Pre-Analysis Plan Addendum: Temperature and Cooperation

Jesse Anttila-Hughes Alessandra Cassar Andrew Hobbs
Travis Lybbert Irvin Rojas Bruce Wydick

January 28, 2023

PAP Addendum (1/28/23)

We create this addendum after having done a preliminary analysis of results on the effect of temperature in Mexico and India, and of gender in India, but before data has been compiled for analysis from California and Kenya and before experimental work has been carried out in Colombia, and (potentially) Italy. Changes can be seen in the modified PAP that follows this addendum.

- First, we would like to clarify the ambiguity in our initial pre-analysis plan that we will analyze outcomes related to specific dictator game decisions as well as indices created from the dictator games related to altruism, egalitarianism, selfishness, spite, and competitiveness. To be conservative in our approach with respect to Type 1 errors, this will not only be tested on the sample as a whole, but verified on the latter three countries independently.
- Second, we correct an error in our initial pre-analysis plan related to the moderator section 3.3.1 of our initial submission. Specifically, we make the correction in our initial PAP filing which initially showed the interaction as $(moderator \times first_g \times hwbtemp_g)$, when according to our specification in 3.2 it should have read $(moderator \times first_g \times lostcomp_i \times hwbtemp_g)$.
- Third, moving forward to the second phase of our data analysis and fieldwork, we would like to upgrade to primary hypotheses the independent effects of the moderators in section 3.2 themselves–gender, income, life satisfaction, social ties, and social trust–on altruism, egalitarianism, selfishness, spite, and competitiveness. Here our specific hypotheses are that female gender, lower income, higher life satisfaction, higher social ties, and higher social trust will each exhibit a positive relationship with altruism and egalitarianism, and a negative relationship with selfishness, spite, and competitiveness. We will test that outcomes from data analyzed in our first two countries extends in terms of external validity to our remaining countries.

Contents

1 Introduction

A large body of research has begun to document substantial negative impacts of hot temperatures on economic outcomes. This pre-analysis plan is modeled after Almås et al. (2017) due to its substantial similarity to Almås et al. (2019).

2 Research Strategy

2.1 Sampling

2.1.1 Sampling Frame

We will conduct this study preliminarily at four sites. First, in Archer's Post, Kenya the target population will be young people with at least a high school degree. In Davis, California the target population will be students at the University of California, Davis. In Mexico City the target population will be undergraduate students at the University of Chapingo, and in Delhi, India, the target population will also be students of the university. Likely sites in the future will be at a campus in the southern United States and/or a university site in South America.

2.1.2 Statistical Power

We estimate the minimum detectable standardized effect size (MDES) as follows:

$$MDES = (t_\alpha + t_{1-\beta}) \times \sqrt{\frac{1}{P(1-P)n}},$$

where α is the significance level, $1-\beta$ is the power; P and $1-P$ are the shares of the sample assigned to the treatment group and the control group, respectively; and n is the sample size. With a sample size on average of 350 per site and with five sites we would have a sample size for the project of approximately 1,750. In our discrete-treatment estimation, we would have temperatures 30-34°C as treatment and temperatures 18-24°C as control. By the proposed design in these estimations, we would eliminate the approximately one-quarter of our sample from 24-29°C, giving us a sample size here of 1,312 with 2/3 in treatment and 1/3 in control. At 80% power, this would give us

$$MDES = (1.96 + 2.84)(0.0587) = 0.166\sigma,$$

meaning that if the true effect on different behaviors were 0.166σ, we would be able to reject a null of no impact of the warmer temperature 80% of the time. It is difficult to gauge the extent to which intraclass correlation of behavioral outcomes would be present within sites. (This also may be less important with the use of site fixed effects when we carry out estimations on the whole sample.) With 80% power and alpha equal to 0.05 ($\rho = 0$) and an $MDES = 0.20$, the required sample size is 317. With $MDES = 0.10\sigma$, required sample size rises

to 1,266, and with $MDES = 0.30\sigma$ it falls to 141, putting us within the range of our target samples for the project. At the specific site level, a sample in the range of 300-400 per site would thus leave us within an MDES of target of 0.20σ.

2.1.3 Assignment to Treatment

Sessions will be randomly assigned to treatment/ control using temperature modulations that will vary daily in a manner that is as-if random with respect to observable and unobservable characteristics of subjects who sign up for sessions at different times. In our dry-climate sites we may increase humidity levels such that they are more consistent with our more humid site areas for the purpose of creating similar wet bulb temperatures.

2.1.4 Attrition from the Sample

We do not expect large numbers of participants to leave the experiment once they have entered. People will, however, be allowed to leave the experiment at any time. In order to minimize this, we will encourage participants to use the bathroom and drink water during the experiment as needed, and snacks will be provided.

2.2 Fieldwork

2.2.1 Instruments

In all sites we will use:

- Space heaters to heat temperatures. Humidifiers in dry sites to adjust humidity in the room closer to a level of humidity at other sites.
- Air conditioners to reduce room temperatures.
- Thermometers and humidity monitors to record ambient temperatures and humidity levels.

2.2.2 Data Collection

Prior to the experiment we will run a series of pilots which we may use to adjust the experimental procedures, which are described in section 4. Data from the pilots will not be used in the analysis.

We expect the data collection process from the first four sites to take several months. Data from the experiment will be recorded by in spreadsheets by USF Masters students and/ or enumerators without identifying information and hosted in a Dropbox folder shared with the research team.

2.2.3 Data Processing

Data processing includes cleaning, managing, and analyzing the data. We expect to do the majority of the data processing once all data is recorded, and it will be done by the study team and USF Masters students.

3 Empirical Strategy

3.1 Main Hypotheses

Primary (Alternative) Hypotheses:

- Generosity and egalitarianism, decrease while selfishness, spite, and competitiveness increase with higher wet-bulb room temperatures. The first five of these will be measured by outcomes from the dictator games and competitiveness will be measured by subjects' choice of piece-rate vs. competition in their third puzzle game. (Null hypothesis: no difference.) A pro-social behavior Kling et al. (2007) index will be created weighted positively by generosity and egalitarianism, and negatively by selfishness, spite, and competitiveness and tested as a summary of these behaviors.
- Generosity, egalitarianism, and the index of pro-social behavior will decrease while selfishness, spite, and competitiveness increase with the interaction of losing the puzzle competition with higher wet-bulb room temperatures (when puzzle competition precedes the dictator games).
- Cognitive capacity, as measured by performance in the piece-rate puzzle task declines with increased wet-bulb room temperature.

Primary hypotheses are not subject to correction of the family-wise error rate; tests of rejections of nulls will be established by two-tailed p-values.

3.2 Empirical Specification

The main specification for our analysis can be written

$$\begin{aligned} y_{igc} =& \alpha + \tau_1 wbtemp_g + \gamma_1 first_g + \gamma_2 lostcomp_i + \gamma_3(first_g \times lostcomp_i) + \\ & \gamma_4(first_g \times wbtemp_g) + \tau_2(first_g \times lostcomp_i \times wbtemp_g) + \theta_c + X_i'\beta + \epsilon_{igc} \end{aligned}$$

where y_{igc} represents one of the five behavioral characteristics from the four dictator games (altruism, egalitarianism, selfishness, and spitefulness) of individual i in experimental group g, and country c, $wbtemp_g$ is the wet bulb temperature measured at the start of the experiment for experimental group g, first is whether a group was randomly assigned to perform the competition exercise first (before the dictator games), $lostcomp_i$ is whether subject i was in the losing half of the competition, θ_c is a country-level fixed effect, and X_i is a k-dimensional vector of controls for personal characteristics. The coefficients τ_1 and τ_2 represent tests of our hypotheses, the former testing whether heat affects behavior, and

the latter testing whether heat interacted with the trigger affects behavior. In some specifications we will include a control for outdoor temperature during the session. In addition we will run two estimations in which we substitute a dummy variable indicating the warmest session temperatures. A second will substitute a wetbulb temperature above 25C ($hwbtemp_g$)for $wbtemp_g$ and where we will omit temperatures from 20-24.99C from the data. A second will substitute a dummy variable for (standard) temperature above 30C for $wbtemp_g$ and where we will omit temperatures from 24-29C from the data, where we will include a control for humidity.

3.3 Heterogeneity

We will examine treatment effect heterogeneity over two groups: moderators and mediators. This section details the hypotheses to be tested, methods, and variables of interest.

3.3.1 Moderators

- The effects of temperature and the interaction of temperature with losing the puzzle competition are lower among female subjects than male subjects. This test will be performed with the interaction of $gender \times hwbtemp_g$ and the interaction of $(gender \times first_g \times lostcomp_i \times hwbtemp_g)$.
- The effects of temperature and the interaction of temperature with losing the puzzle competition is higher among those with lower income as measured by occupation. This test will be performed with the interaction of $occupation \times hwbtempg$ and the interaction of $(occupation \times first_g \times lostcomp_i \times hwbtemp_g)$.
- The effects of temperature and the interaction of temperature with losing the puzzle competition is higher when outside wet-bulb temperature is higher. This test will be performed with the interaction of $OWBT \times hwbtempg$ and the interaction of $(OWBT \times first_g \times lostcomp_i \times hwbtemp_g)$.
- The effects of temperature and the interaction of temperature with losing the puzzle competition is higher when social trust is lower. This test will be performed with the interaction of $social_trust \times hwbtempg$ and the interaction of $(social_trust \times first_g \times lostcomp_i \times hwbtemp_g)$.
- The effects of temperature and the interaction of temperature with losing the puzzle competition is lower when acquaintance with other subjects in the same session is higher. This test will be performed with the interaction of $know_others \times hwbtempg$ and the interaction of $(know_others \times first_g \times lostcomp_i \times hwbtemp_g)$.
- The effects of temperature and the interaction of temperature with losing the puzzle competition is lower when general dissatisfaction with one's

life is higher (as measured by a Kling et al. (2007) index of our life satisfaction questions. This test will be performed with the interaction of $dissatisfaction \times hwbtempg$ and the interaction of $(dissatisfaction \times first_g \times lostcomp_i \times hwbtemp_g)$.

3.3.2 Mediators

The effect of heat on changes in the pro-social behavior Kling index is mediated by

- Lower cognitive ability (as measured by piece-rate Raven's performance)
- Lower levels of happiness (as measured by our survey questionnaire)
- Higher levels of discouragement (as measured by our survey questionnaire)
- Lower levels of energy (as measured by our survey questionnaire)
- Greater levels of anxiety (index created from questionnaire module)

Mediators will be tested through the following procedure: In a regression of the pro-social behavior index on high wet-bulb temperature, the coefficient on high wet-bulb temperature is positive and statistically significant. In a regression of the mediator on high wet-bulb temperature, the coefficient on high wet-bulb temperature is positive and statistically significant. We will use moderators to identify mediators:

1. Perform a first LASSO regression with each mediator on (high wet-bulb) temperature and temperature interacted with each moderator and two-way interactions of all moderators, penalizing only the interaction terms.
2. Perform a second LASSO regression of the pro-social behavior index on temperature and temperature interacted with each moderator and two-way interaction of moderators, penalizing only the interaction terms.
3. Carry out regressions of the pro-social behavior index on both wet-bulb room temperature and an individual mediator, using all of the interaction terms retained by LASSO in that mediator's regression in (1) but not retained in (2) to instrument for the mediator. The coefficient on temperature should be lower than in (2).
4. Carry out a regression of the pro-social behavior index on both temperature and all mediators in question, using all of the interaction terms retained in all of the mediator regressions in (1) but not retained in (2) to instrument for the mediators jointly. Check that the rank condition for instrumental variables is satisfied. The coefficient on high temperature should be lower than in (2) and the coefficient on the particular mediator should be statistically significant if it is to satisfy the criteria for mediation.

3.3.3 Nonlinear effects

We hypothesize that the relationship between temperature and behavior is nonlinear and plan to explore nonlinearities/ breakpoints in the relationship between the two both parametrically and nonparametrically with the goal of determining if there is a wet bulb temperature beyond which behavior begins to change.

3.3.4 Controlling for Multiple Inference

For our primary analysis, we plan to follow Benjamini et al. (2006) to control the false discovery rate (FDR). We will first apply this on all hypothesis being tested in our main specification across all outcomes, and will follow Anderson (2008) in reporting the minimum q-value for which each hypothesis is rejected. We will report both minimum q-values and p-values.

The second family to include adjustments is the heterogeneity analyses described above. We will calculate q-values following Benjamini et al. (2006) for each of the moderators coming from each outcome (in other words, each outcome will constitute a family).

3.4 Balance Checks

We will use Students t-tests to check for balance between treatment and control groups. The variables that will be checked for balance across treatment arms include

4 Experimental Procedure

4.1 Recruitment

Subjects will be recruited via email and in-person flyers. All subjects will be aged 18-39. Teams may use an online sign-up sheet to schedule subjects for sessions. At no time will recruiters mention that the study is related to climate change or mention the temperature treatment in their recruiting of subjects. Subjects should be asked "Would you like to participate in an experiment in which you can earn up to $X (which will vary by country site) that involves puzzle-solving exercises and how people make strategic economic decisions?"

4.2 Assignment to Treatment

The experimental room will be set to varying temperatures potentially ranging from quite cool (18°C, 64°F) to very warm (34°C, 93°F). For statistical power considerations and to also give us the ability to identify at which temperature points behaviors change, approximately half of the room temperatures in each country site will be set (uniformly) between 18°C and 29°C and half will be set (uniformly) between 30°C and 34°C across all of the sessions. Temperatures will

revolve around the following repeating four-day cycle: Day 1, room temperatures will start in the morning at 30°C and be gradually heated to 34°C by the last experimental session of the day. Day 2, room temperatures will start in the morning at 34°C and be reduced to 30°C by the last experimental session of the day. Day 3, room temperatures will start in the morning at 18°C and be gradually heated to 29°C by the last experimental session of the day. Day 4, room temperatures will start in the morning at 29°C and be reduced to 18°C by the last experimental session of the day. (Repeat four-day cycle on Days 5, 9.) Outside air temperature and humidity, and experimental room air temperature and humidity will be recorded at the start of each experiment.

4.3 Experimental Sessions

Subjects will be welcomed into the experimental room and students and field coordinators will explain the experiment to subjects. Experimenters will *not* explain the purpose or hypotheses involved in the experiment—only that the experiment is intended to add to our understanding of puzzle-solving skills and strategic economic behavior. Subjects will be assured that they can quit the experiment or leave the room at any time if they feel uncomfortable.

Two-thirds of the time, the experiment will begin with two Raven's matrix exercises consisting of 10 puzzles each, each lasting two minutes. A piece-rate exercise, in which subjects will be paid $Y for every correctly solved puzzle over the two minutes. A competition exercise in which subjects will be randomly matched with another subject, which subjects will not know which other subject in the session is their competition partner. After solving as many puzzles as possible at the end of two minutes, each subject will then toss a six-sided die, where the die roll is private information between the subject and experimenter. The number appearing face up on the die will then be added to the subject's Raven's test score to create a total score for each subject. In each competition, pair winners will be paid a bonus of $Z while losers receive no bonus. Winners and losers will be private information between the experimenter and the subject. If there is an odd number of subjects, declare the last "odd" subject a winner if and only if their score is higher than the player to their right. Immediately after the competition exercise, subjects will be given three questions:

- How happy would you say that you feel right now? (0 = unhappy to 10 very happy)
- How frustrated would you say you feel right now? (0 = not frustrated to 10 very frustrated)
- How much energy would you say that you feel you have right now? (0 = no energy to 10 highest energy)

In two-thirds of cases, the subjects will then make their four dictator game choices, allocating an amount to "self" and an amount to "other" for each of the four choices. Note: One-third of the time (every third session) the order of

the competition exercises and that of the dictator games, steps 4 and 5, will be reversed.

Subjects will be given two final tasks:

- A quick risk-aversion task in which they make a single optimal choice across greater levels of risk with higher expected payoffs.
- A third play of the Raven's matrix puzzles, where subjects choose whether they would prefer to be paid by piece-rate, the number of correct puzzles solved in 2 minutes to enter into a competition with another random session member in which only the winner will be paid a bonus.

Subjects will lastly fill out the survey that that will include name, gender, marital status, number of children, number of siblings, religion/denomination, employment status, type of employment, approximate income level, leadership positions in community, number of years of formal education, a few basic nutrition questions related to breadth of diet and meal sufficiency, and other questions. The section on general psychological well-being will then follow, questions related to general happiness, alertness, energy level, social trust and potentially other measures of mood.

References

Almås, I., Auffhammer, M., Bold, T., Bolliger, I., Dembo, A., Hsiang, S., Kitamura, S., Miguel, E., and Pickmans, R. (2017). Pre-analysis plan: Temperature and economic choices.

Almås, I., Auffhammer, M., Bold, T., Bolliger, I., Dembo, A., Hsiang, S. M., Kitamura, S., Miguel, E., and Pickmans, R. (2019). Destructive behavior, judgment, and economic decision-making under thermal stress. Technical report, National Bureau of Economic Research.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.