

MIT Open Access Articles

Temperature and Cognitive Performance: Evidence from Mental Arithmetic Training

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Krebs, B. Temperature and Cognitive Performance: Evidence from Mental Arithmetic Training. Environ Resource Econ (2024).

As Published: 10.1007/s10640-024-00881-y

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/155223>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution





Temperature and Cognitive Performance: Evidence from Mental Arithmetic Training

Benjamin Krebs¹

Accepted: 30 April 2024
© The Author(s) 2024

Abstract

With rising average temperatures and extreme heat events becoming more frequent, understanding the ramifications for cognitive performance is essential. I estimate the effect of outside air temperature on performance in mental arithmetic training games. Using data from 31,000 individuals and 1.15 million games played, I analyze frequent engagement in a cognitively challenging task in a non-stressful and familiar environment. I find that, above a threshold of 16.5 °C, a 1 °C increase in outside air temperature leads to a performance reduction of 0.13%. The effect is mostly driven by individuals living in relatively cold areas, who are less adapted to hot temperatures.

Keywords Air temperature · Cognitive performance · Climate change

JEL Classification Q54 · J24

1 Introduction

Climate change entails vast economic and social consequences. An extensive body of research has identified adverse effects of rising temperatures on economic growth and production, agricultural output, labor productivity, mortality, conflict, migration, and

For comments and suggestions, I am grateful to Benedikt Janzen, Michael Keller, Christopher Knittel, Simon Luechinger, Matthew Neidell, and the participants of the 2023 Summer Conference of the Association of Environmental and Resource Economists. Further, I thank Silvia Ferrini and two anonymous reviewers for many great suggestions on how to improve this paper and for giving me the opportunity to revise it.

✉ Benjamin Krebs
bkrebs@mit.edu
<http://ceep.mit.edu/krebs/>

¹ Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, MIT Building E19-411, 400 Main Street, 4th Floor, Cambridge, MA 02142-1017, USA

others.¹ Under the prospect of rising average temperatures and more frequent extreme weather events in most places of the world, this literature naturally predicts an intensification of these effects.

A critical component of any human activity is cognitive performance. It determines labor productivity and serves as a prerequisite for human capital accumulation. As laboratory experiments in neurological research document the detrimental effect of heat on human physiology and brain functioning,² understanding how temperature affects cognitive performance in the real world and, ultimately, what climate change means for this relationship is essential.

In this paper, I estimate the effect of outside air temperature on cognitive performance. I use data from an online mental arithmetic training game called Raindrops from the Lumosity platform, with over 31,000 individuals and 1.15 million games played in 748 three-digit U.S. ZIP Codes between 2015 and 2019. The data include only paying subscribers who regularly train their mental arithmetic skills. This context provides a unique opportunity to investigate how temperature affects people in a non-stressful, recurring, and familiar setting, a context currently not covered in the literature.

I find that hot temperatures significantly reduce mental arithmetic performance. Using piecewise-linear regressions allowing for different slopes in two distinct temperature ranges, above a threshold value of 16.5 °C, a 1 °C increase in the average air temperature during the 24 h preceding a play lowers the score by 0.084 correct answers or 0.13%. Below the threshold, the corresponding coefficient is insignificant and close to zero. 3 °C-bins regressions confirm these findings. They indicate that people attain significantly lower scores when playing in the bins above 21 °C, compared to the bin with the highest average performance (15–18 °C). The scores decrease by 0.484 (21–24 °C), 0.588 (24–27 °C), and 0.954 (≥ 27 °C) correct answers. These figures amount to a drop of 0.73%, 0.90%, and 1.46%, respectively. Similar to the linear regressions, I do not find consistent evidence for adverse effects of cold temperatures.

The results exhibit important heterogeneity, implying potential adaptive behavior. Individuals from relatively cold ZIP Codes (below-median 2015–2019 average temperatures) experience a larger performance drop than individuals from relatively hot ZIP Codes (above-median 2015–2019 average temperatures). The estimate from the piecewise-linear regressions for the above-threshold range is -0.142 for cold ZIP Codes (-0.21%) but only -0.042 for hot ZIP Codes (-0.07%) and is not statistically significant. The results from the bin regressions corroborate these findings. I discuss the possible interpretations of this heterogeneity.

Since subscribers choose when to use the software, this raises concerns about bias from two types of potential selection mechanisms, at the extensive and the intensive margin. First, people might be less likely to play when temperatures are high. This will only be an issue if more temperature-sensitive individuals (those who experience a larger performance drop due to extreme temperatures) are less likely to play than others during such temperature extremes. In this case, the coefficients will be biased toward zero. While I cannot directly test this, I show that individuals do not engage less on hot days. However, they do use the software more often when it is cold. Due to this

¹ E.g., Schlenker and Roberts (2009); Deschênes and Greenstone (2011); Lobell et al. (2011); Barreca (2012); Dell et al. (2012); Hsiang et al. (2013); Graff Zivin and Neidell (2014); Burke et al. (2015); Mis-sirian and Schlenker (2017); Mullins and White (2020); Heutel et al. (2021); Somanathan et al. (2021); Carleton et al. (2022). See Carleton and Hsiang (2016) for an overview.

² See (Hancock and Vasmatazidis 2003; Wang et al. 2021) for reviews.

potential selection issue at the lower end of temperatures, I focus on the results for hot temperatures.

Second, people might play fewer times on particularly hot or cold days. Since they improve their performance with the number of plays on a given day, the average score from a very hot or very cold day will be lower than from a mild day if they actually play fewer times due to extreme temperatures. This potential issue is independent of the selection of people who are affected. It would bias the coefficients downwards, i.e., my results would overestimate the true adverse effect. However, I show that neither hot nor cold temperatures seem to impact the intensive margin substantially.

I identify five main contributions this paper makes. First, key components of the setting I analyze make this paper arguably more representative of people's everyday work lives than previous papers. The individuals in this study engage in a non-stressful but cognitively challenging task. They presumably take it seriously, nonetheless, as they pay a monthly subscription fee of USD 15. Further, many of the observed individuals use the software very frequently. The previous literature (discussed in Sect. 2) investigates individuals' response to temperature either in stressful, non-everyday situations with potentially long-lasting effects on their career path or in unfamiliar settings in which they play a more passive role. Most studies observe their population only infrequently.

Second, by investigating mental arithmetic, I focus on numeracy skills, an essential measure of cognitive performance. Using data from the Program for the International Assessment of Adult Competencies (PIAAC), Hanushek et al. (2015) find that a one-standard-deviation increase in numeracy skills is associated with a 28 percent wage increase among prime-age workers in the U.S. Thus, I analyze a component of cognitive performance that determines productivity and wages.

Third, as I observe the individuals' number of wrong answers per game, I add to our understanding of the temperature-cognition relationship by differentiating between problem-solving speed and the error rate. This is an important distinction, as speed decline and error proneness have distinct ramifications. In many settings, errors are arguably more costly than speed, e.g., in medical procedures, air traffic control, or the assembly of consumer goods. If a higher error rate is the main driver of a lower cognitive performance, protecting people from heat exposure in settings where mistakes are costly seems critical. This paper is the first to make this distinction.

Fourth, my analysis covers a large age range from 18 to 80 years from all over the contiguous U.S. The previous literature almost exclusively focuses on adolescents. Therefore, the average effects I estimate in this paper are arguably more indicative of how temperature affects the general population. The broad age range also allows me to look into effect heterogeneity between younger and older people.

Fifth, the individuals observed in this analysis play during a wide range of temperature realizations. The lowest percentile of the temperature distribution is -11.1 °C and the highest percentile is 31.5 °C. This enables me to investigate both hot and cold temperatures in the same context. Most previous studies analyze only one end of temperature extremes, predominantly heat.

2 Previous Literature

This study contributes to the recent literature on the effect of outside temperature on performance in cognitive tests. With one exception, thus far, all of these papers have investigated the temperature-cognition relationship either in survey tests or in academic exams.

This literature uses performance data from individuals completing their tasks indoors while temperature is measured outdoors. Equivalent to the concentration-response vs. exposure-response function in pollution contexts (Graff Zivin and Neidell 2013), estimates from these papers amount to a combination of the physiological effect of temperature on performance and the impact of short-term adaptive behavior, like air conditioning, intended to avert the consequences of the physiological effect. While the two components cannot be separated, such reduced-form estimates are highly relevant, as they give insight into the real-world temperature-performance relationship. They investigate whether the physiological effect materializes despite defense investments, and they represent a lower bound of the societal costs of extreme temperatures, excluding the costs of adaptive behavior. Although observing both outdoor and indoor temperatures would be valuable, non-experimental studies usually lack information on indoor temperatures.

In contrast, experimental studies use controlled settings to improve our understanding of the dynamics behind the physiological effect. Hancock and Vasmatzidis (2003) and Wang et al. (2021) summarize major contributions showing, e.g., that temperature-induced performance impacts depend on task complexity, task type, and their requirement of different cognitive functions, such as memory and attention. Despite providing important insight, this strand of literature falls short of indicating whether the reported physiological effects actually materialize in the real world. Air conditioning has been shown to prevent the negative consequences of heat (Park 2022). Furthermore, Cook and Heyes (2020) find that cold temperatures negatively affect university students' exam results despite constant indoor temperatures. These results underpin the necessity for outdoor temperature-indoor performance studies.

I identify three papers analyzing survey tests. Graff Zivin et al. (2018) estimate the effect of temperature on mathematics and reading assessments from the U.S. National Longitudinal Survey of Youth (NLSY) based on 8003 children interviewed in their homes between 1988 and 2006. The authors find a decline in math performance for contemporaneous temperatures above 21 °C (significant above 26 °C) but not for reading, and they identify smaller, imprecisely estimated effects for temperature realizations between tests. Garg et al. (2020) use math and reading test results from the Annual Status of Education Report (ASER), a repeated cross-section dataset at the district level of 4.5 million tests across every rural district in India over the period of 2006 to 2014, and the Young Lives Survey (YLS), a panel from the state of Andhra Pradesh with tri-annual household visits between 2002 and 2011. They report 10 extra days with an average daily temperature above 29 °C, compared to 15–17 °C, to reduce math and reading test performance by 0.03 and 0.02 standard deviations in ASER, and they find qualitatively similar effects in YLS. Yi et al. (2021) employ verbal and math test scores from 5404 individuals above 40 years old from the China Health and Retirement Longitudinal Study (CHARLS), covering two rounds in 2013 and 2015. They find a 2.4 °C increase in interview day heat stress, a measure for temperature over 25 °C, to lower math scores by 2.2% and verbal scores by 2.4%. Further, the authors examine cumulative effects, showing that not only the current-day temperature matters but also the accumulation of heat over time.

Investigating survey data, Graff Zivin et al. (2018), Garg et al. (2020), and Yi et al. (2021) concentrate on tasks participants perform rarely. NLSY and CHARLS are bi-annual surveys, YLS is tri-annual, and ASER is a repeated cross-section dataset. Thus, the tasks in ASER are most likely entirely unfamiliar to participants. As beginners, they do not solve tasks at their personal performance limits. Graff Zivin et al. (2020) show that heat has stronger effects on the more proficient, a finding that is consistent with the impacts of another environmental stressor, air pollution (Graff Zivin et al. 2020; Krebs and Luechinger 2024). Further, participants in surveys take a passive role. While they can decline to take part, they do not actively decide to perform these tasks. Presumably, people are less engaged in these tasks and take them less seriously than tasks they perform, e.g., on their job. Similar to practice and proficiency, the effect of temperature on performance might depend on people's engagement.

In contrast, this study focuses on a task people are familiar with. Many individuals use the software day to day and accumulate hundreds of completed games. This frequent engagement makes this data more representative of everyday duties since most tasks we do are repeated. Furthermore, since users pay a USD 15 monthly subscription fee, presumably, they aim to train and improve their skills and take the task seriously. As susceptibility to environmental stressors is likely to depend on proficiency and engagement, studying performance in this context represents a valuable complement.

Eight studies investigate the effect of temperature on college admission or other academic exams. Cho (2017) estimate the effect of summer heat on school-level scores of 1.3 million Korean college entrance exams taken in November in the years 2009 to 2013. They find an additional day above 34 °C daily maximum temperature to lower math and English scores by 0.0042 and 0.0064 standard deviations, respectively, relative to the 28–30 °C range. Graff Zivin et al. (2020) analyze overall test scores from 14 million administered exams at the National College Entrance Examination (NCEE) in China between 2005 and 2011. They report a linear effect of 0.68% lower scores from a 2 °C increase in exam period temperature. Cook and Heyes (2020) investigate results from 67,000 students taking over 600,000 exams at the University of Ottawa between 2007 and 2015. Unlike the other papers, they concentrate on cold temperatures and find a 10 °C colder outdoor temperature to cause a reduction in performance of 8% of a standard deviation. Park et al. (2020) use Preliminary Scholastic Aptitude Test (PSAT) test scores of 10 million American high school students who took the test at least twice between 2001 and 2014. Controlling for test-day weather, the authors find a 1% reduction of an average student's learning gain over a school year from a 1 °F hotter school year. The effect is non-linear, with hotter years having higher impacts. They also provide evidence for air conditioning to alleviate these effects. Park et al. (2021) employ Programme for International Student Assessment (PISA) test scores from 58 developed and developing countries between 2000 and 2015 and district-level annual mathematics and English test scores from the Stanford Educational Data Archive (SEDA) for over 12,000 U.S. school districts. They report an additional day above 26.7 °C in the three years before the exam to reduce scores by 0.18% standard deviations in PISA and an additional day above 26.7 °C to lower achievement by 0.04% standard deviations in the SEDA data. Melo and Suzuki (2023) analyze natural sciences, social sciences, Portuguese, and mathematics exam scores from 8 million Brazilian high-school seniors in about 1,800 municipalities. They find a one-standard-deviation temperature increase to decrease exam scores by 0.036 standard deviations. Roach and Whitney (2021) use SEDA math and English scores from 2008 to 2015 and find non-linear negative effects on achievement. Park (2022) use individual exam-level information from over 4.5 million exam records for almost 1 million students taking the Regents Exams in New York City between

1999 and 2011. They report a one-standard-deviation increase in exam-time temperature (6.2 °F) to decrease performance by 5.5% of a standard deviation. Further, they provide evidence for persistent impacts on educational attainment, showing that on-time graduation likelihood decreases by three percentage points for every one-standard-deviation increase in average exam-time temperature (4.4 °F).

These eight studies generate a holistic picture of how temperature affects children's and adolescents' exam performance and learning. Test-taking students find themselves in a particularly stressful, non-everyday situation with potentially long-lasting effects on their career path. While understanding the role of environmental factors in this context is highly relevant, individuals might be more sensitive to these factors when writing a test than they usually are. From pollution contexts, we know that environmental stressors have a stronger effect the higher the stakes (Künn et al. 2023). Thus, the external validity for less stressful situations people encounter daily might be limited.

To my knowledge, the only paper that does not use survey or academic test data is Bao and Fan (2020), who use Chinese data from an online role-playing game called Dragon Nest from March 1 to March 31, 2011, and find a performance drop for temperatures below 5 °C.

A core feature of almost all papers investigating outdoor temperature and indoor cognitive performance is, with the exception of Yi et al. (2021), they investigate children or adolescents. Thus, the samples are highly selective on a younger population. My sample includes individuals between 18 and 80 years old, although, with a median of 56 years, the sample represents an older population than the U.S., with a median of 39 years.³ This allows me to investigate heterogeneity across age groups. Additionally, none of the mentioned papers investigates both cold and hot temperatures, as most focus on heat only. I observe individuals at both ends of the temperature distribution.

While this study fills some central gaps in the literature, it naturally has shortcomings. The most prominent drawback is sample selection. In contrast to representative surveys, Lumosity users are most likely not a random draw from the overall population. Disclosed characteristics in my data reveal a higher median age and female proportion (61%) than the U.S. The monthly subscription fee suggests relatively wealthy users. Other socioeconomic features might differ as well. This needs to be kept in mind when interpreting the results. Another drawback is the representativeness of mental arithmetic for cognitive challenges people face in their everyday lives. Some tasks may be related quite closely, others not at all. This concern remains despite the somewhat alleviating evidence that numeracy skills are determinants of productivity and wages. Despite these shortcomings, this paper departs from the previous literature in key aspects, thereby contributing crucially to our understanding of the temperature-cognition relationship.

3 Data

3.1 Cognitive Performance

To measure cognitive performance, I use data from Lumosity's online mental arithmetic training game called Raindrops. Individuals have to solve arithmetic problems that fall down in raindrops before they hit the water at the bottom of the screen (see Fig. 4). The

³ <https://www.census.gov/newsroom/press-releases/2023/population-estimates-characteristics.html>, accessed February 14, 2024.

problems disappear when solved correctly. After the third raindrop hits the water, the game is over. The problems get more difficult, and the raindrops fall faster over time.

The data only include individuals who have used the software at least five times. Consistent with the data delivery agreement, I drop the first four plays of every individual because users see a tutorial before the first play but not before any other plays. The data only cover web plays (no mobile apps). The spatial resolution is three-digit ZIP Codes, of which there are 748 included in the data.⁴ I end up with a dataset that includes 31,029 players and 1,151,059 plays from 2015 to 2019.

The main outcome variable is the number of correct answers. In an extension, I also use the error rate, measured as the number of correct entries divided by the total number of entries. As people's performance heavily depends on their play behavior (e.g., how many times they have played, how long ago they last played), I include the following three control variables: the log number of plays an individual has played so far, the log number of plays (+1) an individual has played since taking a break of at least one hour, and the log number of hours (+1) since the last play. Further, I interact each of the three variables with three age range indicators (50–64, 65–74, and 75 and older).

3.2 Air temperature and weather covariates

To generate the outside air temperature variable and weather covariates, I use data from the NOAA National Centers for Environmental Information (NCEI) Global Hourly Integrated Surface Database for the years 2015–2019.⁵ I only consider weather monitoring stations that were operational throughout the whole period, and I include air temperature, air dew point temperature, wind speed, atmospheric pressure, and precipitation. To generate ZIP Code averages, I perform the following three steps: First, I drop all variables with more than 25% missing observations for each station. Second, I interpolate missing values of all available variables at the station level with an inverse distance-weighted average of all stations within a radius of 50 kms and a power parameter of 2. And third, I generate the ZIP Code average from all stations within a ZIP Code. In case there is no station within a ZIP Code, I use the inverse distance-weighted average of all stations within 50 kms of the ZIP Code centroid, again with a power parameter of 2, to attribute the hourly weather variables. I then drop ZIP Codes with more than 25% missing values for any variable.

The main explanatory variable is the average air temperature during the 24 h preceding a play. In a robustness check, I use the average air temperature during the 48 h preceding a play and the average heat index during the 24 h preceding a play as exogenous variables. The covariates are the 24-hour averages of relative air humidity, wind speed, atmospheric pressure, and precipitation and the quadratic function of each of those variables. The 24-hour averages include the current hour when the game was played. If less than 18 h (75%) are non-missing values, I code the observation as missing. I calculate the relative

⁴ The average three-digit ZIP code is about 8497 km² (92 × 92 km), or the size of Idaho County. (There are 902 three-digit ZIP codes in the contiguous U.S. [https://en.wikipedia.org/wiki/List_of_ZIP_Code_prefixes], and the total land area of the contiguous U.S. is 7,663,941 km² [https://en.wikipedia.org/wiki/Contiguous_United_States].) Temperature variation and the resulting measurement error could potentially bias estimates towards zero. However, most observations come from densely populated and, thus, smaller three-digit ZIP codes, diminishing this concern.

⁵ <https://www.ncei.noaa.gov/data/global-hourly/archive/>

Table 1 Summary statistics

| Variable | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------------------|----------------------|-----------------------|----------------------|
| Number of correct answers | 65.417 (37.441) | 68.446 (39.804) | 63.228 (35.476) |
| Error rate | 0.095 (0.071) | 0.092 (0.070) | 0.097 (0.072) |
| Nth play | 254.766 (995.460) | 397.220 (1472.048) | 151.842 (340.635) |
| Nth play since one-hour break | 2.921 (6.814) | 2.978 (7.242) | 2.880 (6.486) |
| Hours since last play | 193.754 (795.622) | 190.326 (782.582) | 196.231 (804.904) |
| Average air temperature past 24 h in °C | 14.076 (9.940) | 9.748 (10.109) | 17.204 (8.545) |
| Average air temperature past 48 h in °C | 14.083 (9.812) | 9.757 (9.956) | 17.209 (8.422) |
| Average heat index past 24 h in °C | 13.429 (11.105) | 8.614 (11.155) | 16.907 (9.679) |
| Average relative humidity past 24 h in % | 67.362 (17.068) | 69.173 (15.632) | 66.053 (17.921) |
| Average wind speed in m/s | 3.033 (1.583) | 3.317 (1.641) | 2.828 (1.506) |
| Average atmospheric pressure past 24 h in hPa | 1016.755 (6.642) | 1016.908 (7.460) | 1016.645 (5.979) |
| Total precipitation past 24 h in mm | 0.091 (0.263) | 0.090 (0.230) | 0.092 (0.284) |
| Observations | 1,151,059 | 482,812 | 668,247 |

Means and standard deviations (in parentheses) of the two main cognitive outcome variables, the variables used to construct the play covariates, the temperature variables, and the variables used to construct the weather variables in the main regression, separately for the full sample and the cold- and hot-ZIP Codes sample

humidity from a function of air temperature and dew point temperature,⁶ and the average heat index according to the National Weather Service.⁷

3.3 Summary statistics

Table 1 shows the means and standard deviations of the main two dependent variables (number of correct answers and the error rate), the variables used to construct the play covariates (Nth play, Nth play since one-hour break, hours since last play), the main independent variable (average air temperature during the 24 h preceding a play), two alternative explanatory variables used in robustness checks (average air temperature during the 48 h

⁶ <https://bmcnoldy.rsmas.miami.edu/Humidity.html>

⁷ https://www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml

and heat index during the 24 h preceding a play), and the variables used to construct the weather controls (average relative humidity, average wind speed, average atmospheric pressure, and total precipitation during the 24 h preceding a play). The table has separate columns for the whole sample, the cold-ZIP Codes sample (below-median 2015–2019 average temperatures), and the hot-ZIP Codes sample (above-median 2015–2019 average temperatures). Figure 5 displays a map of the contiguous United States with different colors for cold and hot ZIP Codes.

The overall average number of correct answers is 65.4. People living in cold ZIP Codes score roughly 5 points higher than those living in hot ZIP Codes, which amounts to 8.3% more correct answers. The difference in the error rate is 5.4% (9.7% erroneous entries in hot ZIP Codes vs. 9.2% in cold ZIP Codes). Both these differences are statistically significant at the 99% level. The average temperature is 14.1 °C. The difference between hot and cold ZIP Codes is 7.5 °C. The heat index variable is slightly lower, mainly due to the cold ZIP Codes sample, which indicates a higher relative humidity in hot ZIP Codes.

4 Identification

I use two different approaches to estimate the effect of outside air temperature on performance in the mental arithmetic game. Equation 1 represents the piecewise-linear regression model.

$$P_{izt} = \beta T_{jt} + \gamma A_{jt} + \delta(T_{jt} \times A_{jt}) + \mathbf{W}'_{zt}\eta + \mathbf{G}'_{izt}\theta + \mathbf{u}_i + \boldsymbol{\tau}_t + \epsilon_{izt} \tag{1}$$

P_{izt} is performance (number of correct answers and error rate) of individual i in ZIP Code z at time t . T_{jt} is the average air temperature during the 24 h preceding the play in °C. A_{jt} is an indicator equal to 1 if the air temperature during the past 24 h was above a certain threshold value. \mathbf{W}'_{zt} is a vector of weather variables described in Sect. 3.2. \mathbf{G}'_{izt} is a vector of play covariates described in Sect. 3.1. \mathbf{u}_i absorbs the individual effects to control for any time-invariant differences between individuals, such as innate ability. $\boldsymbol{\tau}_t$ absorbs year, month-of-year, day-of-week, hour-of-day effects. These fixed effects flexibly control for time trends, seasonal patterns, and differences across the timing of a play between weekdays and the hours of a day. Finally, ϵ_{izt} is the error term.

The two coefficients of interest are β and δ . β represents the effect of temperature on performance below the threshold value. To investigate the effect of temperature on performance above the threshold value, I run a linear combination test for $\beta + \delta \neq 0$ and report the results in the regression tables. The threshold value is 16.5 °C and is determined as the arithmetic mean of the temperature bin with the highest performance from the indicator regression (Eq. 2 below).

A major advantage of Eq. (1) is that the coefficients report average marginal effects over the full temperature range above and below the threshold value. However, the model hinges on the linearity assumption. Therefore, I also estimate the temperature bin regression model presented in Eq. (2).

$$P_{izt} = \mathbf{T}'_{jt}\theta + \mathbf{W}'_{zt}\eta + \mathbf{G}'_{izt}\theta + \mathbf{u}_i + \boldsymbol{\tau}_t + \epsilon_{izt} \tag{2}$$

\mathbf{T}'_{jt} is a vector of temperature bin indicators. Temperature is, as in Eq. (1), the average air temperature during the 24 h preceding the play in °C. The bins are 3 °C temperature steps from 0 to 27 °C, plus one for temperatures below 0 °C and one for temperatures above

27 °C. The bottom bin corresponds to the rounded 10th percentile of the temperature distribution (with the exact value being 0.47 °C). Likewise, 27 °C is the closest rounded number to the 90th percentile that is dividable by 3 (with the exact value being 26.1 °C). The bin with the highest average performance serves as the reference (15–18 °C).

The model in Eq. (2) flexibly allows for non-linearity in the effect of temperature on performance. It is, therefore, a suited complement to Eq. (1) to test the linearity assumption. A potential concern is that the highest bin might be significantly different from other bins by chance, which would make it seem like people perform worse in all other bins. As the linear decline with higher temperatures in Sect. 5.1 does not support this concern, I refer to Eq. (1) as my main results.

5 Results

5.1 Performance

Table 2 shows the coefficients from the piecewise-linear regression model outlined in Eq. (1). The coefficients in the first row (“Air temperature”) correspond to β .⁸ The linear combination test in the second row (“Air temp. + air temp. \times above threshold”) reports the effect of air temperature on performance when temperature is above 16.5 °C. This corresponds to a test of $\beta + \delta \neq 0$ from Eq. (1). I do not report the coefficient from the air temperature \times above threshold interaction, δ , as it is simply equal to the coefficient reported in the linear combination test minus β .

When focusing on all ZIP Codes included in my dataset (column 1), cold temperatures do not seem to affect people’s performance in the Raindrops game. The coefficient of air temperature below the threshold value is positive but very close to zero (0.010) and statistically insignificant. Above the threshold, temperatures negatively affect performance. An increase of 1 °C decreases the number of correct answers by 0.084, which amounts to 0.13%.

To evaluate potential adaptation effects, I run separate regressions for relatively cold and relatively hot ZIP Codes.⁹ Columns 2 and 3 of Table 2 report the respective results. They show that the cold ZIP Codes drive the overall result. In the cold-ZIP Codes sample, a temperature increase of 1 °C lowers the number of correct answers by 0.142, which is 0.21%. The effect for hot ZIP Codes is -0.042 (0.066%) and statistically insignificant.

The baseline result implies that, with rising temperatures and no adaptation, people will perform below their capacity more often. Running separate analyses for cold and hot ZIP Codes gives insight into how adaptation to climate change might mitigate these adverse effects (Dell et al. 2014; Auffhammer 2018). This result suggests that hot ZIP Codes are better equipped to cope with high temperatures. As temperatures rise, adaptation investments in colder regions will potentially reduce this gap. People living in colder regions might, therefore, respond less to hot temperatures in the future, similar to people currently living in hotter regions. However, as climate change will also result in more extreme temperatures, realizations that are currently rare will occur more frequently. Thus, even the better-adapted, hotter regions might experience larger performance drops. As these two

⁸ Note that air temperature is measured in °C and, if not indicated otherwise, is calculated as the average air temperature during the 24 h preceding the play.

⁹ I differentiate based on the ZIP Code mean temperature from 2015 to 2019. I define cold (hot) ZIP Codes as the ones that had a below-median (above-median) average temperature during these years.

Table 2 Air temperature and number of correct answers: piecewise-linear regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|---------------------------------------------------------------------|---------------------|----------------------|-------------------|
| Air temperature | 0.010 (0.021) | 0.010 (0.017) | 0.012 (0.017) |
| Linear combination test: Air temp. + air temp. × above threshold | -0.084** (0.038) | -0.142*** (0.048) | -0.042 (0.030) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from piecewise-linear regressions of the number of correct answers on the average air temperature during the 24 h preceding the play and an interaction with the above-threshold indicator as outlined in Eq. (1). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C, and the linear combination test amounts to the slope above 16.5 °C. Column 1 includes all observations. Columns 2 and 3 show results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

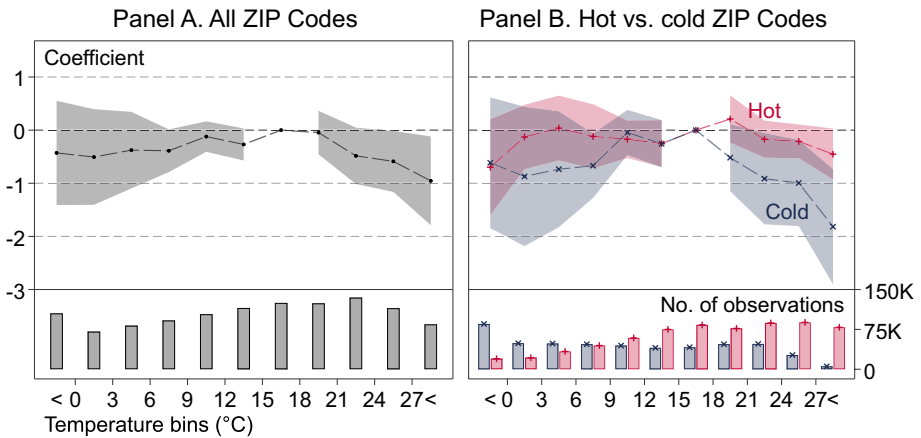


Fig. 1 Air temperature and number of correct answers: 3 °C-bins regressions. Notes: Coefficients with 95% confidence intervals (left y-axis) and number of observations (right y-axis) from regressions of the number of correct answers on 3 °C bins of the average air temperature during the 24 h preceding the play (x-axis) as outlined in Eq. (2), equivalent to Table 3. The standard errors are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in Panel A includes all observations. Panel B shows the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4)

effects run in opposite directions, predictions about how climate change affects our cognitive performance hinge on central assumptions about the degree of potential adaptation.

The temperature bin regressions broadly confirm the results from the piecewise-linear regressions. Panel A of Fig. 1 (equivalent to column 1 of Table 3) shows that the coefficients for the temperature bins below the reference of 15–18 °C are mostly small and insignificant, with two of them being significant at the 90% level. The coefficient from the 18–21 bin is almost zero. Above that, there is a close-to-linear trend with the coefficients and t-values becoming larger in absolute terms. These results support the linearity assumption made in Eq. (1).

Panel B of Fig. 1 (equivalent to columns 2 and 3 of Table 3) confirms the findings from the cold-hot differential piecewise-linear regressions. There are no significant effects for below-reference temperatures, except for one outlier (6–9 °C) for cold ZIP Codes. This outlier does not withstand the robustness checks I present in Sect. 5.3. All bins above 21 °C are statistically significant at the 95% level or higher for cold ZIP Codes. They are roughly twice the size as the coefficients from all ZIP Codes and also seem to follow a linear trend. The results from the hot-ZIP Codes regressions are not significant, except for the top bin, which is significant at the 90% level. While there is a slight downward trend above 19.5 °C, it seems likely that the 18–21 °C bin is an upward outlier.

How do these results compare to the results from the literature? Table 4 summarizes the main results from the relevant papers. I include all papers discussed in Sect. 2 that investigate short-term effects and use a linear approach. The table shows the temperature variable, the outcome variable, the temperature cutoff, the mean of the outcome variable, and the estimated coefficient. From the latter two, I calculate the percentage change of the outcome variable for a 1 °C increase in temperature (last column) to make a comparison across papers more convenient.

The five papers that meet the mentioned requirements all estimate larger effects than my main result by a magnitude of 3 to 13. What might be a possible explanation for this? In contrast to previous studies using academic tests, the stakes in this setting are low. As discussed in Sect. 2, people might be less affected by heat in non-stressful, everyday settings. Another potential explanation is age. Most of the discussed studies focus on adolescents, while the median age in this study is 56 years and ranges from 18 to 80 years. As my dataset includes only relatively few adolescents, I lack the power to test for differential effects between narrow age groups. Yet, in an extension in Sect. 5.3, I estimate effect heterogeneity across two age groups (below and above the median age of 56 years) and find no evidence for differences between those groups.

I see two limiting factors concerning the interpretations of why I estimate smaller effects than the previous literature. First, all studies differ in a multitude of dimensions. E.g., the choice of temperature cutoff might affect the estimated slope. Second, some of the previous studies focus on a relatively constrained location. These locations might be better or worse adapted to heat compared to the average effect across the contiguous U.S., I estimate in this paper.

5.2 Selection

One central difference to the previous literature is that individuals using the training software choose when to play. In contrast, college admission exams have predetermined dates, while surveys usually allow for some participant discretion. In this section, I provide two selection tests mentioned in Sect. 1.

The first test concerns the extensive margin. I construct a dataset that includes an observation for every hour from each individual's first to last play in the data. The observations contain the respective ZIP Code's weather variables and the individual's play covariates. The main outcome variable is an indicator equal to 1 if an individual played during a specific hour and 0 otherwise. Table 5 shows the summary statistics for this dataset. As mentioned, I am not able to test whether susceptible individuals are more or less likely to refrain from playing than others on a particularly hot or cold day. Instead, I test whether people are, on average, more (or less) likely to play depending on the temperature.

Table 6 shows the results from the bin regressions similar to the main result. I multiply the outcome variable by 100 to make the coefficients more readable. As the bin regressions imply a different underlying functional form compared to the main results, I do not estimate a linear model. The results in column 1 for the whole sample suggest that temperatures do not affect the probability of playing, except for the coldest bins. The coefficient of 0.0107 for temperatures below 0 °C means that people have a 3.06% higher probability of playing ($0.0107 / 100 / 0.0035$) compared to the reference bin (15–18 °C). Separating between cold and hot ZIP Codes, I find no effects for hot temperatures either. The positive effect of cold temperatures seems to kick in earlier in cold ZIP Codes.

These findings suggest that there is no extensive margin selection for hot temperatures. However, people are more likely to play when it is cold. This could potentially bias the results if, e.g., it is primarily non-susceptible people who are more likely to play during cold weather. In that case, a potential negative effect of cold temperatures would be biased toward zero. Therefore, I only cautiously interpret the null result from Sect. 5.1.

To address intensive margin selection, I construct a dataset with an observation for every hour an individual uses the software. As the dependent variable, I use the log number of plays for each of these observations. Thus, a coefficient of 0.01 corresponds to a 1% increase in the number of plays. This allows me to test whether people play more (or less) often, conditional on playing at all when temperatures are extreme. The summary statistics for this dataset are in Table 7.

Table 8 shows the 3 °C-bins regression results. Similar to the extensive margin, I do not estimate the linear model, as the bin regressions do not support a corresponding functional form. The results do not provide consistent evidence for a change in the number of plays for hot temperatures. While the coefficient for the 18–21 °C bin is significantly negative, purely driven by the hot ZIP Codes, there is no trend, and the pattern indicates that this is likely an outlier. People in cold ZIP Codes play somewhat more during cold weather, but the extent is small. All in all, I find little evidence for selection at the intensive margin.

5.3 Robustness

To test the robustness of my estimates, I run three alternations of my baseline results: First, instead of averaging temperature over 24 h, I take the average during the 48 h preceding each play. The degree to which past temperatures affect people's performance has an underlying function that is ex-ante unknown. A 24-hour cutoff is random and neglects that temperatures from longer ago might still affect people's performance. If temperatures in the 24-to-48-hour range do not affect performance, the coefficients should be closer to zero than in the main results regression due to attenuation bias.

The results are robust to using the 48-hour average. Table 9 shows the linear regression results, and Table 10 the coefficients from the bin regressions. The coefficients for

the linear combination test are somewhat larger in absolute terms for the whole sample (-0.093 vs. -0.084) as well as for the cold-ZIP Codes sample (-0.177 vs. -0.142). This suggests that temperatures in the 24-to-48-hour range are indeed relevant.

In the second robustness check, I use a heat index instead of air temperature. The effect of hot temperatures on the human body varies with relative humidity.¹⁰ The heat index, a function of temperature and relative humidity, takes that into account directly. Again, the results stay qualitatively unchanged. I present the piecewise-linear regression results in Table 11 and the bin regression results in Table 12. The coefficient from the cold-ZIP Code sample is somewhat smaller (-0.117 vs. -0.142), while the coefficient from the hot-ZIP Code sample is slightly larger (-0.050 vs. -0.042), both in absolute terms.

Finally, I use the natural logarithm of 1 plus the number of correct answers as the dependent variable. As temperature effects on the number of correct answers might depend on individuals' baseline performance or average score, this specification would be a more suited functional form. Again, there are no qualitative changes to the results. The coefficients for the linear combination test in Table 13 can be interpreted as percentage changes for a 1 °C change in temperature above the threshold of 16.5 °C. The figures are smaller than in Sect. 5.1 (-0.08% vs. 0.13% for the whole sample, -0.11% vs. -0.21% for the cold-ZIP Codes sample, and -0.06% vs. 0.07% for the hot-ZIP Codes sample). While the cold-ZIP Codes sample coefficient is less precisely estimated, the precision of the hot-ZIP Codes sample coefficient increases. The bin regression results in Table 14 also look fairly similar.

5.4 Extensions

The main finding of this paper raises the question about the underlying mechanism. Do people simply solve the arithmetic problems more slowly and, thus, achieve fewer points, or do they make more mistakes and, thereby, lose time to find the right answer?

One way to investigate these mechanisms is to look at the error rate, i.e., the number of erroneous entries per total number of entries. Figure 2 (equivalent to Table 16) reports the bin regressions results. None of the estimated bin indicators returns a significant coefficient, and there seems to be no trend, neither for cold nor hot temperatures. This is confirmed in the linear regression results in Table 15. These results provide suggestive evidence that hot temperatures affect the solving speed rather than the error rate.

Another central aspect is how cumulative heat exposure affects cognitive performance. Specifically, is ongoing heat for multiple days worse than a single heat day alone? Fig. 3 (equivalent to Table 17) summarizes the results of different heat period lengths. I run regressions similar to Eq. (2), except I use only one dummy variable equal to 1 if the average air temperature during the 24 h preceding a play is greater than 21 °C (baseline). This is the lower end of the bin with a negative coefficient at the 90% level in the main results. The coefficients for this dummy are -0.489 (all ZIP Codes), -0.666 (cold ZIP Codes), and -0.263 (hot ZIP Codes), all significant at the 95% level. These numbers correspond to a performance reduction of 0.75%, 0.97%, and 0.42%, respectively.

To disentangle, I calculate the average temperature of seven different time periods: period 1 is hours 0 to 23 preceding a play (the same as above and throughout the paper), period 2 is hours 24 to 47 preceding a play, etc. Accordingly, period 7 is hours 144 to 167. I then generate three indicator variables based on the average temperature of these seven

¹⁰ <https://www.weather.gov/ama/heatindex>

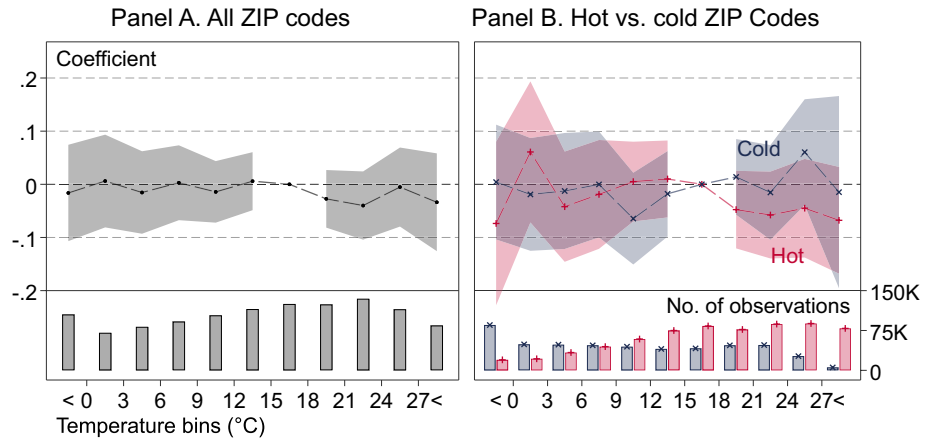


Fig. 2 Air temperature and error rate: 3 °C-bins regressions. Notes: Coefficients with 95% confidence intervals (left y-axis) and number of observations (right y-axis) from regressions of the error rate (number of incorrect answers / total answers) on 3 °C bins of the average air temperature during the 24 h preceding the play (x-axis) as outlined in Eq. (2), equivalent to Table 15. The standard errors are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in Panel A includes all observations. Panel B shows the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4)

periods. The first indicator is equal to 1 if the average temperature in period 1 but not 2 or in periods 1 and 2 but not 3 was greater than 21 °C. This indicator refers to hot temperatures for one or two days, but not longer (“1-2 days...” in Fig. 3). The second indicator is equal to 1 if the average temperature in periods 1 to 3 but not 4, or in periods 1 to 4 but not 5, or in periods 1 to 6 but not 7 was greater than 21 °C. This indicator refers to hot temperatures for at least three but not more than six days (“3-6 days...” in Fig. 3). Finally, the third indicator is equal to 1 if the average temperature in all seven periods was greater than 21 °C. This indicator refers to hot temperatures for at least seven consecutive days. I then run similar regressions to the baseline but with these three heat period length indicators instead of just one indicator. The reference is a day with an average temperature below or exactly 21 °C.

The estimated coefficient strictly increases with the length of the heat period, from -0.220 (-0.34%) for 1 or 2 days to -0.568 (-0.87%) for 3 to 6 days and -0.789 (-1.21%) for 7 or more days with average temperatures above 21 °C. This pattern is consistent for both cold and hot ZIP Codes. The performance drop is -0.413 (-0.60%), -0.857 (-1.25%), and -1.075 (-1.57%) for cold ZIP Codes, and -0.014 (-0.02%), -0.27 (-0.43%), and -0.479 (-0.76%) for hot ZIP Codes, respectively. While these coefficients are not statistically different from each other, all coefficients increase with the length of the heat period.

Analyzing the channels for this potential increase in the effect of heat on performance is beyond the scope of this paper. At least two mechanisms could be at play: First, a recent paper shows that people sleep less when temperatures are high (Minor et al. 2022). The observed pattern would be what to expect if the lack of sleep is a principal reason

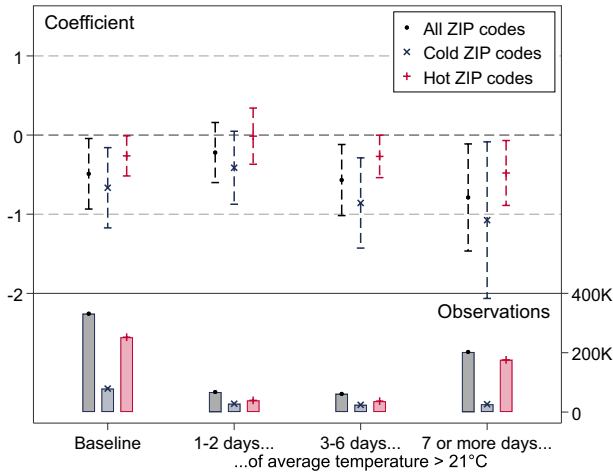


Fig. 3 Effect accumulation. Notes: Coefficients with 95% confidence intervals (left y-axis) and number of observations (right y-axis) from regressions of the number of correct answers on an indicator = 1 if the average temperature was above 21 °C during different temporal periods before a play (x-axis), equivalent to Table 17. The standard errors are clustered on ZIP Codes. The baseline is 24 h preceding a play. The reference is a day with an average temperature below or exactly 21 °C. I run separate regressions for all observations and the cold- and the hot-ZIP Codes sample. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4)

for reduced cognitive performance and continuing sleep deprivation makes it worse. The second channel is more mechanical: Buildings have temperature inertia. Without air conditioning, homes will usually be hotter after more days of heat. Hence, the effect might worsen with consecutive hot days.

The final extension I provide is on effect heterogeneity across age. I estimate separate regressions for people who are below, or exactly, 56 years old, and people who are above 56 years old, as 56 is the median age. Table 18 shows the piecewise-linear results and Table 19 show the temperature bins results, respectively. While the coefficients from the piecewise-linear regressions reveal a slightly larger effect for older people, the bins regressions coefficients are somewhat higher (in absolute terms) for younger people. The differences are generally very small, and the coefficients are close to the baseline results for both age groups. They are, however, less precisely estimated. These results do not provide any evidence for effect heterogeneity.

6 Conclusion

I estimate the effect of outside air temperature on cognitive performance for a large sample in the U.S. using a rich dataset on mental arithmetic training. Hot temperatures reduce people's performance, with larger effects in colder, less adapted regions. The driver for the lower performance seems to be slower problem-solving rather than higher error proneness. Consecutive hot days worsen the effect. I do not find any significant effects of cold temperatures. The results for the cold temperature range should be taken with a grain of salt due to potential selection issues.

This paper fills a gap in the literature by focusing on non-stressful, familiar, and repeated tasks, a context arguably more representative of everyday life situations than academic exams and survey tests. The estimated coefficients are small compared to the previous literature. Using the estimates from Hanushek et al. (2015), they translate to 0.11% lower wages for every 1 °C increase in temperature in colder areas.¹¹ However, these estimates ignore the cost of adaptive behavior and the psychological costs of adaptation. Further, the representativeness of mental arithmetic for tasks people perform on their jobs and in their everyday lives is limited. Therefore, more research is needed to better understand the drivers behind temperature sensitivity in cognitively demanding settings and, ideally, to consider behavioral and psychological costs.

This paper raises the question of climate change adaptation potential. Heat affects individuals' cognitive performance in colder areas more than in hotter areas, implying a policy focusing on currently under-adapted, colder regions to reduce temperature vulnerability through air conditioning, building insulation, and cultivating urban vegetation. However, as hotter regions will experience temperature realizations that are currently not observed, this paper and the relevant literature underestimate the potential costs of future climate change for these regions. Therefore, to address climate change adaptation policies solely towards colder regions would be myopic. While providing important insight into how heat affects individuals today, discussing optimal policy is beyond the scope of this paper.

Appendix. Figures

See Figs. 4, 5.

¹¹ This is, of course, a very crude approach assuming similar short- and long-term effects.

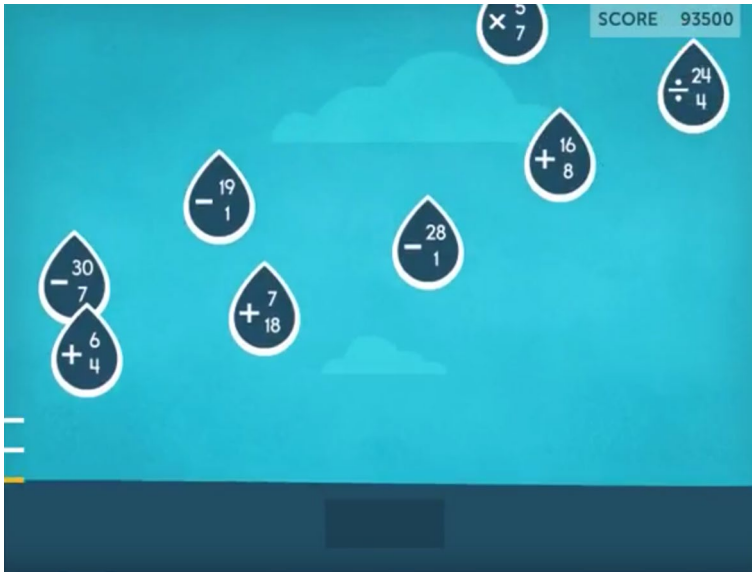


Fig. 4 The Raindrops game. Source: <https://www.youtube.com/watch?v=M8mASg4KOS0>, accessed June 3, 2021, as used in Krebs and Luechinger (2024)

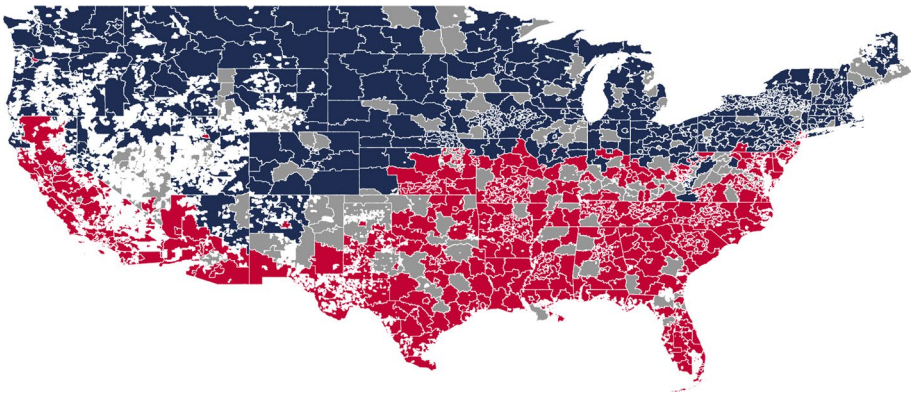


Fig. 5 Hot and cold three-digit ZIP Codes. Notes: Map of three-digit ZIP Codes of the contiguous United States with the cold-ZIP Codes sample (below-median 2015–2019 average temperatures) in blue and the hot-ZIP Codes sample (above-median 2015–2019 average temperatures) in red. Gray areas are ZIP Codes without a play observation. The average temperature of each ZIP Code is based on the NOAA data described in Sect. 3.2

Appendix. Tables

See Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20.

Table 3 Air temperature and number of correct answers: 3 °C-bins regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------------------|---------------------|----------------------|--------------------|
| Air temp. < 0 °C | -0.428 (0.500) | -0.614 (0.625) | -0.699 (0.459) |
| Air temp. 0–3 °C | -0.506 (0.458) | -0.872 (0.667) | -0.130 (0.308) |
| Air temp. 3–6 °C | -0.376 (0.367) | -0.736 (0.555) | 0.037 (0.311) |
| Air temp. 6–9 °C | -0.388* (0.205) | -0.670** (0.309) | -0.119 (0.308) |
| Air temp. 9–12 °C | -0.121 (0.146) | -0.045 (0.216) | -0.171 (0.180) |
| Air temp. 12–15 °C | -0.270* (0.154) | -0.256 (0.226) | -0.253 (0.224) |
| Air temp. 18–21 °C | -0.046 (0.209) | -0.518 (0.321) | 0.208 (0.227) |
| Air temp. 21–24 °C | -0.484* (0.270) | -0.913** (0.437) | -0.171 (0.178) |
| Air temp. 24–27 °C | -0.588** (0.293) | -0.995** (0.412) | -0.212 (0.162) |
| Air temp. ≥ 27 °C | -0.954** (0.425) | -1.817*** (0.551) | -0.452* (0.244) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from regressions of the number of correct answers on 3 °C bins of the average air temperature during the 24 h preceding the play as outlined in Eq. (2) and depicted in Fig. 1. The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 4 Results from previous literature

| Paper | Temperature variable | Outcome variable | Temperature cutoff | Mean of outcome var | Coefficient (Std. err.) | Δ outcome var. in % |
|---------------------------|------------------------------------------|------------------------------------------------|--------------------|---------------------|-------------------------------|----------------------------|
| This paper | Average temp. during 24 h preceding play | Mental arithmetic performance | 16.5 °C | 65.4 | -0.084 (0.038) | -0.13 |
| Graff Zivin et al. (2018) | Same-day number of degree days | National Longitudinal Survey of Youth | 21 °C | 49.7 | -0.211 (0.090) | -0.42 |
| Graff Zivin et al. (2020) | Same-day number of degree days | China's National College Entrance Examination | 14 °C | 519 | -0.0034 ⁱ (0.0004) | -0.34 |
| Park (2022) | Average temp. around test time | Regents Exams in New York City | None ⁱⁱ | 64.9 | -0.290 ⁱⁱⁱ | -0.45 |
| Melo and Suzuki (2023) | Temp. during exam | Exame Nacional do Ensino Médio in Brazil | None ⁱⁱ | 506.5 | -0.879 ⁱⁱⁱ | -1.73 |
| Yi et al. (2021) | Same-day heat stress degree days | China Health and Retirement Longitudinal Study | 25 °C | 2.8 | -0.025 (0.013) | -0.89 |

ⁱ Coefficient and standard error from log-lin regression. ⁱⁱ Tests taken during relatively warm weather. ⁱⁱⁱ No standard error is provided because the coefficient is translated from a regression on the standardized test score

Notes: Comparison of the main result of this paper to the main result of five other papers that focus on short-term effects and apply a linear approach. The table shows the temperature variable, the outcome variable, the temperature cutoff, the mean of the outcome variable, and the estimated coefficient. I calculate the percentage change of the outcome variable in percent for a 1 °C increase in temperature, displayed in the last column

Table 5 Summary statistics of extensive margin selection

| Variable | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------|---------------------|----------------------|---------------------|
| Played | 0.0035 (0.0589) | 0.0036 (0.0596) | 0.0034 (0.0583) |
| Temperature past 24 h in °C | 14.5942 (9.7932) | 10.3918 (10.1184) | 17.5256 (8.3951) |
| Observations | 226,926,601 | 93,249,007 | 133,677,594 |

Means and standard deviations (in parentheses) of the extensive margin selection variable (indicator variable = 1 if an individual played during a specific hour), and the average air temperature during the 24 h preceding a play in °C. See Sect. 5.2 for a sample construction description

Table 6 Air temperature and probability of playing: 3 °C-bins regressions

| Dependent variable: play | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|--------------------------|-----------------------|-----------------------|----------------------|
| Air temp. < 0°C | 0.0107*** (0.0035) | 0.0120*** (0.0043) | 0.0130** (0.0055) |
| Air temp. 0–3 °C | 0.0069** (0.0031) | 0.0081** (0.0040) | 0.0085* (0.0048) |
| Air temp. 3–6 °C | 0.0058** (0.0028) | 0.0093** (0.0039) | 0.0032 (0.0039) |
| Air temp. 6–9 °C | 0.0005 (0.0027) | 0.0054 (0.0039) | −0.0034 (0.0036) |
| Air temp. 9–12 °C | −0.0001 (0.0025) | 0.0086*** (0.0032) | −0.0059* (0.0032) |
| Air temp. 12–15 °C | 0.0008 (0.0021) | 0.0069* (0.0037) | −0.0024 (0.0026) |
| Air temp. 18–21 °C | −0.0025 (0.0017) | −0.0022 (0.0028) | −0.0024 (0.0022) |
| Air temp. 21–24 °C | −0.0024 (0.0022) | −0.0008 (0.0036) | −0.0034 (0.0028) |
| Air temp. 24–27 °C | −0.0008 (0.0025) | −0.0008 (0.0041) | −0.0015 (0.0029) |
| Air temp. ≥ 27 °C | −0.0007 (0.0032) | 0.0049 (0.0057) | −0.0030 (0.0034) |
| Observations | 226,926,601 | 93,249,007 | 133,677,594 |
| Individuals | 30,081 | 12,344 | 17,737 |
| ZIPs | 745 | 373 | 372 |

Results from regressions of play (indicator variable = 100 if an individual played during a specific hour) on 3 °C bins of the average air temperature during the 24 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 7 Summary statistics of intensive margin selection

| Variable | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------|---------------------|---------------------|---------------------|
| Number of plays | 1.4541 (1.5234) | 1.4475 (1.5149) | 1.4590 (1.5295) |
| Temperature past 24 h in °C | 14.0711 (9.9291) | 9.7705 (10.1390) | 17.2051 (8.4982) |
| Observations | 789,305 | 332,720 | 456,585 |

Notes: Means and standard deviations (in parentheses) of the intensive margin selection variable (the number of plays an individual engaged in during an hour in which the individual played), and the average air temperature during the 24 h preceding a play in °C. See Sect. 5.2 for a sample construction description

Table 8 Air temperature and frequency of playing: 3 °C-bins regressions

| Dependent variable: Log no. of plays | All ZIP Codes | Cold ZIP Codes | Hot ZIP Codes |
|-----------------------------------------|-----------------------|-----------------------|-----------------------|
| Air temp. < 0°C | -0.0010 (0.0032) | 0.0052 (0.0047) | -0.0070 (0.0052) |
| Air temp. 0–3 °C | 0.0073** (0.0030) | 0.0127*** (0.0047) | 0.0040 (0.0045) |
| Air temp. 3–6 °C | 0.0023 (0.0028) | 0.0062 (0.0045) | 0.0013 (0.0040) |
| Air temp. 6–9 °C | 0.0007 (0.0024) | 0.0063* (0.0038) | -0.0028 (0.0037) |
| Air temp. 9–12 °C | 0.0026 (0.0021) | 0.0042 (0.0034) | 0.0025 (0.0031) |
| Air temp. 12–15 °C | -0.0015 (0.0019) | -0.0038 (0.0032) | 0.0004 (0.0025) |
| Air temp. 18–21 °C | -0.0051** (0.0021) | -0.0037 (0.0034) | -0.0065** (0.0027) |
| Air temp. 21–24 °C | -0.0022 (0.0021) | -0.0051 (0.0035) | -0.0010 (0.0027) |
| Air temp. 24–27 °C | -0.0013 (0.0029) | 0.0006 (0.0049) | -0.0013 (0.0034) |
| Air temp. ≥ 27 °C | -0.0038 (0.0034) | -0.0042 (0.0075) | -0.0021 (0.0038) |
| Observations | 789,305 | 332,720 | 456,585 |
| Individuals | 30,081 | 12,344 | 17,737 |
| ZIPs | 745 | 373 | 372 |

Results from regressions of the log number of plays (the log number of plays an individual engaged in during the current hour) on 3 °C bins of the average air temperature during the 24 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 9 Air temperature past 48 h and number of correct answers: piecewise-linear regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------------------|------------------|------------------|------------------|
| Air temperature | 0.013 (0.023) | 0.016 (0.019) | 0.013 (0.018) |
| Linear combination test | -0.093** | -0.177*** | -0.045 |
| Air temp. + air temp. × above threshold | (0.041) | (0.054) | (0.030) |
| Observations | 1,150,343 | 482,500 | 667,843 |
| Individuals | 31,027 | 12,706 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from piecewise-linear regressions of the number of correct answers on the average air temperature during the 48 h preceding the play (in °C) and an interaction with the above-threshold indicator as outlined in Eq. (1). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C, and the linear combination test amounts to the slope above 16.5 °C. Column 1 includes all observations. Columns 2 and 3 show results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 10 Air temperature past 48 h and number of correct answers: 3 °C-bins regressions

| Dependent variable: No. of correct answers | All ZIP Codes | Cold ZIP Codes | Hot ZIP Codes |
|-----------------------------------------------|---------------------|----------------------|--------------------|
| Air temp. < 0°C | -0.327 (0.468) | -0.519 (0.512) | -0.688 (0.550) |
| Air temp. 0–3 °C | -0.330 (0.533) | -0.784 (0.659) | 0.328 (0.299) |
| Air temp. 3–6 °C | -0.364 (0.328) | -0.616 (0.439) | -0.087 (0.308) |
| Air temp. 6–9 °C | -0.264 (0.176) | -0.487** (0.215) | -0.051 (0.295) |
| Air temp. 9–12 °C | -0.009 (0.161) | 0.034 (0.202) | -0.011 (0.191) |
| Air temp. 12–15 °C | -0.074 (0.111) | 0.024 (0.150) | -0.091 (0.157) |
| Air temp. 18–21 °C | 0.158 (0.213) | -0.282 (0.212) | 0.372 (0.299) |
| Air temp. 21–24 °C | -0.477* (0.288) | -1.124** (0.446) | -0.048 (0.190) |
| Air temp. 24–27 °C | -0.553** (0.246) | -0.925*** (0.349) | -0.204 (0.178) |
| Air temp. ≥ 27 °C | -0.945** (0.390) | -1.855*** (0.571) | -0.466* (0.262) |
| Observations | 1,150,343 | 482,500 | 667,843 |
| Individuals | 31,027 | 12,706 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from regressions of the number of correct answers on 3 °C bins of the average air temperature during the 48 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 11 Heat index and cognitive performance: piecewise-linear regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------------------|------------------|------------------|------------------|
| Heat index | 0.011 (0.020) | 0.012 (0.016) | 0.012 (0.015) |
| Linear combination test | -0.081*** | -0.117*** | -0.050* |
| Air temp. + air temp. × above threshold | (0.031) | (0.043) | (0.026) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from piecewise-linear regressions of the number of correct answers on the average heat index during the 24 h preceding the play (in °C) and an interaction with the above-threshold indicator as outlined in Eq. (1). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C, and the linear combination test amounts to the slope above 16.5 °C. Column 1 includes all observations. Columns 2 and 3 show results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 12 Heat index and cognitive performance: 3 °C-bins regressions

| Dependent variable: No. of correct answers | All ZIP Codes | Cold ZIP Codes | Hot ZIP Codes |
|-----------------------------------------------|----------------------|----------------------|---------------------|
| Heat index < 0 °C | -0.573 (0.506) | -0.641 (0.590) | -0.582* (0.330) |
| Heat index 0–3 °C | -0.516 (0.432) | -0.722 (0.543) | -0.066 (0.322) |
| Heat index 3–6 °C | -0.386 (0.267) | -0.384 (0.300) | -0.284 (0.450) |
| Heat index 6–9 °C | -0.570*** (0.211) | -0.624* (0.323) | -0.389 (0.243) |
| Heat index 9–12 °C | -0.141 (0.124) | 0.099 (0.193) | -0.253 (0.172) |
| Heat index 12–15 °C | -0.404* (0.215) | -0.308 (0.218) | -0.393 (0.318) |
| Heat index 18–21 °C | -0.122 (0.178) | -0.495 (0.335) | 0.085 (0.151) |
| Heat index 21–24 °C | -0.574** (0.257) | -0.913** (0.435) | -0.314* (0.166) |
| Heat index 24–27 °C | -0.617** (0.281) | -1.082*** (0.397) | -0.229 (0.184) |
| Heat index ≥ 27 °C | -0.914** (0.378) | -1.159** (0.544) | -0.520** (0.237) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from regressions of the number of correct answers on 3 °C bins of the average heat index during the 24 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 13 Air temperature and log cognitive performance: piecewise-linear regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------------------|---------------------|---------------------|--------------------|
| Air temperature | -0.003 (0.016) | 0.008 (0.016) | -0.010 (0.018) |
| Linear combination test | -0.079** (0.031) | -0.113** (0.049) | -0.057* (0.030) |
| Air temp. + air temp. × above threshold | | | |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from piecewise-linear regressions of $\log(\text{number of correct answers} + 1)$ on the average air temperature during the 24 h preceding the play (in °C) and an interaction with the above-threshold indicator as outlined in Eq. (1). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C, and the linear combination test amounts to the slope above 16.5 °C. Column 1 includes all observations. Columns 2 and 3 show results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 14 Air temperature and log cognitive performance: 3 °C-bins regressions

| Dependent variable: Log (no. of correct answers + 1) | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|---------------------------------------------------------|---------------------|---------------------|--------------------|
| Air temp. < 0 °C | -0.054 (0.389) | -0.285 (0.522) | -0.082 (0.427) |
| Air temp. 0–3 °C | -0.327 (0.356) | -0.596 (0.527) | -0.131 (0.378) |
| Air temp. 3–6 °C | 0.046 (0.301) | -0.224 (0.467) | 0.366 (0.343) |
| Air temp. 6–9 °C | -0.187 (0.227) | -0.381 (0.323) | 0.026 (0.353) |
| Air temp. 9–12 °C | -0.002 (0.181) | 0.297 (0.256) | -0.189 (0.232) |
| Air temp. 12–15 °C | -0.182 (0.182) | -0.054 (0.235) | -0.223 (0.253) |
| Air temp. 18–21 °C | -0.038 (0.185) | -0.526** (0.257) | 0.220 (0.228) |
| Air temp. 21–24 °C | -0.585** (0.247) | -0.780** (0.383) | -0.427* (0.248) |
| Air temp. 24–27 °C | -0.575** (0.255) | -0.898** (0.399) | -0.307 (0.234) |
| Air temp. ≥ 27 °C | -0.820** (0.349) | -1.332** (0.573) | -0.504 (0.306) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from regressions of $\log(\text{number of correct answers} + 1)$ on 3 °C bins of the average air temperature during the 24 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 15 Air temperature and error rate: piecewise-linear regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|-----------------------------------------------|------------------|-------------------|------------------|
| Air temperature | 0.000 (0.002) | -0.001 (0.002) | 0.002 (0.004) |
| Linear combination test | 0.001 | 0.004 | -0.002 |
| Air temp. + air temp. × above threshold | (0.004) | (0.006) | (0.004) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from piecewise-linear regressions of the error rate (number of incorrect answers / total answers) on the average air temperature during the 24 h preceding the play (in °C) and an interaction with the above-threshold indicator as outlined in Eq. (1). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C, and the linear combination test amounts to the slope above 16.5 °C. Column 1 includes all observations. Columns 2 and 3 show results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 16 Air temperature and error rate: 3 °C-bins regressions

| Dependent variable: Error rate | All ZIP Codes | Cold ZIP Codes | Hot ZIP Codes |
|--------------------------------|-------------------|-------------------|-------------------|
| Air temp. < 0 °C | -0.016 (0.046) | 0.004 (0.055) | -0.074 (0.078) |
| Air temp. 0–3 °C | 0.006 (0.044) | -0.019 (0.054) | 0.061 (0.067) |
| Air temp. 3–6 °C | -0.015 (0.040) | -0.013 (0.055) | -0.042 (0.053) |
| Air temp. 6–9 °C | 0.003 (0.036) | -0.000 (0.051) | -0.019 (0.052) |
| Air temp. 9–12 °C | -0.014 (0.029) | -0.065 (0.044) | 0.005 (0.038) |
| Air temp. 12–15 °C | 0.006 (0.028) | -0.018 (0.041) | 0.010 (0.037) |
| Air temp. 18–21 °C | -0.027 (0.028) | 0.014 (0.036) | -0.048 (0.037) |
| Air temp. 21–24 °C | -0.040 (0.033) | -0.015 (0.045) | -0.058 (0.042) |
| Air temp. 24–27 °C | -0.005 (0.038) | 0.060 (0.051) | -0.045 (0.047) |
| Air temp. > 27 °C | -0.034 (0.047) | -0.015 (0.092) | -0.068 (0.051) |
| Observations | 1,151,059 | 482,812 | 668,247 |
| Individuals | 31,029 | 12,708 | 18,321 |
| ZIPs | 748 | 374 | 374 |

Results from regressions of the error rate (number of incorrect answers / total answers) on 3 °C bins of the average air temperature during the 24 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 2 and 3 show the results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 17 Effect accumulation

| Dep. var.: No. of correct answers | All ZIP codes | | Cold ZIP codes | | Hot ZIP codes | |
|-----------------------------------------------|-----------------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|
| Av. air temp. past 24 h > 21 °C (baseline) | -0.4894** (0.2270) | | -0.6657** (0.2579) | | -0.2630** (0.1290) | |
| Av. air temp. > 21 °C for 1 or 2 days | | -0.2206 (0.1934) | | -0.4130* (0.2345) | | -0.0139 (0.1806) |
| Av. air temp. > 21 °C for 3 to 6 days | | -0.5682** (0.2288) | | -0.8578*** (0.2899) | | -0.2696** (0.1364) |
| Av. air temp. > 21 °C for 7 days or more | | -0.7886** (0.3446) | | -1.0750** (0.5032) | | -0.4790** (0.2086) |
| Observations | 1,142,089 | 1,142,089 | 478,616 | 478,616 | 663,473 | 663,473 |
| Individuals | 31,006 | 31,006 | 12,702 | 12,702 | 18,304 | 18,304 |
| ZIPs | 747 | 747 | 374 | 374 | 373 | 373 |

Results from regressions of the number of correct answers on and indicator = 1 if the average temperature was above 21 °C during different temporal periods before a play as presented in Fig. 3. The standard errors are clustered on ZIP Codes. The reference is a day with an average temperature below or exactly 21 °C. I run separate regressions for all observations and the cold- and hot-ZIP Codes. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 18 Air temperature and number of correct answers by age: piecewise-linear regressions

| Dependent variable: No. of correct answers | Below or 56 years old | Above 56 years old |
|-----------------------------------------------|--------------------------|--------------------|
| Air temperature | -0.012 (0.021) | 0.024 (0.030) |
| Linear combination test | -0.068 (0.049) | -0.091* (0.049) |
| Air temp. + air temp. × above threshold | | |
| Observations | 431,871 | 719,188 |
| Individuals | 15,758 | 15,271 |
| ZIPs | 722 | 728 |

Results from piecewise-linear regressions of the number of correct answers on the average air temperature during the 24 h preceding the play (in °C) and an interaction with the above-threshold indicator as outlined in Eq. (1). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C, and the linear combination test amounts to the slope above 16.5 °C. Columns 1 and 2 show the results from separate regressions for individuals below or 56 years old and individuals above 56 years old. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 19 Air temperature and number of correct answers by age: 3 °C-bins regressions

| Dependent variable: No. of correct answers | Below or 56 years old | Above 56 years old |
|-----------------------------------------------|-----------------------|--------------------|
| Air temp. < 0 °C | 0.155 (0.468) | -0.794 (0.715) |
| Air temp. 0–3 °C | -0.183 (0.472) | -0.724 (0.671) |
| Air temp. 3–6 °C | -0.032 (0.472) | -0.584 (0.501) |
| Air temp. 6–9 °C | -0.379 (0.380) | -0.396* (0.231) |
| Air temp. 9–12 °C | -0.107 (0.297) | -0.130 (0.159) |
| Air temp. 12–15 °C | -0.102 (0.251) | -0.366* (0.194) |
| Air temp. 18–21 °C | -0.162 (0.280) | 0.015 (0.281) |
| Air temp. 21–24 °C | -0.491 (0.351) | -0.472 (0.369) |
| Air temp. 24–27 °C | -0.627* (0.363) | -0.564 (0.412) |
| Air temp. ≥ 27 °C | -1.057** (0.513) | -0.890 (0.586) |
| Observations | 431,871 | 719,188 |
| Individuals | 15,758 | 15,271 |
| ZIPs | 722 | 728 |

Results from regressions of the number of correct answers on 3 °C bins of the average air temperature during the 24 h preceding the play as outlined in Eq. (2). The standard errors (in parentheses) are clustered on ZIP Codes. The reference bin is 15–18 °C. The regression in column 1 includes all observations. Columns 1 and 2 show the results from separate regressions for individuals below or 56 years old and individuals above 56 years old. The control variables include individual effects, time effects, play controls, and weather controls (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Table 20 Air temperature and number of correct answers by work hours: piecewise-linear regressions

| Dependent variable: No. of correct answers | All ZIP codes | Cold ZIP codes | Hot ZIP codes |
|---------------------------------------------------------------------|-------------------|-------------------|------------------|
| Air temperature | -0.009 (0.018) | 0.011 (0.021) | 0.013 (0.029) |
| Air temperature × work hours | -0.002 (0.015) | -0.024 (0.017) | 0.024 (0.028) |
| Linear combination test | -0.076** | -0.162** | -0.064 |
| Air temp. + air temp. × above threshold | (0.037) | (0.068) | (0.047) |
| Linear combination test: | -0.065* | -0.134** | -0.056 |
| Air temp. + air temp. × above threshold + air temp. × work hours | (0.037) | (0.068) | (0.048) |
| + air temp. × above threshold × work hours | | | |
| Observations | 705,629 | 298,399 | 407,230 |
| Individuals | 21,956 | 9,075 | 12,881 |
| ZIPs | 738 | 370 | 368 |

Notes: Results from piecewise-linear regressions of the number of correct answers on T_{jt} , A_{jt} , and their interaction, as outlined in Eq. (1), further interacted with an indicator for work hours (9 am to 5 pm on weekdays). The standard errors (in parentheses) are clustered on ZIP Codes. “Air temperature” denotes the coefficient below 16.5 °C during non-work hours, and “air temperature × work hours” refers to the coefficient below 16.5 °C during work hours. Similarly, the first linear combination test shows the effect above 16.5 °C during non-work hours, while the second linear combination test shows the effect above 16.5 °C during work hours. Column 1 includes all observations. Columns 2 and 3 show results from separate regressions for the cold- and hot-ZIP Codes samples. The control variables include individual effects, time effects, play controls, weather controls, and an indicator for above threshold temperature (see Sect. 4). The significance levels are: * $p \leq .1$; ** $p \leq .05$; *** $p \leq .01$

Appendix. Additional results

A further interesting question is whether temperature has a higher effect at the workplace vs. in people’s homes.¹² Unfortunately, since the data only includes information on the three-digit ZIP code fixed at the individual level, I cannot discern between users’ home and work locations. However, I run the main analysis for people under 65 and interact T_{jt} , A_{jt} , and the interaction of these two variables with an indicator equal to 1 for plays during the hours of 9 am to 5 pm on weekdays. Table 20 presents the results. I find that heat affects people in colder ZIP codes somewhat less during work hours (-0.162 vs. -0.134).

Of course, these results have several limitations. First, as mentioned, it is unclear where people actually are. Second, I have no information about the HVAC system, neither at people’s homes nor at their work locations. Most likely, for some, climate control is better at work, and for others, it is worse at work, especially for people working outdoors. Third, even if we had information about people’s indoor temperature at home and at work, it is unclear to what extent a potentially differential effect at work is attributable to exposure at work vs. exposure during the commute.

¹² I thank an anonymous reviewer for this suggestion.

Funding 'Open Access funding provided by the MIT Libraries'

Declarations

Conflict of interest The author declares no Conflict of interest regarding the publication of this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Auffhammer M (2018) Quantifying economic damages from climate change. *J Econ Perspect* 32(4):33–52
- Bao X, Fan Q (2020) The impact of temperature on gaming productivity: evidence from online games. *Empir Econ* 58(2):835–867
- Barreca AI (2012) Climate change, humidity, and mortality in the United States. *J Environ Econ Manag* 63(1):19–34
- Burke M, Hsiang SM, Miguel E (2015) Global non-linear effect of temperature on economic production. *Nature* 527(7577):235–239
- Carleton T, Jina A, Delgado M, Greenstone M, Houser T, Hsiang S, Hultgren A, Kopp RE, McCusker KE, Nath I et al (2022) Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *Quart J Econ* 137(4):2037–2105
- Carleton TA, Hsiang SM (2016) Social and economic impacts of climate. *Science* 353(6304)
- Cho H (2017) The effects of summer heat on academic achievement: a cohort analysis. *J Environ Econ Manag* 83:185–196
- Cook N, Heyes A (2020) Brain freeze: outdoor cold and indoor cognitive performance. *J Environ Econ Manag* 101:102318
- Dell M, Jones BF, Olken BA (2012) Temperature shocks and economic growth: evidence from the last half century. *Am Econ J Macroecon* 4(3):66–95
- Dell M, Jones BF, Olken BA (2014) What do we learn from the weather? The new climate-economy literature. *J Econ Lit* 52(3):740–98
- Deschênes O, Greenstone M (2011) Climate change, mortality, and adaptation: evidence from annual fluctuations in weather in the US. *Am Econ J Appl Econ* 3(4):152–85
- Garg T, Jagnani M, Taraz V (2020) Temperature and human capital in India. *J Assoc Environ Resour Econ* 7(6):1113–1150
- Graff Zivin J, Solomon M H, Neidell M (2018) Temperature and human capital in the short and long run. *J Assoc Environ Resour Econ* 5(1):77–105
- Graff Zivin J, Tong L, Yingquan S, Tang Q, Peng Z (2020) The unintended impacts of agricultural fires: human capital in China. *J Dev Econ* 147:102560
- Graff Zivin J, Matthew N (2013) Environment, health, and human capital. *J Econ Lit* 51(3):689–730
- Graff Zivin J, Matthew N (2014) Temperature and the allocation of time: implications for climate change. *J Law Econ* 32(1):1–26
- Graff Zivin J, Yingquan Song Q, Tang PZ (2020) Temperature and high-stakes cognitive performance: evidence from the national college entrance examination in China. *J Environ Econ Manag* 104:102365
- Hancock PA, Vasmatazidis I (2003) Effects of heat stress on cognitive performance: the current state of knowledge. *Int J Hyperth* 19(3):355–372
- Hanushek EA, Schwerdt G, Wiederhold S, Woessmann L (2015) Returns to skills around the world: evidence from PIAAC. *Eur Econ Rev* 73:103–130
- Heutel G, Miller NH, Molitor D (2021) Adaptation and the mortality effects of temperature across us climate regions. *Rev Econ Stat* 103(4):740–753
- Hsiang SM, Burke M, Miguel E (2013) Quantifying the influence of climate on human conflict. *Science* 341(6151):1235367

- Krebs B, Simon L (2024) Air pollution, cognitive performance, and the role of task proficiency. *J Assoc Environ Resour Econ*, forthcoming
- Steffen K, Juan P, Nico P (2023) Indoor air quality and strategic decision making. *Manag Sci* 69(9):5354–5377
- Lobell DB, Schlenker W, Costa-Roberts J (2011) Climate trends and global crop production since 1980. *Science* 333(6042):616–620
- Melo AP, Mizuhiro S (2023) Temperature, effort, and achievement. *Unpublished manuscript*
- Minor K, Bjerre-Nielsen A, Jonasdottir SS, Lehmann S, Obradovich N (2022) Rising temperatures erode human sleep globally. *One Earth* 5(5):534–549
- Missirian A, Schlenker W (2017) Asylum applications respond to temperature fluctuations. *Science* 358(6370):1610–1614
- Mullins JT, White C (2020) Can access to health care mitigate the effects of temperature on mortality? *J Public Econ* 191:104259
- Park RJ (2022) Hot temperature and high-stakes performance. *J Human Resour* 57(2):400–434
- Park RJ, Behrer AP, Goodman J (2021) Learning is inhibited by heat exposure, both internationally and within the United States. *Nat Hum Behav* 5(1):19–27
- Park RJ, Goodman J, Hurwitz M, Smith J (2020) Heat and learning. *Am Econ J Econ Pol* 12(2):306–339
- Roach T, Withney J (2021) Heat and learning in elementary and middle school. *Educ Econ*, 1–18
- Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proc Natl Acad Sci* 106(37):15594–15598
- Somanathan E, Somanathan R, Sudarshan A, Tewari M (2021) The impact of temperature on productivity and labor supply: evidence from Indian manufacturing. *J Polit Econ* 129(6):1797–1827
- Wang C, Zhang F, Wang J, Doyle JK, Hancock PA, Mak CM, Liu S (2021) How indoor environmental quality affects occupants' cognitive functions: a systematic review. *Build Environ* 193:107647
- Yi F, Zhou T, Luyun Yu, McCarl B, Wang Y, Jiang F, Wang Y (2021) Outdoor heat stress and cognition: effects on those over 40 years old in China. *Weather Clim Extremes* 32:100308

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.