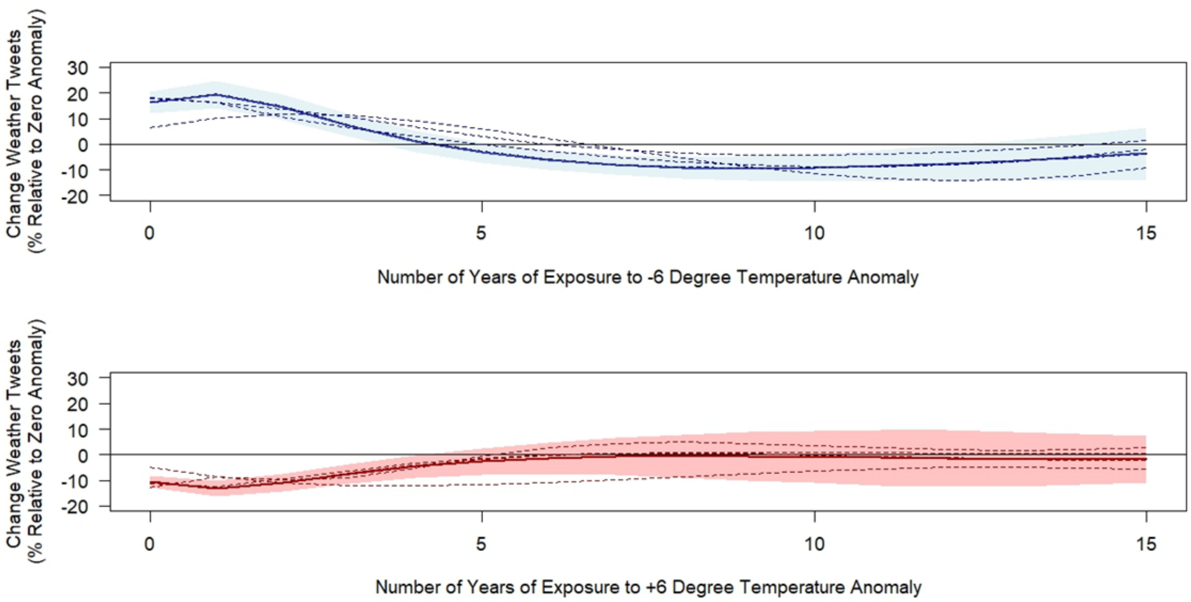
Supplementary Figure 5: a) Three-dimensional representation of the fitted response surface for the cold sample of reference temperatures. Shows the fractional change in the number of tweets about weather, relative to an instantaneous zero anomaly as a function of both temperature anomaly (Realized Tmax – Reference) and lagged exposure. b) As Figure 3 in the main text except showing the estimated lagged effects rather than the cumulative sum of lagged effects with 95% confidence intervals. The marked areas show the period defined as the “learning period” and used to estimate the weights of the learning model. The learning period is defined as the period over which experience of temperature anomalies reduces the instantaneous effect of those anomalies. For both warm and cold anomalies, this is the period between 2 and 8 years ago.



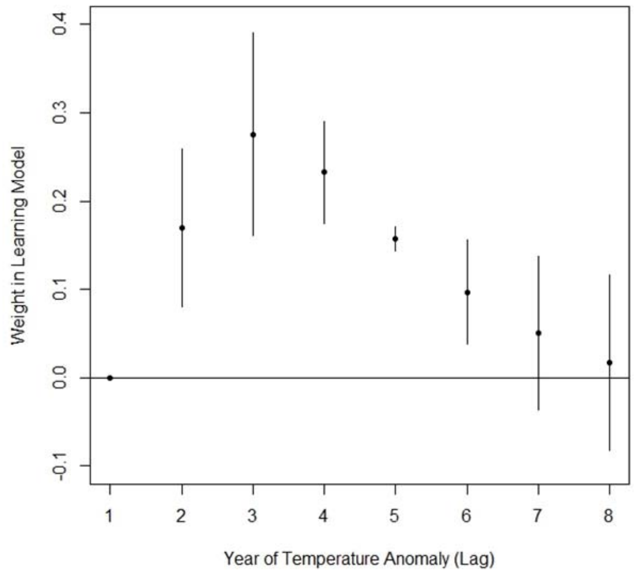
Supplementary Figure 6: As Figure 3 in main text, but estimated using the sample in the hottest third of reference temperatures.



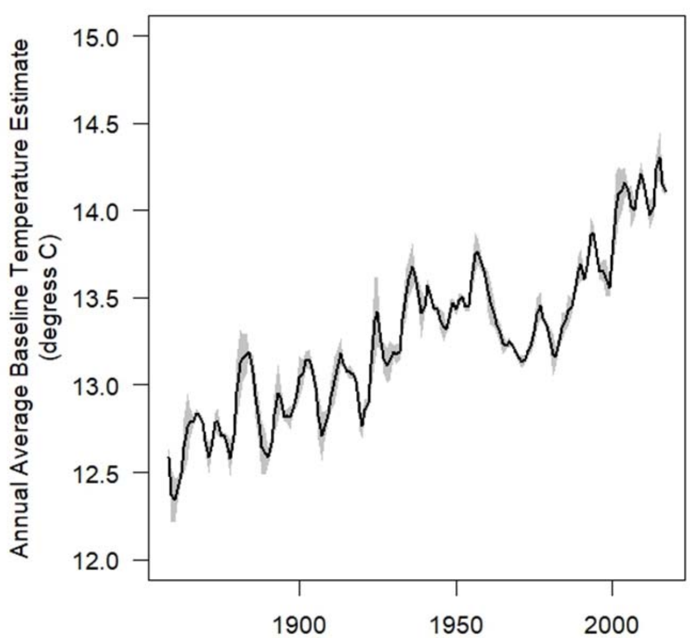
Supplementary Figure 7: As Figure 3 in main text, but estimated using the full sample.



Supplementary Figure 8: As Figure 3 in main text, but with additional results from alternative specifications of the dynamic lag model (dashed lines). The four alternative specifications include a quadratic rather than cubic response to the temperature anomaly and changing the number of internal knots in the lag spline from 2 to 1, 3, and 4.



Supplementary Figure 9: Weights in learning model based on coefficient values during the learning period defined using the estimated lag coefficients shown in Supplementary Figure 4. Standard errors show the 95% confidence interval calculated using the delta method.



Supplementary Figure 10: Population-weighted annual average baseline temperatures over the continental United States, updated based on the learning process derived from the dynamic lag model and observed temperatures[3]. Shaded areas show the 95% confidence interval associated with statistical uncertainty in the coefficients of the lag model, calculated using the delta method.

References

1. PRISM. Oregon State University.
2. Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M. & Potter, G. L. NCEP-DOE AMIP-II Reanalysis (R2). *Bull. Am. Meteorol. Soc.* **83,** 1631–1643 (2002).
3. BEST. Berkeley Earth Data. (2018). Available at: http://berkeleyearth.org/data/. (Accessed: 5th March 2018)
4. CIESIN, University, C. & CIAT. *Gridded Population of the World, Version 3: Population Density Grid*.
5. KNMI. Climate Explorer. (2015). Available at: climexp.knmi.nl. (Accessed: 1st July 2016)
6. Gasparrini, A. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *J. Stat. Softw.* **43,** 1–20 (2011).
7. Almon, S. The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica* **33,** 178–196 (1965).
8. Oehlert, G. W. A Note on the Delta Method. *Am. Stat.* 27–29 (1992).
9. Kay, J. E., Deser, C., Philips, A. S., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., *et al.* The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bull. Am. Meteorol. Soc.* **96,** 1333–1349 (2013).
10. Gasparrini, A., Armstrong, B. & Scheipl, F. Package ‘dlnm’. (2017). Available at: https://cran.r-project.org/web/packages/dlnm/dlnm.pdf.