

NBER WORKING PAPER SERIES

DESTRUCTIVE BEHAVIOR, JUDGMENT, AND ECONOMIC DECISION-MAKING  
UNDER THERMAL STRESS

Ingvild Almås  
Maximilian Auffhammer  
Tessa Bold  
Ian Bolliger  
Aluma Dembo  
Solomon M. Hsiang  
Shuhei Kitamura  
Edward Miguel  
Robert Pickmans

Working Paper 25785  
<http://www.nber.org/papers/w25785>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue

Cambridge, MA 02134

April 2019, Revised December 2020

We thank Botond Kozegi, Bertil Tungodden, and seminar participants at the Santa Fe Institute, Stockholm University, University of California, Berkeley, and the East African Social Science Training Annual Summit for useful comments. This project was supported by the Peder Sather Center for Advanced Study and the Swedish Research Council. The authors declare no competing financial interests. A randomized controlled trials registry entry is available at <https://www.socialscienceregistry.org/trials/1361>. This study has complied with all relevant ethical regulations. The University of California, Berkeley Committee for the Protection of Human Subjects and the Kenya Medical Research Institute Scientific and Ethics Review Unit approved the study protocols. Informed consent was obtained from all participants. All authors contributed equally to this work, and are listed in alphabetical order. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w25785.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2019 by Ingvild Almås, Maximilian Auffhammer, Tessa Bold, Ian Bolliger, Aluma Dembo, Solomon M. Hsiang, Shuhei Kitamura, Edward Miguel, and Robert Pickmans. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Destructive Behavior, Judgment, and Economic Decision-making under Thermal Stress  
Ingvild Almås, Maximilian Auffhammer, Tessa Bold, Ian Bolliger, Aluma Dembo, Solomon M. Hsiang,  
Shuhei Kitamura, Edward Miguel, and Robert Pickmans

NBER Working Paper No. 25785

April 2019, Revised December 2020

JEL No. C91,D74,O1,Q50

## **ABSTRACT**

Accumulating evidence indicates that environmental temperature substantially affects economic outcomes and violence, but the reasons for this linkage are only partially understood. While factors external to human beings (such as agricultural production) are known to respond adversely to high temperatures, extreme temperatures could also directly influence the internal mental processes governing decision-making. We study this by systematically evaluating the effect of thermal stress on multiple dimensions of economic decision-making, judgment, and destructive behavior with 2,000 participants in Kenya and the US who were randomly assigned to different temperatures in a laboratory. We find that heat significantly affects individuals' willingness to voluntarily destroy other participants' assets, with pronounced increases among those experiencing heightened political conflict in Kenya. We find that other major dimensions of economic decision making are largely unaffected by temperature.

Ingvild Almås

Institute for International Economic Studies

Stockholm University

SE-106 91 Stockholm

Sweden

Ingvild.Almas@nhh.no

Maximilian Auffhammer

Agricultural and Resource

Economics Department

University of California, Berkeley

207 Giannini Hall

Berkeley, CA 94720-3310

and NBER

auffhammer@berkeley.edu

Tessa Bold

Stockholm, Sweden

IIES

tessa.bold@iies.su.se

Ian Bolliger

Energy and Resources Group

University California Berkeley

Berkeley, CA 94720

bolliger@berkeley.edu

Aluma Dembo

Department of Economics

Manor Road Building

Oxford Nuffield College

New Road

Oxford, OX1 1NF

United Kingdom

aluma.dembo@economics.ox.ac.uk

Solomon M. Hsiang

Goldman School of Public Policy University

of California, Berkeley

2607 Hearst Avenue

Berkeley, CA 94720-7320

and NBER

shsiang@berkeley.edu

Shuhei Kitamura

Osaka School of International Public Policy

Osaka University

1-31 Machikaneyama, Toyonaka

Osaka, 560-0043

Japan

kitamura@osipp.osaka-u.ac.jp

Edward Miguel

Department of Economics

University of California, Berkeley

530 Evans Hall #3880

Berkeley, CA 94720

and NBER

emiguel@econ.berkeley.edu

Robert Pickmans

Agricultural and Resource

Economics Department

University of California, Berkeley

207 Giannini Hall

Berkeley, CA 94720-3310

rpickmans@berkeley.edu

# 1 Introduction

Scholars have long attempted to understand the role that climate plays in human societies (de Secondat, 1748; Huntington, 1922; Lee, 1962). We have seen a recent acceleration of findings enabled by improved data, computing, and statistical techniques (Carleton and Hsiang, 2016). This research has established an empirical link between high temperature and heightened risk of violence and conflict (Hsiang et al., 2013) as well as a link between extreme temperatures and worse economic and societal outcomes (Burke et al., 2015; Jones and Olken, 2010; Graff Zivin and Neidell, 2014; Obradovich et al., 2018).<sup>1</sup>

The most pressing and challenging task remains to identify the causal mechanisms underlying this array of reduced-form findings (Dell et al., 2014). Two hypotheses are typically proposed, at least informally. First, it is possible that factors *external* to human beings respond to high temperature, which then directly or indirectly affect outcomes. For example, climate-induced loss of agricultural income has been proposed as a mechanism linking extreme climatic events with decisions to engage in political conflict (Schlenker and Roberts, 2009; Miguel et al., 2004), but also to suicide (Carleton, 2017), transactional sex (Burke et al., 2015), and migration (Missirian and Schlenker, 2017; Feng et al., 2010).

Alternatively (and not mutually exclusively), extreme temperatures could directly influence the *internal mental processes* governing decision-making. In this vein, the medical and psychological literature has posited that heat increases aggression due to its negative impact on emotional affect, because lowered affect is misattributed to anger, or because of a physiological effect of heat that triggers aggression (Bell and Baron, 1976; Zillmann, 2006), though experimental evidence of a heat-aggression relationship is mixed (Anderson et al., 2000). In the economics literature, it has been suggested that violent crime and group conflicts become more likely at high temperatures at least in part because individual decision-making is altered – for instance, through diminished

---

<sup>1</sup>In addition, high temperatures are found to reduce physical activity (Obradovich and Fowler, 2017), and increase the incidence of suicide (Baysan et al., 2019), HIV transmission (Burke et al., 2015) and interpersonal violence. High temperatures are also found to reduce fertility (Barreca et al., 2018), while increasing domestic migration (Bohra-Mishra et al., 2014), and energy use (Wenz et al., 2017).

consideration for future consequences, lowered risk-aversion, or heightened risk of strategic miscalculation (Anderson, 1989; Baysan et al., 2019). Beyond scholarly interest in these issues, determining the relative influence of these various external factors and internal mental processes is crucial to developing policies or technologies to address the damages caused by high temperatures. One explanation would suggest directing resources towards adaptation of external factors such as infrastructure and agricultural biotechnology, while the other might favor investments in cooling indoor environments at home and work.

This study conducts the first systematic laboratory evaluation to our knowledge of high temperature on the internal mental processes that govern economic decision-making, judgment and destructive behavior. We randomly assign subjects to Hot (30°C target) or Control indoor environments (22°C target) in two parallel and simultaneous experiments conducted in Berkeley, California and in Nairobi, Kenya, from September 2017 to February 2018. We deploy a battery of assessments designed to measure components of individual cognition, private preferences and decision-making quality, as well as social preferences and destructive behavior. For the cognition domain, we measure the effects of temperature manipulation on *precision*, *fluid intelligence*, and *cognitive reflection* (Gill and Prowse, 2012; Penrose and Raven, 1936; Frederick, 2005). In the domain of private preferences, we measure *risk taking* and *patience* (Eckel and Grossman, 2008; Andreoni et al., 2015). Decision-making quality is assessed by checking the *transitivity of risk preferences* and *consistency of time preferences* (Eckel and Grossman, 2008; Andreoni et al., 2015). We measure social preferences through behavior in *fairness (altruism)*, *trust*, *public goods provision*, and *charity* games (Almås et al., 2010; Johnson and Mislin, 2011; Fischbacher et al., 2001; Andreoni, 2006) and destructive behavior in a *joy of destruction* game (Abbink and Sadrieh, 2009). Finally, we record individual affect and arousal, namely, *happiness* and *alertness* (Russell and Barrett, 1999).

We find that nearly all dimensions of judgment and decision-making are largely unaffected by high temperatures. The precisely estimated null results provide the strongest available evidence against previous theories that changes in decision-making are responsible for many population-

scale responses to a warming environment. This finding is encouraging as it suggests that the behavioral framework underlying mainstream economics largely withstands changes in the environmental characteristic that is arguably most central to human existence, namely, temperature.

A second set of results provides novel evidence for a decision-making channel linking high temperatures to destructive or violent outcomes. The data also indicate that this effect is context-specific. Specifically, we estimate that exposure to high temperatures increases destructive behavior among Kenyan participants by +0.30 of a standard deviation on average, providing support for prior theories that cognitive responses to thermal stress could contribute to the systematic rise in violence observed due to warming. Interestingly, this effect is driven by the participants who self-identify as members of ethnic groups that had been politically marginalized in a highly contested national election in Kenya, a major event for which timing only coincidentally aligned with our experiment. The finding is striking as one of the most hotly debated topics in the larger climate-society literature is the apparent linkage between political conflicts in Africa and climatic extremes (Hsiang et al., 2013), with some scholars (von Uexkull et al., 2016) arguing that ethnic divisions are at least partly responsible. Among California participants, on the other hand, where the associations with heat may be more positive, if anything we document a moderate reduction in destruction in the treatment condition. As such, our study provides new evidence that thermal stress and the political context together might be jointly responsible for generating observed patterns of violence through a mechanism embedded in the human mind.

The analysis reported in this paper aims to attain high standards for transparency in study design, implementation, and analysis. The replicability of social science research has recently been questioned, with a number of studies showing that publication bias and specification searching may produce misleading bodies of research, including in economics lab experiments (Camerer et al., 2018; Andrews and Kasy, 2019; Christensen and Miguel, 2018). This has motivated proposals for better and more transparent research practices, such as pre-registration of experiments and analysis and adjusting p-values for multiple hypothesis testing (Miguel et al., 2014; Christensen et al., 2019). This study's data analysis was pre-registered to limit researcher bias. The finding of

a direct link from temperature to destruction in Kenya was part of this pre-registered analysis, is significant even relative to the more stringent level of p-value < 0.005 proposed (Benjamin et al., 2018), and survives adjustment for multiple hypothesis testing. For transparency, we also note whenever the analysis goes beyond the originally specified plan, which is the case, for instance, for the ethnic subgroup analysis in Kenya mentioned above. We also view our complete reporting of the null findings obtained on the wide range of pre-specified outcomes as important for avoiding the publication bias that has characterized some bodies of research evidence.

Additionally, we note that the study was designed in advance to be highly statistically powered (N=2,000) in case the analysis recovered null results, a decision that turned out to be crucial in allowing us to reject substantive effect sizes for many economic outcomes. Specifically, the study had 80% power to detect non-trivial effects (0.20 or more of a standard deviation) for the whole battery of laboratory experiments, and at least 80% power to detect small to medium effect sizes (0.30 or more of a standard deviation) even with the most conservative multiple testing adjustment.

In terms of the broader literature, some previous work has suggested a link between temperature and internal mental processes. First, scholars have established a link between high temperature and cognitive performance, specifically, deciphering Morse code (Mackworth, 1946) and solving math problems (Graff Zivin et al., 2018). Second, there is evidence that “aggressive” behaviors – such as sports violence (Lerrick et al., 2011), horn-honking (Kenrick and MacFarlane, 1986) and swearing on social media (Baylis et al., 2018) – increase with temperature. In fully controlled laboratory environments, the evidence is decidedly more mixed, as discussed in Anderson et al. (2000). We hope to contribute to this body of work through (i) experiments conducted in a tightly controlled laboratory environment (most of the studies noted above are non-experimental), (ii) a wide range of standard economic outcome measures, (iii) large samples at two sites, which allow us to detect modest effect sizes and establish precise null results where appropriate, and (iv) a pre-registered econometric analysis.

In general, there are useful complementarities between studying behavior in the field and in a more controlled laboratory setting. Field experiments have the advantage of examining naturally

occurring settings, but often struggle to pin down causal mechanisms. A leading strength of an experimental laboratory setting like ours, relative to most field experiments or observational data, is the ability to “shut down” other factors that might confound the relationship between higher temperatures and the outcomes we consider, in order to better establish underlying channels. We believe the results of this study add valuable evidence to an ongoing debate, especially in light of the dearth of empirical evidence on the mechanisms underlying the well-established reduced-form link between extreme temperatures and societal outcomes.

It is also important to note the relevance of our findings to the understanding of how climate change affects economic outcomes. Climate change will drastically increase the frequency of hot and very hot days (Field et al., 2012). The most direct avenue humans will use to adapt to hotter ambient temperatures is to cool down the indoor environment to a more comfortable temperature, which has been shown to lead to significant improvements in health outcomes – though at a cost. Increased operation of existing cooling equipment and possible purchasing of new air conditioners will lead to increased fixed and variable expenditures for households and firms (Davis and Gertler, 2015). Our findings are most directly relevant to decision-making in a counterfactual cooled versus a non-cooled room setting, as the laboratory experiments were conducted indoors. This would include many settings where humans make economically consequential decisions, including the home, offices, shops, and factory floor. Decisions made indoors may have consequences for what happens outdoors, including the choice to engage in organized political or criminal violence.

## 2 Research Design

### 2.1 Sample and timeline

The experiment took place during the period September 2017- February 2018 in two sites: Berkeley, California and Nairobi, Kenya. During this period, we conducted 144 experimental sessions with 12 participants each, and each participant was randomly assigned to either the heat treatment

(described below) or the control group.<sup>2</sup>

Participants were drawn from the student body of the University of California, Berkeley for the California sample and the University of Nairobi for the Kenya sample.<sup>3</sup> Although one should bear in mind that university students are not necessarily representative of the population as a whole (Sears, 1986), the use of student subjects should not pose a problem for the study's external validity per se (Druckman and Kam, 2011).

To be able to detect moderate and even relatively small standardized differences between treatment and control groups, we aimed for a sample of at least 1600 participants, with 800 at each site.<sup>4</sup> A test with size  $\alpha = 0.05$  and power  $(1 - \beta) = 0.80$  would allow us to detect differences of 0.158 standard deviations in the pooled sample, and of 0.223 standard deviations in each site separately (see Figures 1, 3 and 5 in the pre-analysis plan).<sup>5</sup> Since we want to compare many outcomes between treatment and control, we also account for the fact that the false discovery rate increases mechanically with the number of comparisons. We therefore adjust the minimum detectable effect (MDE) sizes for multiple comparisons under two scenarios: (i) the proportion of tested null hypotheses that are true is 0.50, and (ii) the proportion of tested null hypotheses that are true is 0.80. In the first scenario, the MDE is essentially unchanged by the adjustment. In the second scenario, it rises to 0.188 of a standard deviation in the pooled sample, and to 0.264 of a standard deviation in each site separately. The eventual sample size exceeded the target with 903 participants in Berkeley and 1,015 participants in Nairobi.

---

<sup>2</sup>In practice, while we over-recruited at 50% to better ensure 12 participants per session, we were not always able to secure 12 participants. In Nairobi, we had 79 full sessions of 12, 5 sessions of 11, and 2 sessions of 6. In California, we had 65 sessions of 12, 9 sessions of 11, and 4 sessions of 6. For the sessions with 6 participants, we randomized all into either the heat treatment or the control group. For the sessions with 11, we randomized which treatment group would have 5 participants.

<sup>3</sup>Due to the University of Nairobi temporarily closing because of political events, students were also recruited from Strathmore University for a period of time. Please see Experimental Protocol, Design, and Timeline Section J for more details.

<sup>4</sup>As noted above, the experimental design, protocol and core analyses were pre-registered in a pre-analysis plan on the American Economic Association RCT registry, at <https://www.socialscienceregistry.org/trials/1361>.

<sup>5</sup>In these calculations, we assumed equal sample sizes as well as equal variance across treatment arms. Standard errors were clustered by estimating the intra-session correlation of the outcomes of interest from early experimental pilot data.

## 2.2 Experimental temperature manipulation and measurement

The experimental treatment consisted of a temperature manipulation. Each session was randomly selected to take place in either a ‘hot’ treatment room or in a ‘cool’ control room. The target temperature in the control room was 22°C, while the target temperature in the treatment room was 30°C. To expose participants to the treatment, they were placed in the temperature-controlled laboratory rooms for a 20 minute waiting period before beginning the experimental modules. The entire lab lasted for roughly one hour (further information and photos of the experimental set up are included in Experimental Protocol, Design, and Timeline Section I).

In California, temperature in the treatment room was manipulated by a combination of two radiator heaters and an air conditioning (A/C) unit. The control room contained an identical A/C unit. In Nairobi, the temperature was controlled via a joint heating plus A/C wall-unit in both treatment and control rooms, plus an additional heater in the treatment room.<sup>6</sup>

During the experiment, we recorded both indoor air temperature as well as operative temperature, since the latter is considered the main driver of indoor thermal comfort. Indoor air temperature was measured by two sensors in each of the rooms, installed at either end of the room. Operative temperature, which is essentially an average of air temperature and mean radiant temperature, was measured using a sensor placed inside a ping pong ball coated in paint hanging in the middle of the room.<sup>7</sup>

In addition to temperature, we also measured a number of other environmental characteristics that have been shown to potentially affect cognitive performance (see Wargocki et al., 2006). To select the list of indicators, we draw upon extensive research in building science on the effects of indoor environment on cognitive performance and workplace productivity. Specifically, we recorded relative humidity, carbon dioxide, lighting, and background noise (see Experimental Protocol, Design, and Timeline Sections H.3-H.5 for details). All temperature equipment and sensors

---

<sup>6</sup>For a more detailed description of the laboratory environment and the types and brands of temperature equipment used, see Experimental Protocol, Design, and Timeline Section H.1.

<sup>7</sup>For the types and brands of sensors used in the experiment, see Experimental Protocol, Design, and Timeline Section H.2.

except for the operative temperature sensor were hidden from participants' view.

To be able to describe and control for the ambient climate at each experimental site as a robustness check, we also recorded daily outdoor temperature and relative humidity directly at both sites for the duration of the experiment.

## 2.3 Study sites

The two experimental sites, Kenya and California, were intentionally selected because their socio-economic characteristics are distinct and might indicate whether or not the response to thermal stress differs across locations and contexts. Kenya is a lower middle-income country in East Africa, which had a GNI per capita of 3,970 PPP\$ and was ranked 142 out of 189 countries on the Human Development Index at the time of the experiment. California is a federal state of the United States of America, a high income country with a GNI per capita almost fifteen times that of Kenya (61,020 PPP\$) and which is ranked 13th on the Human Development Index (all statistics from 2017).

Both sites are located in the temperate climate zone. The daily temperature in Nairobi was fairly tightly distributed around a median of 20°C for the duration of the experiments (Figure 1 Panel B), while the temperature in Berkeley, California, was more variable and at a lower median of 15.7°C (Figure 1 Panel D). Daily median relative humidity was 65.4% in Nairobi and 57.8% in California (see Appendix Sections A.2 and B.2, as well as Appendix Table C.1 Panel C for further details).

The political environment in the two sites also differs in ways germane to our experiment. The United States is considered a relatively stable and mature democracy, while Kenya is more politically volatile with a recent history of controversial national elections marred by election fraud accusations and resulting ethnic violence (Mebane, 2017). Our experiment occurred during a period of heightened instability in Kenya immediately following the presidential election between Raila Odinga (a member of the Luo ethnic group) and the incumbent Uhuru Kenyatta (a member of the Kikuyu ethnic group) on the 8<sup>th</sup> of August 2017. Kenyatta won the general election, followed by an annulment of the results by the Supreme Court due to election discrepancies (September

1). Kenyatta won a second round of voting that was boycotted by the opposition (October 26) and generated accusations of voter intimidation and fraud. Violent clashes resulted, including in Nairobi, with many victims reported to be ethnic Luos supporting Odinga (The Economist, 2017; Kiruga, 2018). This political context was salient during our study, as many experimental sessions in October were postponed because violence caused university closures. No similar disruption occurred in Berkeley, California during the study (see Figure 2 for a timeline of events and Appendix Sections A.1 and B.1 for details).

Beyond these broad differences between sites, there are also differences specific to the two student samples. Subjects in California are on average slightly younger ( $20.0 \pm 2.7$  yrs vs.  $22.0 \pm 2.4$  yrs in Nairobi), more likely to be female (60% vs. 35%), and more likely to have at least one parent with university education (78% vs. 25%) and higher income levels (89% vs. 8% earn  $> \$24,000$  US, PPP adjusted; see Appendix Table C.1). They also generally completed experimental tasks somewhat faster in the lab.

## 2.4 Experimental Outcomes

We measure the effect of temperature on a broad range of economic preferences, social behavior, and cognition outcomes. We use experimental modules that are standard and well-known in the literature (see Panel E of Figure 1).<sup>8</sup> We elicit standard economic preferences in both risk and time settings. We observe social behaviors such as fairness, cooperation, trust, and destruction in how participants interact with each other; additionally we measure participants' inclination for charitable giving. We also elicit subjects emotional state or *affect* using a self-reported survey. We measure cognitive performance with incentivized tasks on precision, fluid intelligence, and cognitive reflection.

Subjects are never deceived and their performance and choices during the experiment determine their monetary earnings. In the social behavior modules, individual earnings depend on interactions with other participants in the same room, anonymized and mediated through a com-

---

<sup>8</sup>The ordering of modules was fixed across treatments and sites, see Figure 1, Panels A and C.

puter terminal. The experimental interface incentivizes decisions with tokens, which are converted into local currency at the end of the experiment. The exchange rate is 3.5 tokens to 1 Kenyan shilling in Nairobi and 2.1 tokens to 1 cent in Berkeley.<sup>9</sup> After all experiments were completed, average earnings in California were around \$35, while average earnings in Nairobi were around 2000 Kenyan shillings.

In describing the modules and their outcomes, we follow our pre-specification and distinguish between “primary outcomes” and “secondary outcomes” of interest. Below is an overview of the experimental outcomes; further information about the design of each module (including the full transcript of the instructions) can be found in Experimental Protocol, Design, and Timeline Sections I.1-I.4.

#### 2.4.1 Economic Preferences

The risk module elicits risk preferences using choices over lotteries from two different menus (Eckel & Grossman, 2008). A primary outcome is the participant’s risk preferences, where choosing a lottery with higher variance is defined as being more *risk-taking*.<sup>10</sup> Risk preferences are measured by the outcome variance in the chosen lottery from the first menu. Another primary outcome is the quality of decision-making which is measured by *violation(s) of rational choice*. Comparing a participant’s choices from both menus, the decision-quality outcome of interest is an indicator for a transitivity violation. A secondary outcome is an indicator for a first-order stochastic dominance violation in the first menu.

The time preferences module consists of the standard protocol for eliciting so-called “beta-delta” preferences, namely a convex time budget (CTB) design following Andreoni et al. (2015). The primary outcomes of interest from the time preferences module include aggregate estimates of *time inconsistency* (parameter beta) and *patience* (discounting parameter delta). Secondary outcomes include an indicator for a violation of rational choice (GARP violation), as well as individual

---

<sup>9</sup>For context, 1 US dollar was equal to about 100 Kenyan shillings in 2017. Note that the exchange rate was previously 1.75 tokens to 1 cent in Berkeley, but was changed early in the experimental timeline (on September 28 2017), both to keep at the original target payment of \$25 an hour and to sustain recruitment.

<sup>10</sup>These are also sometimes called risk-tolerant or risk-loving preferences.

rankings of choices from each menu.

### 2.4.2 Social Behavior

The fairness module uses a real effort dictator game to measure preferences between oneself and another. Participants first earn tokens from a precision task (see Gill and Prowse, 2012, and the short description below). They are then paired with an anonymous partner who earns the same amount, and decide how to distribute the shared earnings. Such earned income generates clear entitlements for the dictator game (Kahneman et al., 1986; Forsythe et al., 1994), and the fair outcome given by entitlements is always an equal split. All participants act as dictators and know that their decision is implemented with fifty percent probability (Almås et al., 2010). The primary outcome of interest is *fairness* measured as the share of the total initial amount of tokens that is allocated to the other participant.

The cooperation module consists of a public goods game with three players, which encourages cooperation by doubling all contributions to a common pool (Fischbacher et al., 2001). The primary outcome of interest is *public contribution*, defined as the amount of tokens the individual placed into the group fund. A secondary outcome is having correct beliefs, which is defined through an indicator variable for whether the individual guesses correctly about another's contribution within the group.

In the trust module, one player in a pair is given an initial amount of tokens that can be sent to the other player. As in the public goods game, pro-social behavior is encouraged through a multiplier: any transfer sent is multiplied by three. The other player then decides to return any share of the amount received back to the first mover, including zero (see Johnson and Mislin, 2011, for a similar design). The primary outcome of interest is *trust*, defined as the share of the total initial amount sent to the other participant. A secondary outcome is trustworthiness, defined as the share of tokens sent back to the first participant. We also use trust elicited from a non-incentivized survey question asking whether other people in general can be trusted as an exploratory outcome; this subjective measure of trust is given as a value coming from a 0 to 10 scale (see also Falk et al.,

2018).

To study anti-social behavior, we include a destruction module. The prevalence of destruction of others' assets has been established even in situations absent of any clear personal gain by the use of the Joy of Destruction (JOD) game (Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011).<sup>11</sup> Participants are anonymously matched in pairs and told how much their partner has earned from a fluid intelligence task (described below). Each participant can destroy any amount of their assigned partner's earnings, including 0. The computer can also destroy some of the remaining earnings: it destroys nothing with fifty percent probability, and any amount of the remaining earnings with equal probability. The partner does not know whether her earnings are destroyed because of the computer or because of the participant's decision. Thus, a participant's purposeful destruction is partially concealed by the computer's random destruction. Destruction of currency in both Kenya and the United States is illegal so earnings are in units of vouchers or gift cards, thus actual destruction (and not reallocation) can take place.<sup>12</sup> The primary outcome of interest is a measure of *joy of destruction*, defined as the proportion of the partners' earnings purposely destroyed.

The final experimental module measures charitable donation outside the lab, which is conducted after participants have been informed about their monetary earnings. This module asks the participant if they want to donate some of their payment to a charity organization of their choice. The primary outcome from this module is the amount of tokens participants choose to donate.

### 2.4.3 Affect

We measured self-reported emotional state using core affect, a combination of happiness and alertness (referred to in the literature as valence and arousal, respectively; see Russell and Barrett, 1999). We measure *happiness* by asking "On a scale from 1-7, with 1 being sad and 7 being happy, how do you feel right now?" We also measure *alertness* by including the question: "On a scale

---

<sup>11</sup>Anti-social behavior and punishment of others have by now been studied quite extensively, both in competitive settings where people cheat and sabotage competitors (Charness et al., 2013) and in studies of cooperation where participants punish non-cooperators (Ambrus and Greiner, 2012; Dreber et al., 2008; Wu et al., 2009; Gächter and Herrmann, 2008).

<sup>12</sup>In Berkeley we physically destroy \$1 Amazon gift cards, while in Nairobi we destroy Airtime vouchers each worth 50 Ksh.

from 1-7, with 1 being tired and 7 being alert, how do you feel right now?” Both measures were included as secondary outcomes.

#### 2.4.4 Cognition

The experiment also includes modules to measure individual cognition and precision. First, we use a slider task in which participants are asked to place a slider on an assigned number on a line ranging from 1-100 (Gill and Prowse, 2012). The primary outcome of interest is *precision*, measured by the number of sliders correctly placed within three minutes. Second, *fluid intelligence* is measured using Raven’s matrices and the primary outcome of interest is the proportion of puzzles solved correctly (Penrose and Raven, 1936). Finally, we use a (non-incentivized) *cognitive reflection* test (CRT) consisting of five survey questions (Frederick, 2005). The primary outcome of interest from the CRT is the share of survey questions answered correctly, and the secondary outcome is the time spent answering the questions.

### 2.5 Survey Data

In addition to the data generated in the experimental modules, we conduct a non-incentivized demographic survey, which collects information on each participant’s height, weight, age, gender, and ethnicity (Kenya) or in-state residence status (California), as well as information about their parental employment, income and education.<sup>13</sup> The survey also collects a subjective trust measure (See “Experimental Protocol, Design, and Timeline” Section I.5 for the full demographic survey).

We also use a post-experiment debriefing survey, which asks participants about their perceptions of the laboratory environment (Experimental Protocol, Design, and Timeline Section I.6 contains the full post-experiment survey). The purpose is to gauge whether participants in the treatment condition could infer that room temperature is the study’s treatment. The survey includes questions specifically asking about the comfort of their seats, computer screen brightness,

---

<sup>13</sup> Appendix Table C.2 displays response rates to survey questions. For the pooled sample, we cannot reject balanced non-response rates to the demographic questions across treatment arms at traditional confidence levels.

air quality, room lighting, space, noise, and room temperature. *Discomfort* is measured by asking participants whether they experience the temperature in the room as “too cold,” “comfortable,” or “too hot.” This is followed by measuring *treatment salience* in which participants are asked which topics they think the experiment is studying, where respondents could choose from the aforementioned list, as well as “generosity” and “gender discrimination.” The post-experiment debriefing survey was introduced half-way through the study and data is thus available for 913 of the 1,918 study participants.

## 2.6 Balance

At each site, subjects are randomly assigned to the Hot or Control condition, such that within-site treatment effect estimates are not confounded by differences in individual-level characteristics. As pre-specified, we check for successful randomization by comparing demographic survey averages between treatment and control in each site using a two-sided t-test.<sup>14</sup> Randomization was successful in that subject populations in the Hot condition are statistically indistinguishable from the Control population in California along all dimensions, and differ in Nairobi only by being slightly heavier on average ( $63.0 \pm 9.3$  kg vs  $61.3 \pm 7.8$  kg, significant at the 5% level), see Appendix Table C.1. Panel B of the same table shows that treatment and control sessions were also balanced in terms of outdoor temperature and relative humidity, which is to be expected since we almost always ran treatment and control sessions simultaneously.

## 3 Estimation

The main specification estimates the treatment effect of temperature on the outcome of interest as

$$Y_i^k = \beta_0^k + \beta_1^k Heat_i + \beta_2^k Male_i + \beta_3^k Nairobi_i + \varepsilon_i^k \quad (1)$$

---

<sup>14</sup>The following variables from the survey are tested for balance: the participant’s weight, height, age, gender, in-state residency (Berkeley) or ethnicity (Nairobi), father’s occupation status, mother’s occupation status, combined parental income, father’s highest level of education, and mother’s highest level of education.

where  $i$  refers to the individual,  $k$  to the module outcome (e.g., trust, time preference, etc.),  $Y_i$  to an outcome of interest,  $Heat_i$  is the treatment indicator for being placed in the hot room,  $Male$  is an indicator for the individual's self-reported gender being male,  $Nairobi$  is a fixed effect for the experimental location being Nairobi, and  $\varepsilon_i$  is an idiosyncratic error term. Standard errors are clustered at the session level in all regressions and tests.<sup>15</sup> The specification described in (1) is employed to estimate the treatment effect of heat on precision (number of correct sliders), fairness (share of tokens earned from the precision task given freely to the participant's partner) risk preferences (variance of the chosen coin toss and rational choice violation), trust (share of tokens sent by individual A), public contribution (amount transferred to the public fund in the public goods game), fluid intelligence (number of Raven's matrices correctly solved), cognitive reflection (share of correct answers and time spent on the cognitive reflection test) and destruction (share of gift cards purposefully destroyed in the JOD game). The treatment effect of heat on self-reported alertness, happiness and willingness to trust others is also estimated in this way.

As prespecified, we depart from this specification for two families of outcomes: (i) when estimating the treatment effect of heat on charitable donations we include an additional indicator variable in eqn. (1) that takes the value 1 if the individual donates to an organization based locally for the California sample or based in their home region for the Nairobi sample, and we additionally control for total lab earnings; and (ii) for the time preference module, we structurally estimate time inconsistency (aggregate estimate of beta and its standard error) and discounting (aggregate estimate of delta and its standard error) using non-linear least squares, as in Andreoni et al. (2015). Differences in treatment and control are then compared using a t-test.

To understand whether treatment effects differ by gender or by site, we also estimate a specifi-

---

<sup>15</sup>The main specification in the pre-analysis plan includes  $X_i$ , a vector of covariates that differ significantly (at the 5% level) between treatment and control. Only self-reported weight in the Nairobi sample is unbalanced. However, we chose not to include weight as a regressor in the results reported in Section 4 because of a non-trivial amount of missing data for the weight variable. We examine robustness of the results with respect to inclusion of weight as a regressor in Supplement Section F.2, and demonstrate that its inclusion does not alter the results.

cation with heterogeneous effects:

$$Y_i = \beta_0^k + \beta_1^k Heat_i + \beta_2^k Male_i + \beta_3^k Male_i \times Heat_i + \beta_4^k Nairobi_i + \\ \beta_5^k Nairobi_i \times Heat_i + \varepsilon_i^k \quad (2)$$

where the regressors and outcome variables are the same as above and "×" denotes an interaction between two variables.<sup>16</sup> Again, the charitable donation regression includes additional covariates controlling for in-group bias. The heterogeneous treatment effect for time inconsistency and discounting is estimated by comparing (i) differences in means between treatment and control among males to differences in means between treatment and control among females, and (ii) differences in means between treatment and control among Berkeley participants to differences in means between treatment and control among Nairobi participants.

Finally, in addition to the main specification and the heterogeneity specification, we also employ a "California specification" and a "Nairobi specification" that simply carry out the main specification among participants at each site. Several regressions involving more exploratory outcomes, which we discuss below, include additional covariates to the main and heterogeneity specification.

As pre-specified, we adjust statistical significance levels accounting for the total number of null hypotheses considered. Specifically, alongside traditional p-values, we present false discovery rate adjusted "q-values" that limit the expected proportion of false rejections of a true null within a family of hypotheses, as in Benjamin et al. (2018). Rather than specifying a particular q-value (and thus false discovery rate), we report the minimum q-value at which each hypothesis is rejected (see Anderson, 2008). The first family of hypotheses for which we perform multiple hypothesis adjustments are the set of p-values associated with the treatment effect of heat  $\beta_1^k$  in the main specification across all  $k$  primary outcomes of interest (Panel E of Figure 1) for the pooled sample as well as each site sample. The second family of hypotheses comes from the heterogeneity speci-

---

<sup>16</sup>As in the main specification, the pre-analysis plan included in (2) additional covariates that differ significantly between treatment and control at the 5% level. Since the results are robust to exclusion of these covariates, we relegate the heterogeneity specification with these additional controls to Supplement Section F.2 for primary outcomes and Section F.3 for additional outcomes.

fication, where we report sharpened q-values for heterogeneous treatment effects of temperature related to gender ( $\beta_3^k$ ) and site ( $\beta_5^k$ ) over all  $k$  primary outcomes of interest in the pooled sample.

## 4 Results

Here we present the treatment effects of heat on the primary outcomes of interest as well as the secondary exploratory outcomes. To preview, we find substantial “first stage” effects as the heat treatment had a meaningful effect on actual measured room temperature as well as on participants’ happiness and alertness (Section 4.1). Yet nearly all components of economic decision-making and judgment are unaffected by the increase in thermal stress, with the exception of destructive behavior, which increases significantly in the heat treatment in Nairobi and decreases somewhat in California (Section 4.2). In Section 4.3, we further examine the mechanisms underlying these destruction results. Section 4.4 demonstrates robustness of the core results with respect to an array of sensitivity checks.

### 4.1 Was the Temperature Treatment Meaningful?

Whether a study design is powered to detect statistically significant effects of a temperature manipulation on participants’ behavior depends on the “dosage” of heat and on participant sample size. We now discuss where on the dosage-sample size spectrum our study is located, and argue that the design leaves us well-powered statistically.

Regarding the temperature manipulation, our treatment lies towards the higher end of the ranges of temperatures examined in related research. On one extreme, the temperature and aggression studies by Paul Bell and Robert Barron, (see e.g. Bell and Baron, 1976) administer hot treatments between 32°C to 35°C relative to control treatments of roughly 23°C. However, in the experimental heat-productivity literature, temperature ranges are often much narrower, examining changes of a few degrees in the 20-30°C range, although there are several exceptions, with one study examining temperature up to 39.4°C (Ye et al., 2010). Temperatures in our study are also similar to those

examined in recent observational studies, for example Park (2020) on the effect of temperature on exam performance and Jacob et al. (2007)'s study of crime rates. In the former, 90% of temperature observations lie between 21°C and 35°C, and in the latter there is a larger variation because they also examine colder temperatures. Similarly, the length and coarseness of treatment in our study are also in line with the relevant literature: like us, most experimental studies consider just one heat treatment condition, with exposure times varying from less than an hour up to 2-3 hours.

In the second dimension, our study is a clear outlier (in a good way) in terms of participant sample size relative to the existing experimental literature: the lab studies referred to above generally have sample sizes in the range of 10-60 participants, while we have nearly a thousand participants in each of our two study sites, leaving us far better powered statistically (see calculations in Section 2.1).

Importantly, we find that the temperature manipulation was successful in impacting actual measured room temperature, as well as individual affect, which could be thought of as prerequisites for subsequent behavioral responses. Specifically, mean measured temperature in the hot room was 31.3°C in Nairobi and 30.2°C in Berkeley, while mean temperature in the control rooms were roughly 10°C lower, at 21.7°C in Nairobi and 21.8°C in Berkeley. Consistent with this, in the post-experiment survey 91% of participants in Berkeley and 52% of participants in Nairobi described temperature in the treatment room as ‘too hot’, while effectively no one in the control room gave this description (see Appendix Table C.3). The temperature treatment also had an effect on self-reported alertness, which decreases by a significant 0.24 standard deviation both in the pooled sample and in each of the sites. Respondents in Nairobi additionally reported significantly lower happiness levels (see Appendix Tables D.2.11 and D.2.12).

Taken together, the study design featured a meaningful dosage of high temperature (in terms of level and exposure time), large participant sample sizes (far exceeding most existing work), and we find that important intermediate variables were impacted, as intended. We conclude that our study is well-designed to detect significant behavioral responses to high temperature, if they exist.

## 4.2 Main results

Despite the strong “first stage” effects on room temperature and individual affect, we find that nearly all components of economic decision-making and judgment are largely unchanged by the randomly administered thermal stress (Table 1; Figure 3 displays standardized treatment effects and q-values graphically). The effect sizes themselves are also generally small and precisely estimated. Thus, the failure to detect statistically significant and substantive impacts does not appear to be a consequence of small sample sizes or an insufficient temperature manipulation, as the study is well powered. Further, we cannot reject that subjects in Nairobi and California respond similarly to high temperature across nearly all outcomes. We do, however, find a large, positive, and highly statistically significant effect of heat on destructive behavior in the pooled sample, and importantly, this result is robust to multiple hypothesis testing adjustments. There is heterogeneity in this effect, with a significant positive destructive response to heat in Nairobi and a negative (and marginally significant) response in California. Below, we discuss the findings across primary outcomes and in the different sites in detail.

We find a small positive impact of temperature on individual cognition and precision in our main specification (1). Subjects in the hot treatment increase their precision slightly – they complete half an additional slider task relative to a baseline of 18 completed sliders in the control group – but the effect is small (0.07 of a standard deviation) and no longer significant once we adjust for multiple hypothesis testing. The treatment effects of temperature on fluid intelligence and cognitive reflection follow a similar pattern: there is no negative effect of temperature on either of these measures and the point estimates in the pooled sample are precisely estimated zeros (0.08 of a standard deviation for the former and -0.01 for the latter). We examine heterogeneity in these patterns by estimating equation (2), and find no significant differences in the treatment effects by gender or study site.

We find no evidence that higher temperatures affect social preferences, and estimate treatment effects of negligible size on all outcomes of interest in this domain. In the real-effort dictator game, participants transfer on average 32% of joint earnings to their anonymous partner in the

control group, and this decreases by 1.7 percentage points (1 percentage point in Nairobi and 2.5 percentage points in California) amounting to a small and non-significant reduction of 0.07 of a standard deviation.<sup>17</sup> Heat does not affect how trusting a participant is towards others in the lab: in the trust game control participants entrust on average 42% of their endowment to an anonymous partner and heat increases this transfer by less than 1 percentage point in the pooled sample (resulting from a 2 percentage points decrease in California and a 3 percentage points increase in Nairobi), equivalent to an increase of 2% of a standard deviation.

Similarly, heat has no effect on contributions in the public goods game: contributions to the joint pool decrease by 9 tokens in the hot condition (relative to an average of 530 tokens out of 1200 initial tokens contributed in the control condition). This is equivalent to a decrease of 2% of a standard deviation in the pooled sample (resulting from a decrease of 0.09 standard deviations in California and an increase of 0.04 standard deviations in Nairobi). Post-hoc, we also consider the treatment effect on trustworthiness, measured as the share of tokens given back to the anonymous partner, and find effectively zero treatment effects in both the pooled sample and each site separately.

The final outcome of interest in the pro-sociality domain is charitable behavior, as measured by donations in the charity module. In the control treatment, participants donate 410 tokens (245 in California and 562 in Nairobi), equivalent to roughly \$1 dollar in both sites. Heat reduces donations by just over 1 percentage point (with a decrease of 0.02 of a standard deviation in California and an increase of less than 0.01 in Nairobi). We do not find any evidence of heterogeneity of treatment effects in the pro-social domain either by gender or study site.

Turning to risk and time preferences, we again find no effect of the heat treatment. Risk-taking decreases by less than 1% of a standard deviation in the hot condition and patience by less than a percent.<sup>18</sup> Similarly, we observe no effect of heat on the frequency of rational choice violations in

---

<sup>17</sup>We reach a similar conclusion through an additional pre-specified formulation for the fairness module, in which we include both an indicator for an individual being classified as having high precision from the Precision task as well as the number of correct sliders. See Supplement Table F.4.1 for details.

<sup>18</sup>Our results are robust to measuring risk by 1) assuming a mean-variance utility function, and then use the backed out lambda parameter as the outcome variable with OLS; and 2) using a categorical indicator of choice with an ordered logit and ordered probit estimation. Treatment is not statistically significant within these specifications, in line with

the risk module or time inconsistency in the time preference module. Overall, we conclude that many fundamental elements of economic decision-making and judgment are unaffected by thermal stress, despite the large observed declines in self-reported alertness (all samples) and happiness (in Nairobi) (Figure 3, bottom panel).<sup>19</sup>

There is, however, a large and highly statistically significant effect of thermal stress on destructive behavior in both sites with effect sizes that survive multiple hypothesis testing. In the pooled sample across both sites, individuals choose to destroy more of the winnings of other players in the JOD module ( $p$ -value = 0.069) (Abbink and Sadrieh, 2009). In Nairobi, subjects destroyed an additional 8 percentage points (0.30 of a standard deviation) of other players' earnings in the Hot condition relative to Control – equivalent to an increase of roughly 50% relative to the 14% destruction rate in the control group). In contrast, destruction falls at higher temperatures in California (with a  $q$ -value of = 0.053 after multiple testing adjustment).<sup>20</sup> We also find a significant interaction between the heat treatment and the male indicator variable in the pooled sample: while men destroy more vouchers than women in this task overall, the increase in their destruction levels as a result of the temperature manipulation is smaller than that found among women.

The finding of a large negative effect of heat on destruction in Nairobi without apparent provocation mirrors results from existing experimental studies on temperature and aggression, although the main analysis involving our preferred risk outcome. See Supplement Section G.2 for details.

<sup>19</sup>In Appendix Section D.2, we examine the treatment effect on a range of secondary outcomes. Specifically, we examine the effect of heat on the likelihood of rational choice violations (relating to First-Order Stochastic Dominance in Appendix Table D.2.1 and relating to the Generalized Axiom of Revealed Preference in Appendix Table D.2.2), on non-parametric estimates of time preferences (Appendix Tables D.2.3-D.2.6), on experimental and survey measures of others' trustworthiness (Appendix Tables D.2.7 and D.2.8), on forming correct beliefs about others' contributions in the public goods game (Appendix Table D.2.9) and on time taken to complete the cognitive reflection test (Appendix Table D.2.10). The regressions here follow the format of (1) and (2). We find no significant treatment effects of heat exposure on any of these additional outcomes. The inclusion of weight as a regressor does not change this result (see Supplement Section G.3).

<sup>20</sup>To compare with other studies, we examine the effect of heat on *any* destruction, finding destruction rates (using the binary construction) to be 7 to 11% in California and 31 to 47% in Nairobi (see Supplement Table G.3.1). This aligns with the main results using share of gift cards or vouchers purposefully destroyed. Previous work using the JOD game has documented destruction rates among university students ranging from 8 to 40% in the Netherlands (Abbink and Sadrieh, 2009) and 10 to 26% in Ukraine (Abbink and Herrmann, 2011). Meanwhile, destruction rates ranged from 23 to 40% among pastoralists in Namibia (Prediger et al., 2014), with higher levels among the resource-scarce subsample. Bauer et al. (2018) find destruction rates to be higher among adolescents from rural Uganda (53 to 59%) compared to adolescents from disadvantaged regions in Slovakia (32 to 42%). The differences in destruction rates using a binary construction in this experiment is thus within the range of previously reported non-zero effect sizes.

those documented the effect with far smaller samples. For example, Bell and Baron (1976) find positive treatment effects of heat on aggression for unprovoked participants (which exceed those for provoked participants). On the other hand, and in contrast to the prediction of a U-shaped temperature-aggression relationship in the psychology literature, some small-sample studies find a negative relationship between temperature and aggression (Baron, 1972; Palamarek and Rule, 1979), which we confirm in the California sample. In the next sub-section, we set out to better understand the nature of these treatment effects on destruction behavior.

### 4.3 Understanding the effect of heat on destruction

We examine the factors generating the destructive response to heat in the Nairobi sample and the differential effect in the California sample, exploring potential mechanisms (called mediators in some social science fields) as well as characteristics associated with heterogeneous treatment effects, sometimes called moderating variables (Baron and Kenny, 1986).

This more speculative analysis, which was not pre-specified, builds on existing theoretical models of individual aggression (Bell and Baron, 1976; Zillmann, 2006; Berkowitz, 1989; Anderson et al., 2000) as well as evidence that political and economic marginalization contribute to the empirical link between extreme climatic events and conflict in Sub-Saharan Africa (von Uexküll et al., 2016; Prediger et al., 2014). In particular, we examine whether there is evidence of heterogeneous treatment effects along three pre-determined characteristics, namely, (i) socioeconomic status, (ii) cognition (measured as a weighted average of scores in the fluid intelligence and the cognitive reflection tasks), and, in the case of Nairobi, (iii) ethno-political exclusion. For the latter, we study the response to heat among both politically dominant and marginalized ethnic groups in the sample, given the timing of a controversial national election during data collection (see Section 2.3 for details). We also focus on four potential mechanisms: (i) happiness, (ii) alertness, (iii) discomfort, (iv) and the salience of high temperature as the treatment.<sup>21</sup> Figure 4 summarizes the

---

<sup>21</sup>Happiness is emphasized as an important mechanism or mediator in the negative-affect model (Bell and Baron, 1976), which posits that temperature causes aggression by decreasing affect (happiness). Alertness (arousal) and treatment salience are important variables in excitation transfer theory, which posits that arousal due to tempera-

variables considered as interaction effects and mechanisms in the analysis.

To estimate the heterogeneous treatment effects and begin to identify mechanisms that drive the observed heat-aggression relationship, we estimate a general linear structural equation model (Imai et al., 2010):

$$M_i = \alpha_1 + \beta_1 T_i + \varepsilon_{i1} \quad (3)$$

$$Y_i = \alpha_2 + \beta_2 T_i + \gamma M_i + \kappa T_i \times M_i + \varepsilon_{i2} \quad (4)$$

where  $M_i$  is the mechanism and/or interaction of interest,  $T_i$  is the treatment indicator, and  $Y_i$  is the outcome, for instance, the share of vouchers destroyed in the JOD game. We carry out the analysis separately in each study site.

From regression (4),  $\kappa$  captures the extent of treatment heterogeneity along the dimension of variable  $M_i$ . Similarly, a meaningful  $\beta_1$  in (3), together with a significant difference in destruction as  $M_i$  varies (captured by  $\gamma$  and  $\gamma + \kappa$  in (4) for a binary  $M_i$ ) suggests that the variable  $M_i$  is plausibly a mechanism driving the treatment effect.<sup>22</sup> For pre-determined variables, we only estimate regression (4) since the sample is balanced for these variables as a result of the experimental design.<sup>23</sup>

---

ture may be misattributed to another stimulus (frustration/irritation), and thus trigger aggression, but only if the true source of the initial stimulus, i.e. the temperature increase, is not salient (Zillmann, 2006). Discomfort is relevant in cognitive-neoassociation theory (Berkowitz, 1989), which posits that aggression is associated with discomfort, extreme temperature increases discomfort, and thus discomfort triggers aggression. Finally, cognition is a moderator of the temperature-aggression relationship in stages I and II of the general-aggression-framework (Anderson et al., 2000), which describes episodes of aggression by means of a 4-stage framework whereby personal characteristics – such as cognitive ability – may affect how easily hostile thoughts and scripts are accessed and how an aggressive cue is appraised, and hence whether an aggressive response is triggered or not. Note that this model does not consider the possibility that cognitive ability may itself be affected by the temperature stimulus.

<sup>22</sup>In causal mediation analysis  $\delta(t) = \beta_1(\gamma + \kappa(t))$ , where  $t = T, C$ , denotes the causal mediation effect (conditional on treatment status) for a binary variable  $M_i$ . This measures by how much average destruction levels would change if one varied the mediator  $M_i$  from the average value observed under the control condition to the average value observed under the treatment condition holding differences in destruction levels for  $M_i = 1, 0$  constant at  $t$ . The direct treatment effect, which is estimated conditional on the average value of  $M_i$  at  $T = t$  is then given by  $\zeta(t) = \beta_2 + \kappa(\alpha_1 + \beta_1 t)$ .

<sup>23</sup>The distinction between interactions and mechanisms may be less definitive in practice, and there may be more of a continuum between the two, as observed variation in variable  $M_i$  (below) may be partly driven by both pre-determined differences (say, in participants' underlying cognitive ability) and variation due to the experiment (i.e., if temperature affects cognitive performance). The separation of variables into interactions versus mechanisms below reflects our view regarding which of these sources of variation is more consequential in the study samples.

There are differences between the two study sites in terms of how the treatment effect depends on individual characteristics along some dimensions but not others. Individual socioeconomic status, as proxied by parental education, does not appear to affect the magnitude of the heat treatment effect in either Nairobi or Berkeley (Figure 5, Panels B and E and Appendix, Tables E.1 and E.2). Yet in Kenya we do find evidence that participants with higher individual cognitive ability show a more blunted aggressive response to heat: study participants in Nairobi with below median cognitive scores had a significantly larger ( $p = 0.020$ ) destructive response to heat (12%) than those with above median cognitive scores (4%, Figure 5 Panel C). Conceptually, this could be because it is easier for those with greater cognitive ability to control aggressive thoughts and to appraise the situation in a more level-headed way (Anderson et al., 2000). However, there is no meaningful interaction effect with individual cognitive ability in the California sample (Panel F). The reasons for the differential role played by individual cognitive ability in the two sites, in the context of main effects with opposing signs, would be a useful direction for future research.

There is also evidence that the aggressive response to heat seen in Nairobi is related to the ongoing political marginalization and associated frustration of non-Kikuyu, and especially Luo, ethnic communities: Nairobi subjects that identify with an ethnicity other than Kikuyu – the ethnic group of the president controversially re-elected during the study period – have a much larger increase in destructive behavior under thermal stress (Figure 6, Panel B and Appendix, Table E.1). Ethnic Kikuyu subjects, who made up 20.6% of the Nairobi sample, exhibited no destruction response (Control = 14%, Hot = 12%). The largest response was visible in ethnic Luo subjects (20.8% of the sample), where heat caused destructive behavior to more than double (12% vs. 27%), while other ethnic groups exhibited a response of intermediate magnitude. The difference between ethnic Luo and Kikuyu responses is statistically significant ( $p$ -value = 0.003). Among members of other ethnic groups, those predominantly aligned with the political opposition show a response to high temperature that is somewhat larger than that found among members of largely pro-government ethnic groups, mirroring the pattern among ethnic Luo and Kikuyu participants (see Figure 6 Panel C for additional details, as well as Appendix Table C.4 for how these groups

were constructed from self-identified ethnicity).<sup>24</sup>

Importantly, the results taken together do not suggest that there is anything fundamentally different in the judgment or economic decision-making processes of Nairobi non-Kikuyu participants. The results for all other modules suggest that these subjects exhibit patterns of decision-making that are otherwise statistically indistinguishable from other subjects (see Supplement Section G.5); moreover, average rates of destruction in the Control condition are nearly identical across ethnic groups in Nairobi. We hypothesize that if ethnic Kikuyu – or California – subjects were politically marginalized similarly to the recent experience of ethnic Luos in Kenya, they might exhibit a JOD-temperature response similar to ethnic Luo participants.

We next turn to the analysis of underlying mechanisms. Of course, this analysis is suggestive, since there is just one treatment (the hot room) and we are considering how it affects multiple mechanisms, all of which could be interconnected in complicated ways. Concerns about mediation analysis are also linked to the well-known problem of “bad control” (Angrist and Pischke, 2009). However, the suggestive patterns we present below do shed light on some possibilities that are useful for interpretation and could inspire future research.

First, there is no clear evidence that happiness or alertness are key mechanisms (Appendix Tables E.3 and E.5). As described in Section 3, treatment significantly decreases alertness in both sites ( $-0.28$  points on a scale from 1 to 7,  $p = 0.031$  in Nairobi, and  $-0.50$  points in California,  $p = 0.004$ ), but destruction rates do not covary with either of these variables.

Second, there is some evidence consistent with the neo-cognitive association model in the California sample: treatment strongly and significantly increases discomfort ( $\beta_1 = 70$  percentage points,  $p < 0.001$ ) and those who experience the treatment room as “too hot” destroy more vouchers ( $\gamma + \kappa = 1.8$  percentage points,  $p = 0.610$ ) than those who are comfortable. Moreover, conditional on experiencing discomfort, the treatment effect of temperature on aggression is positive ( $\beta_2 + \kappa = 2.3$  percentage points with  $p = 0.295$ , and the difference in treatment effects between

---

<sup>24</sup>Destruction does not depend on the share of other ethnicities in one’s assigned room; this may be expected, given that the laboratory setting was designed to keep participants’ views towards their tasks (see Supplement Table G.4.1 for details).

comfortable versus uncomfortable participants is  $\kappa = 8.3$  percentage points with  $p = 0.074$ ). These results suggest that the negative average treatment effect in California may be driven in part by behavior in the control room and among those who are comfortable in either condition. Indeed, and consistent with this interpretation, the treatment effect conditional on experiencing comfort is large and negative ( $\beta_2 = -6$  percentage points,  $p = 0.118$ ). Moreover, those who experience temperature as too cold in the control room destroy fewer vouchers than those who say they feel comfortable ( $\gamma = -6.5$  percentage points,  $p = 0.017$ ). Behavioral change in the cold condition is not the focus of this study, and was unexpected, yet suggests that discomfort can affect individual behavior at both extremes of the temperature scale.

The salience of the temperature treatment emerges as a plausible mechanism driving the negative treatment effect (reduction in destruction) in California: destruction rates are lower in the hot room if temperature is salient to the individual ( $\gamma + \kappa = -1.8$  percentage points,  $p = 0.567$ ), and treatment simultaneously meaningfully increases the share of participants for whom this is the case ( $\beta_1 = 36$  percentage points,  $p < 0.001$ ). Together, this implies a large estimated negative treatment effect for those who guessed the experiment's purpose (and thus understood that they were deliberately placed in a hot room, with  $\beta_2 + \kappa = -9.5$  percentage points,  $p = 0.317$ ), while the treatment effect is close to zero for those who do not ( $\beta_2 = -1.6$  percentage points,  $p = 0.592$ ). Thus a leading possibility is that the negative treatment effect in California is generated by a sharp increase in the share of participants for whom temperature is salient, and that those participants reduce aggression when temperature increases, consistent with the predictions of excitation transfer theory (Zillmann, 2006).<sup>25</sup> There is no systematic interaction between awareness of the underlying temperature treatment and the treatment effect for the other primary outcomes of interest (as shown

---

<sup>25</sup>Briefly, excitation transfer theory posits that arousal is non-specific across a range of negative conditions and that arousal due to temperature may be misattributed to another stimulus (frustration/irritation) for which aggression would be an appropriate response. An important prerequisite for this mechanism is that misattribution, and thus an effect from temperature on aggression, can only happen if temperature is not salient as a cause of arousal. In fact, if temperature does become salient, the theory posits that misattribution may happen in the other direction (from non-temperature induced arousal to temperature induced arousal), resulting in less aggression in extreme temperatures than in moderate ones. The finding of higher destruction levels among those for whom temperature is not salient in the treatment room, as well as a large negative treatment effect for this group and a near zero effect for participants who are aware of the experiment's purpose, are all consistent with this theory.

in Supplement Section G.6), suggesting that this is unlikely to be due to a general experimenter demand effect.

In contrast, in Kenya we do not find any evidence that the same channel is at work, as there is no clear treatment effect on the salience of temperature or awareness of the underlying treatment (Appendix Table E.3). Participants were perhaps less struck by the temperature manipulation because steady indoor temperature control via central heating and air conditioning is less common in individuals' daily experience in Kenya than in California.

#### 4.4 Robustness

Additional robustness checks do not meaningfully alter the main conclusions reported above. First, the results are robust to controlling for additional covariates, including self-reported weight (see Supplement Sections F.2 for primary outcomes and F.3 for additional outcomes). As noted in Section 3, there is a small difference in individuals' self-reported weight between the treatment and control groups, significant at the 5% level in the Nairobi sample, and thus we follow our pre-specified strategy and show that all our results are robust to controlling for any characteristic that is not balanced between treatment arms.

Second, the results are robust to controlling for the gender composition of participants in the lab session; we also do not find any evidence of significant heterogenous treatment effects as a function of session gender composition. This analysis was motivated by the fact that the majority of participants in Nairobi were male, while most California participants were female (see Appendix Table C.1 and Supplement Section G.7). Third, to further address concerns regarding effects of any other session features on behavior, we demonstrate that results are robust to including session fixed effects (Supplement Section G.8). Fourth, we investigate whether accounting for outside temperature and relative humidity on the day of the session alter the main results. Nearly all results are unchanged when controlling for these factors (see Supplement Section G.9). Yet when outdoor weather is humid and hot, the additional heating in the room appears to change behavior in a few instances. In particular, there is a significant interaction between treatment and outdoor weather

in the charitable donation module: the treatment effect of heat on tokens donated is negative and significant when outdoor temperatures are above the sample average in the pooled sample, with an additional negative effect in Nairobi at above average outdoor humidity. The temperature treatment also has an additional effect in lowering alertness when outdoor temperatures are higher in Nairobi.

We calculate exact p-values, as specified in the pre-analysis plan, and find that these are quite similar to the asymptotic p-values and do not change the main conclusions, including for the Joy of Destruction results (see Supplement Section F.2). Finally, we conduct a test of joint significance of the treatment effects across nearly all of the primary outcomes, in a system of seemingly unrelated regressions.<sup>26</sup> We cannot reject the null hypothesis of no joint treatment effect in the pooled sample, but do reject the null in the Nairobi sample, which appears to be largely driven by the JOD game (see Supplement Tables F.2.13 and F.2.14).

## 5 Conclusion

Overall, the results indicate that, with the exception of destructive behavior, many fundamental elements of economic decision-making are largely unchanged by thermal stress across populations on two continents facing different baseline climates. This suggests that some population-level effects of high temperatures, such as reduced economic productivity (Burke et al., 2015) and labor supply (Graff Zivin and Neidell, 2014), may be predominantly caused by changes to external factors, such as crop physiology (Schlenker and Roberts, 2009) or other non-cognitive components of the economy (Carleton and Hsiang, 2016).

The finding that a politically disenfranchised population in Kenya exhibits an increased preference for harming others under thermal stress might partially explain previous findings that higher temperatures elevate rates of interpersonal violence and social conflict around the world (Hsiang et al., 2013; von Uexkull et al., 2016). Of course, the possible role of an internal mental channel does not preclude simultaneous influence of external factors, such as altered economic circum-

---

<sup>26</sup>The non-linear estimation approach used for the Patience and Time inconsistency outcomes prevents these from being included in the SUR estimates.

stances (Miguel et al., 2004), in generating human conflict via thermal stress. The finding of a negative treatment effect of temperature on destruction in California provides strong evidence that even the operation of an internal channel from temperature to aggression is context dependent.

We think the results on the relationship between heat and aggression are particularly rich and interesting because they replicate the wide range of treatment effects found across studies with much smaller sample sizes in one unified setting and a large sample. As a result, we are able to document a variety of mechanisms, many of which have been hypothesized in the theoretical literature. These range from ethnopolitical marginalization as a potential driver of the positive effect of heat on aggression in Kenya to temperature salience as psychological mechanisms underpinning the negative treatment effect in California.

The study aims to provide the first evidence on a potential internal channel linking temperature to economic decision-making and behavior using a controlled laboratory set-up that allows for causal inference. However, due to ethical concerns, we were limited in terms of the level and duration of high temperature exposure used in the lab, and as such, we view the study as an initial attempt to establish empirical facts relevant to the ongoing discussion regarding the causal effects of extreme temperature, and thus may inform future analysis. It is possible that different results may be achieved if subjects are exposed over a longer duration to temperatures higher than 30°C (to the extent such interventions are deemed ethical for human subjects). In terms of other extensions, the collection of individual biomarker data (Haushofer and Fehr, 2014) could shed light on any underlying physiological changes. It is also possible that other non-temperature stressors could similarly lead to increases in destructive and anti-social behavior among socially marginalized groups, and this remains a promising direction for future research.

We view this study's ability to provide novel evidence on the mechanisms underlying the link between extreme temperature and economic outcomes as a meaningful contribution to both the active scholarly and policy debates. The ability of human societies to mitigate the adverse impacts of climate through public policy may depend on the underlying channels of impact. Traditional policy responses and technologies can effectively manage several external factors, but appear less

suited to manage internal mental processes. The findings on destruction suggest that interventions that dampen the economic consequences of thermal stress – such as expanded rainfall insurance – alone might not break the link between temperatures and human conflict if it is indeed mediated by internal mental processes. Especially in high-risk contexts, including during episodes of political instability (like the study setting in Kenya), other innovations that instead directly dis-incentivize destructive behavior or insulate individuals from high heat exposure, for instance, through air conditioning (Barreca et al., 2016), may also be necessary.

## Bibliography

- Abbink, K. and B. Herrmann (2011). The moral costs of nastiness. *Economic Inquiry* 49 (2), 631–633.
- Abbink, K. and A. Sadrieh (2009). The pleasure of being nasty. *Economics letters* 105(3), 306–308.
- Almås, I., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2010). Fairness and the development of inequality acceptance. *Science* 328, 1176–1178.
- Ambrus, A. and B. Greiner (2012). Imperfect public monitoring with costly punishment: An experimental study. *American Economic Review* 102(7), 3317–32.
- Anderson, C. A. (1989). Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological bulletin* 106(1), 74.
- Anderson, C. A., K. B. Anderson, N. Dorr, K. M. DeNeve, and M. Flanagan (2000). Temperature and aggression. *Advances in Experimental Social Psychology* 32, 63–133.
- 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.
- Andreoni, J. (2006). Philanthropy. In S.-C. Kolm and J. Mercier Ythier (Eds.), *Handbook of the economics of giving, altruism, and reciprocity*, Volume 2, Chapter 18, pp. 1201–1269. North-Holland.
- Andreoni, J., M. A. Kuhn, and C. Sprenger (2015). Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior & Organization* 116, 451–464.
- Andrews, I. and M. Kasy (2019). Identification of and correction for publication bias. *American Economic Review* 109(8), 2766–94.
- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Baron, R. A. (1972). Aggression as a function of ambient temperature and prior anger arousal. *Journal of Personality and Social Psychology* 21, 183–189.
- Baron, R. M. and D. A. Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology* 51(6), 1173.
- Barreca, A., K. Clay, O. Deschenes, M. Greenstone, and J. S. Shapiro (2016). Adapting to climate change: the remarkable decline in the us temperature-mortality relationship over the twentieth century. *Journal of Political Economy* 124(1), 105–159.
- Barreca, A., O. Deschenes, and M. Guldi (2018). Maybe next month? temperature shocks and dynamic adjustments in birth rates. *Demography* 55, 1269–1293.

- Bauer, M., J. Cahlíková, D. Celik Katreniak, J. Chytilová, L. Cingl, and T. Želinský (2018). Anti-social behavior in groups. Technical Report 11944, IZA Institute of Labor Economics.
- Baylis, P., N. Obradovich, Y. Kryvasheyev, H. Chen, L. Coviello, E. Moro, M. Cebrian, and J. H. Fowler (2018). Weather impacts expressed sentiment. *PloS one* 13(4), e0195750.
- Baysan, C., M. Burke, F. Gonzalez, S. Hsiang, and E. Miguel (2019). Non-economic factors in violence: Evidence from organized crime, suicides and climate in mexico. *Journal of Economic Behavior & Organization* 168, 434–452.
- Bell, P. A. and R. A. Baron (1976). Aggression and heat: The mediating role of negative affect. *Journal of Applied Social Psychology* 6, 18–30.
- Benjamin, D. et al. (2018). Redefine statistical significance. *Nature Human Behaviour* 2, 6–10.
- Berkowitz, L. (1989). Frustration-aggression hypothesis: Examination and reformulation. *Psychological Bulletin* 106(1), 59–73.
- Bohra-Mishra, P., M. Oppenheimer, and S. M. Hsiang (2014). Nonlinear permanent migration response to climatic variations but minimal response to disasters. *Proceedings of the National Academy of Sciences* 111(27), 9780–9785.
- Burke, M., E. Gong, and K. Jones (2015). Income shocks and hiv in africa. *The Economic Journal* 125(585), 1157–1189.
- Burke, M., S. M. Hsiang, and E. Miguel (2015). Global non-linear effect of temperature on economic production. *Nature* 527, 235–239.
- Camerer, C. F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B. A. Nosek, T. Pfeiffer, et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour* 2(9), 637.
- Carleton, T. A. (2017). Crop-damaging temperatures increase suicide rates in india. *Proceedings of the National Academy of Sciences*, 201701354.
- Carleton, T. A. and S. M. Hsiang (2016). Social and economic impacts of climate. *Science* 353(6304), aad9837.
- Charness, G., D. Masclet, and M. C. Villeval (2013). The dark side of competition for status. *Management Science* 60(1), 38–55.
- Christensen, G., J. Freese, and E. Miguel (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science* (1 ed.). University of California Press.
- Christensen, G. and E. Miguel (2018). Transparency, reproducibility, and the credibility of economics research. *Journal of Economic Literature* 56(3)(3), 920–980.
- Davis, L. W. and P. J. Gertler (2015). Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences* 112(19), 5962–5967.

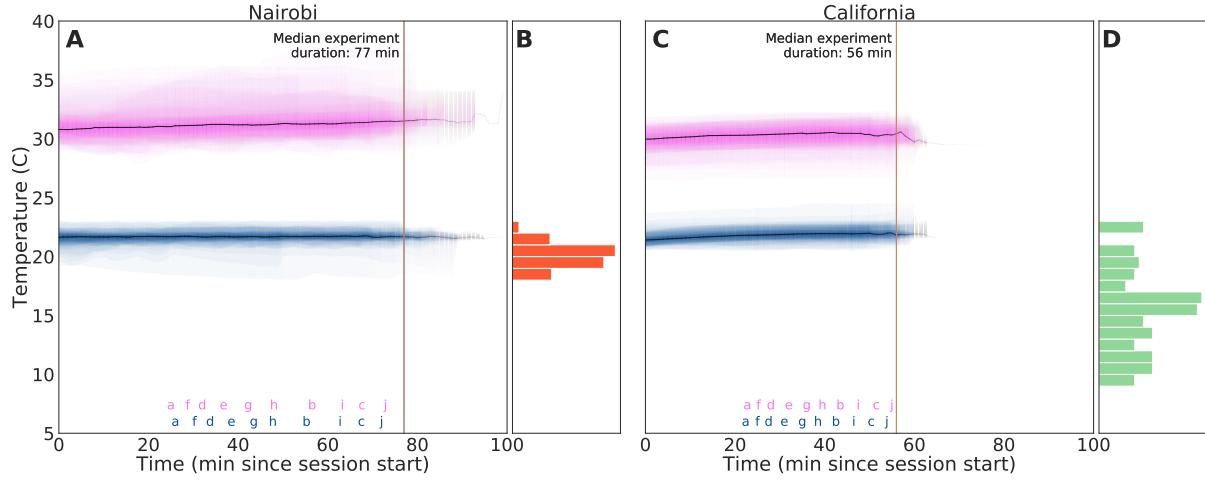
- de Secondat, C. B. D. M. (1748). *The Spirit of Laws*. Hafner Publishers.
- Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature* 52(3), 740–98.
- Dreber, A., D. G. Rand, D. Fudenberg, and M. A. Nowak (2008). Winners dont punish. *Nature* 452(7185), 348.
- Druckman, J. N. and C. D. Kam (2011). Students as experimental participants. In J. N. Druckman, D. P. Green, J. H. Kuklinski, and A. Lupia (Eds.), *Cambridge handbook of experimental political science*, Volume 1, Chapter 4, pp. 41–57. Cambridge UP.
- Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization* 68(1), 1–17.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *The Quarterly Journal of Economics* 133(4), 1645–1692.
- Feng, S., A. B. Krueger, and M. Oppenheimer (2010). Linkages among climate change, crop yields and mexico–us cross-border migration. *Proceedings of the National Academy of Sciences* 107(32), 14257–14262.
- Field, C. B., V. Barros, T. F. Stocker, Q. Dahe, D. J. Dokken, K. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, et al. (2012). Special report on managing the risks of extreme events and disasters to advance climate change adaptation (srex).
- Fischbacher, U., S. Gächter, and E. Fehr (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics letters* 71(3), 397–404.
- Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton (1994). Fairness in simple bargaining experiments. *Games and Economic behavior* 6(3), 347–369.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives* 19(4), 25–42.
- Gächter, S. and B. Herrmann (2008). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1518), 791–806.
- Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review* 102(1), 469–503.
- Graff Zivin, J., S. M. Hsiang, and M. Neidell (2018). Temperature and human capital in the short and long run. *Journal of the Association of Environmental and Resource Economists* 5(1), 77–105.
- Graff Zivin, J. and M. Neidell (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics* 32(1), 1–26.
- Haushofer, J. and E. Fehr (2014). On the psychology of poverty. *Science* 344, 862–867.

- Hsiang, Solomon, M., M. Burke, and E. Miguel (2013). Quantifying the influence of climate on human conflict. *Science* 341(1235367).
- Huntington, E. (1922). *Civilization and climate*. Yale University Press.
- Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological Methods* 15(4), 309–334.
- Jacob, B., L. Lefgren, and E. Moretti (2007). The dynamics of criminal behavior. *The Journal of Human Resources* 42(3), 489–527.
- Johnson, N. D. and A. A. Mislin (2011). Trust games: A meta-analysis. *Journal of Economic Psychology* 32(5), 865–889.
- Jones, B. F. and B. A. Olken (2010). Climate shocks and exports. *American Economic Review* 100(2), 454–59.
- Kahneman, D., J. L. Knetsch, and R. H. Thaler (1986). Fairness and the assumptions of economics. *Journal of business*, S285–S300.
- Kenrick, D. T. and S. W. MacFarlane (1986). Ambient temperature and horn honking: A field study of the heat/aggression relationship. *Environment and behavior* 18(2), 179–191.
- Kiruga, M. (2018, January). The ethnic rivalry that holds kenya hostage. <http://www.theafricareport.com/East-Horn-Africa/the-ethnic-rivalry-that-holds-kenya-hostage.html>.
- Larrick, R. P., T. A. Timmerman, A. M. Carton, and J. Abrevaya (2011). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychological Science* 22(4), 423–428.
- Lee, H. D. P. (1962). *Aristotle: Meteorologica*. Harvard University Press.
- Mackworth, N. H. (1946). Effects of heat on wireless telegraphy operators hearing and recording morse messages. *British Journal of Industrial Medicine* 3(3), 143–158.
- Mebane, Jr., W. R. (2017). Anomalies and frauds(?) in the kenya 2017 presidential election. Working paper.
- Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, et al. (2014). Promoting transparency in social science research. *Science* 343(6166), 30–31.
- Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy* 112(4), 725–753.
- Missirian, A. and W. Schlenker (2017). Asylum applications respond to temperature fluctuations. *Science* 358(6370), 1610–1614.
- Obradovich, N. and J. H. Fowler (2017). Climate change may alter human physical activity patterns. *Nature Human Behaviour* 1(5), 0097.

- Obradovich, N., D. Tingley, and I. Rahwan (2018). Effects of environmental stressors on daily governance. *Proceedings of the National Academy of Sciences*, 201803765.
- Palamarek, D. L. and B. G. Rule (1979). The effects of ambient temperature and insult on the motivation to retaliate or escape. *Motivation and Emotion* 3, 83–92.
- Park, R. J. (2020). Hot temperature and high stakes performance. *Journal of Human Resources*.
- Penrose, L. S. and J. Raven (1936). A new series of perceptual tests: Preliminary communication. *Psychology and Psychotherapy: Theory, Research and Practice* 16(2), 97–104.
- Prediger, S., B. Volland, and B. Herrmann (2014). Resource scarcity and antisocial behavior. *Journal of Public Economics* 119, 1–9.
- Russell, J. A. and L. F. Barrett (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology* 76, 805–819.
- Schlenker, W. and M. J. Roberts (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences* 106(37), 15594–15598.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of personality and social psychology* 51(3), 515–530.
- The Economist (2017, October). Kenya's fresh election is preposterously flawed. <https://www.economist.com/middle-east-and-africa/2017/10/28/kenyas-fresh-election-is-preposterously-flawed>.
- von Uexküll, N., M. Croicu, H. Fjelde, and H. Buhaug (2016). Civil conflict sensitivity to growing-season drought. *Proceedings of the National Academy of Sciences* 113(44), 12391–12396.
- Wargocki, P., O. Seppanen, J. Andersson, A. Federation of European Heating and Air-conditioning Associations., D. Forssan Kirjapaino, K. Fitzner, and S. O. Hanssen (2006). *Indoor climate and productivity in offices how to integrate productivity in life-cycle cost analysis of building services*. Rehva.
- Wenz, L., A. Levermann, and M. Auffhammer (2017). North–south polarization of european electricity consumption under future warming. *Proceedings of the National Academy of Sciences*, 201704339.
- Wu, J.-J., B.-Y. Zhang, Z.-X. Zhou, Q.-Q. He, X.-D. Zheng, R. Cressman, and Y. Tao (2009). Costly punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences* 106(41), 17448–17451.
- Ye, X., H. Chen, and Z. Lian (2010). Thermal environment and productivity in the factory. *ASHRAE Transactions* 2010 116, 590–599.
- Zillmann, D. (2006). Transfer of excitation in emotional behavior. In J. T. Cacioppo and R. E. Petty (Eds.), *Social psychophysiology: A sourcebook*, pp. 215–240. New York: Guilford Press.

# Figures and Tables

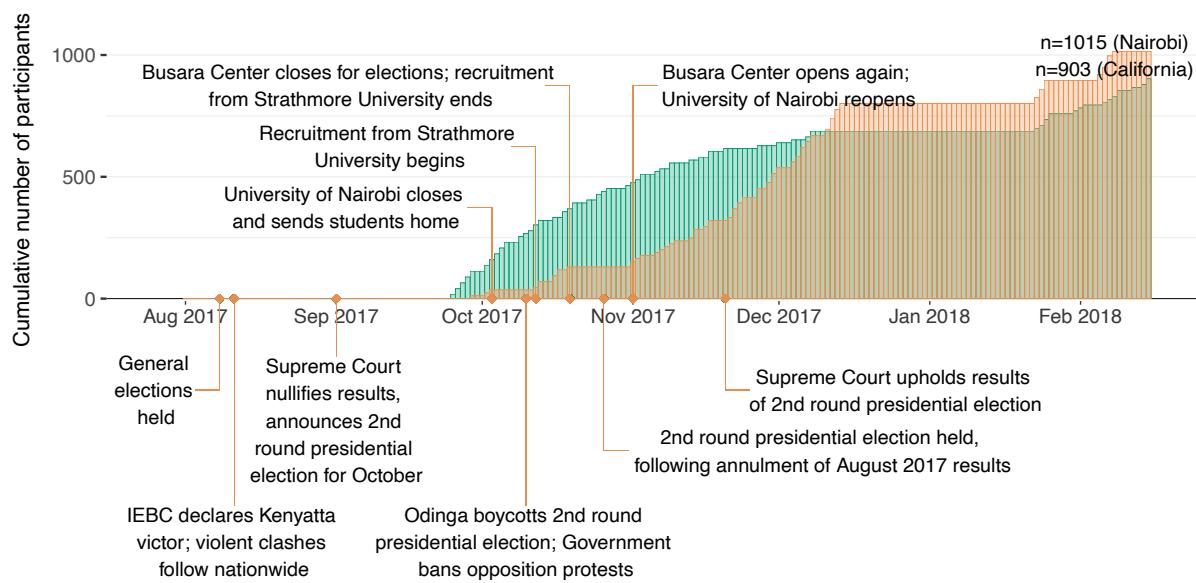
Figure 1: Experimental temperatures, timeline, and description



Group	Label	Measurement	Description
Economic Preferences	d	Risk-taking (Eckel and Grossman, 2008)	Participant-chosen variance in outcome from a coin toss, where choosing a higher variance outcome is risk-loving
	d	Rational choice violation I (Eckel and Grossman, 2008)	When coin toss A is preferred to coin toss B, and coin toss B is preferred to coin toss C, but coin toss A is not preferred to coin toss C
	e	Patience (Andreoni et al., 2015)	How one weighs two future periods (time discounting)
	e	Time inconsistency (Andreoni et al., 2015)	How one weighs the present versus the future (present bias)
Social Behavior	f	Fairness (Almås et al., 2010)	Share of tokens earned from precision task given freely to partner
	g	Trust (Johnson and Mislin, 2011)	Share of endowed tokens entrusted to partner, as a measure of trust
	h	Public contribution (Fischbacher et al., 2001)	Share of endowed tokens placed into the group fund, as a measure of willingness to contribute to a public good
	i	Destruction (Abbrink and Sadrieh, 2009)	Share of partner's vouchers or gift cards voluntarily and anonymously destroyed, as a measure of ill-will
	j	Charitable donation (Andreoni, 2006)	Amount of tokens donated to charity, as a measure of charitableness
Affect	c	Happiness (Russell and Barrett, 1999)	Self-expressed degree of happiness at the end of the experiment
	c	Alertness (Russell and Barrett, 1999)	Self-expressed degree of alertness at the end of the experiment
Cognition	a	Precision (Gill and Prowse, 2012)	Productivity, measured through moving sliders
	b	Fluid intelligence (Penrose and Raven, 1936)	Reasoning ability, measured through Raven's matrices
	c	Cognitive reflection (Frederick, 2005)	Ability to override gut response, measured through five questions

Note: Panels A and C display the median indoor temperature trend during experiments (black) and the distribution of temperature measurements across experiments (colored shading). The opacity of the median line is correlated with the number of experiments that had a measurement recorded at a given time, and the opacity of the shading is correlated with the number of experiments that recorded a temperature measurement falling within a given range. The edge of the shading indicates the minimum and maximum temperatures across experiments at each point in time. The mean temperature experienced by Hot and Control groups was 31.3°C ( $\sigma = 0.3^\circ\text{C}$ ,  $n = 77$ ) and 21.7°C (0.1°C, 79) in Nairobi, respectively, and 30.2°C (0.2°C, 76) and 21.8°C (0.2°C, 76) in California. The midpoint of the median start and end times of participation in each experiment module are indicated at the bottom of Panels A and C for treatment (pink) and control (blue) groups, with labels corresponding to those in Panel E. The vertical green line shows the median experiment length in each location. Panels B and D contain session-weighted histograms of daily mean outdoor temperatures on days in which we ran sessions. On average, the mean daily outdoor temperature was 20.0°C ( $\sigma = 0.9^\circ\text{C}$ ) in Nairobi and 15.7°C (3.7°C) in California. Panel E contains a table of the 14 measurements (outcomes) and brief descriptions of each. These measurements consist of the 12 primary outcomes from the pre-analysis plan plus happiness and alertness.

Figure 2: Experimental timelines for Nairobi (dark orange) and California (light green)



Note: The colored bars display the cumulative number of experiment participants. Events directly relating to the Nairobi site are included above the timeline, while events relating to the Nairobi political context are included below the timeline (see Experimental Protocol, Design, and Timeline Section J and Appendix Section A.1).

Table 1: Experimental results

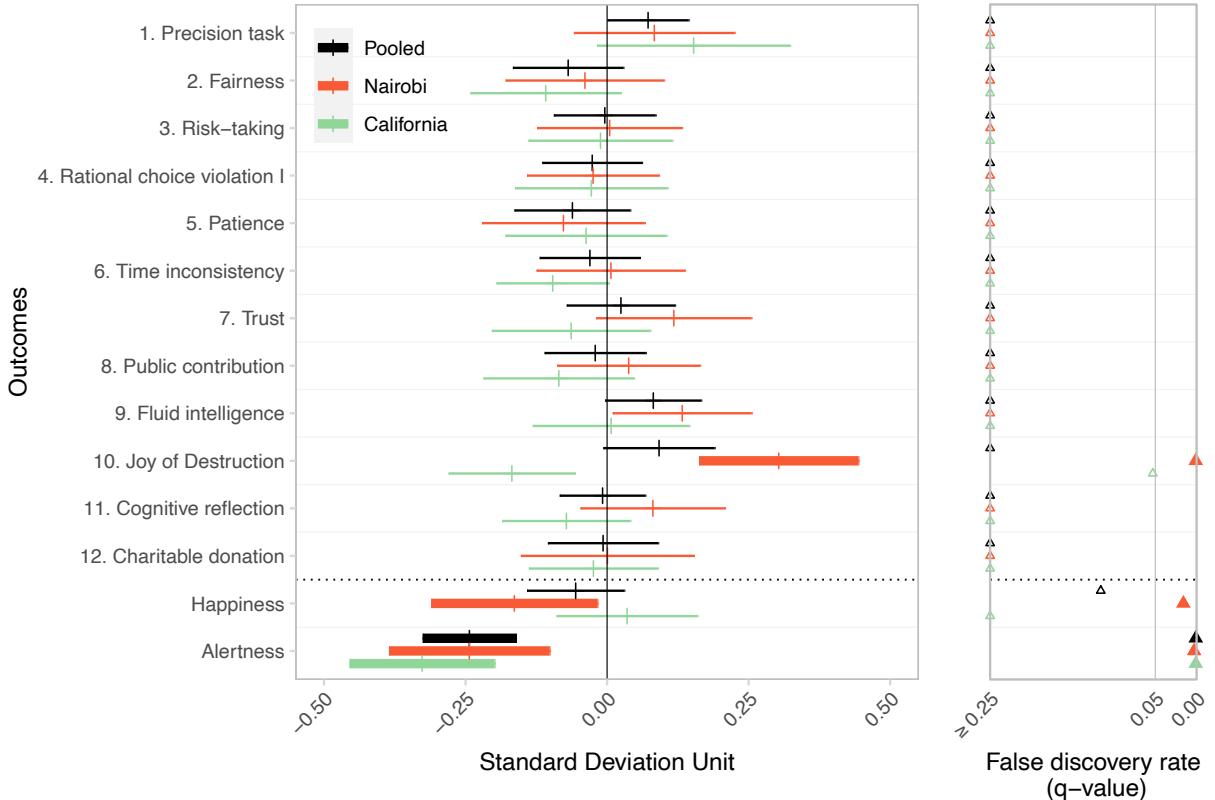
Module		Treatment effect	Standard error	Standardized treatment effect	p-value	q-value	Control Mean (SD)
1. Precision task	Pooled	0.516	0.261	0.072	0.049	0.384	17.9 (7.1)
	Nairobi	0.451	0.391	0.083	0.253	0.722	13.0 (5.4)
	California	0.590	0.336	0.153	0.083	0.457	23.6 (3.9)
2. Fairness		-0.017	0.012	-0.069	0.170	0.618	0.32 (0.24)
		-0.010	0.018	-0.039	0.586	1.000	0.33 (0.26)
		-0.025	0.015	-0.108	0.114	0.457	0.31 (0.23)
3. Risk-taking		-1.835	20.643	-0.004	0.929	1.000	366.3 (449.8)
		1.926	27.568	0.005	0.944	1.000	329.1 (421.5)
		-5.553	30.951	-0.012	0.858	0.850	409.0 (477.2)
4. Rational choice violation I		-0.004	0.007	-0.026	0.561	1.000	0.03 (0.16)
		-0.004	0.010	-0.025	0.678	1.000	0.03 (0.16)
		-0.005	0.011	-0.028	0.687	0.768	0.03 (0.16)
5. Patience		-0.000058	0.000316	-0.061	0.854	1.000	0.994 (0.006)
		0.000961	0.000888	-0.077	0.279	0.722	0.994 (0.013)
		-0.000609	0.000362	-0.037	0.093	0.457	0.994 (0.005)
6. Time inconsistency		0.000480	0.009721	-0.030	0.961	1.000	1.001 (0.202)
		0.0179	0.02575	0.007	0.487	1.000	1.053 (0.400)
		-0.008782	0.012689	-0.096	0.489	0.768	0.989 (0.175)
7. Trust		0.008	0.016	0.024	0.617	1.000	0.42 (0.32)
		0.033	0.020	0.118	0.096	0.471	0.35 (0.28)
		-0.022	0.024	-0.063	0.377	0.759	0.50 (0.34)
8. Public contribution		-8.923	19.444	-0.021	0.647	1.000	529.7 (426.2)
		15.316	25.846	0.038	0.555	1.000	442.5 (401.1)

	-36.881	29.359	-0.085	0.213	0.520	629.6 (432.6)
9. Fluid intelligence	0.015 <b>0.027</b> 0.001	0.008 <b>0.013</b> 0.008	0.082 <b>0.133</b> 0.007	0.061 <b>0.037</b> 0.919	0.384 <b>0.256</b> 0.850	0.86 (0.18) <b>0.80 (0.21)</b> 0.94 (0.12)
10. Joy of Destruction	0.023 <b>0.079</b> -0.041	0.013 <b>0.019</b> 0.014	0.092 <b>0.303</b> -0.168	0.069 <b>&lt;0.001</b> 0.004	0.384 <b>0.001**</b> 0.053	0.11 (0.25) <b>0.14 (0.26)</b> 0.08 (0.24)
11. Cognitive reflection	-0.002 <b>0.016</b> -0.023	0.011 <b>0.013</b> 0.018	-0.008 <b>0.081</b> -0.072	0.837 <b>0.218</b> 0.217	1.000 <b>0.722</b> 0.520	0.33 (0.29) <b>0.21 (0.20)</b> 0.46 (0.32)
12. Charitable donation	-4.443 <b>0.412</b> -12.673	31.228 <b>52.675</b> 30.512	-0.007 <b>0.001</b> -0.024	0.887 <b>0.994</b> 0.679	1.000 <b>1.000</b> 0.768	410.4 (628.0) <b>561.8 (675.3)</b> 245.0 (524.9)
Happiness	-0.079 <b>-0.190</b> 0.049	0.062 <b>0.087</b> 0.088	-0.055 <b>-0.164</b> 0.035	0.208 <b>0.031</b> 0.579	0.116 <b>0.016*</b> 0.408	5.31 (1.42) <b>5.91 (1.16)</b> 4.62 (1.39)
Alertness	<b>-0.420</b> <b>-0.322</b> <b>-0.530</b>	<b>0.073</b> <b>0.096</b> <b>0.106</b>	<b>-0.243</b> <b>-0.243</b> <b>-0.327</b>	<b>&lt;0.001</b> <b>0.001</b> <b>&lt;0.001</b>	<b>0.001**</b> <b>0.003**</b> <b>0.001**</b>	<b>5.25 (1.42)</b> <b>6.09 (1.32)</b> <b>4.28 (1.62)</b>

↑

Note: Treatment effects, as well as associated standard errors, standardized treatment effects, p-values, q-values, control means, and control standard deviations from the main specification in the pooled sample (black;  $n = 1,878$  participants), as well as from the Nairobi (dark orange; 1,004) and California (light green; 874) specifications. These tests were pre-specified in the pre-analysis plan. Multiple hypothesis testing adjustments are performed on the set of per comparison  $p$ -values associated with the treatment across pre-specified primary outcomes, for each specification. Multiple hypothesis testing adjustments are also performed on the set of  $p$ -values associated with the treatment for Happiness and Alertness (not pre-specified), for each specification. Rows are bolded in cases where  $q < 0.05$ . To standardize outcomes, sample control means are subtracted from outcome values, and the result is divided by the sample control standard deviation. The treatment effect for Time Inconsistency is flipped so that a positive treatment effect means more time inconsistency. Individual-level estimates are used to derive the standardized treatment effect values for Patience and Time Inconsistency, while pre-specified aggregate estimates are used to derive the treatment effects and other statistics that feature in the core results. Individuals who responded solely with corner cases were dropped (8% of the sample) as well as those who exhibited GARP violations (4% of the sample) for the calculations of aggregate beta and delta (the outcomes for Patience and Time inconsistency, respectively). \* $p < .05$ , \*\* $p < .01$ .

Figure 3: Experimental results



Note: Treatment effects from the main specification in the pooled sample (black;  $N = 1,878$  participants), as well as from the Nairobi (dark orange; 1,004) and California (light green; 874) specifications (see Experimental Protocol, Design, and Timeline Section Section J for details) in the left panel. These tests were pre-specified in the pre-analysis plan. Treatment effects are presented in terms of standard deviation units for the 12 pre-specified primary outcomes of interest and for the survey questions on participant Happiness and Alertness. For comparability with regression results, individuals who did not respond “Female” or “Male” to the gender survey question were dropped (2% of the sample). Thin standard error bars correspond to the 95% confidence interval around the treatment effect off of the regression for the relevant specification, while thick standard error bars refer to those instances where the multiple testing adjusted False Discovery Rate (FDR) q-value significance level is less than 0.05. Multiple hypothesis testing adjustments are performed on the set of per comparison p-values associated with the treatment across pre-specified primary outcomes, for each specification. Multiple hypothesis testing adjustments are also performed on the set of p-values associated with the treatment for Happiness and Alertness (not pre-specified), for each specification. The right panel records the q-value associated with treatment across each outcome, by specification, where open triangles record  $q \geq 0.05$  and closed triangles record  $q < 0.05$ . The treatment effect for Time Inconsistency is flipped so that a positive treatment effect means more time inconsistency. To graphically display standardized treatment effects for Patience and Time Inconsistency, individual level estimates were developed via non-linear least squares, following Andreoni et al. (2015). Individuals who responded solely with corner cases were dropped (8% of the sample) as well as those who exhibited GARP violations (4% of the sample) for the former calculations. Both individual level estimates for Patience and Time Inconsistency were bottom and top coded to be between 0 and 1 (inclusive) before standardization.

Figure 4: Mechanisms and interactions of temperature's effect on destruction

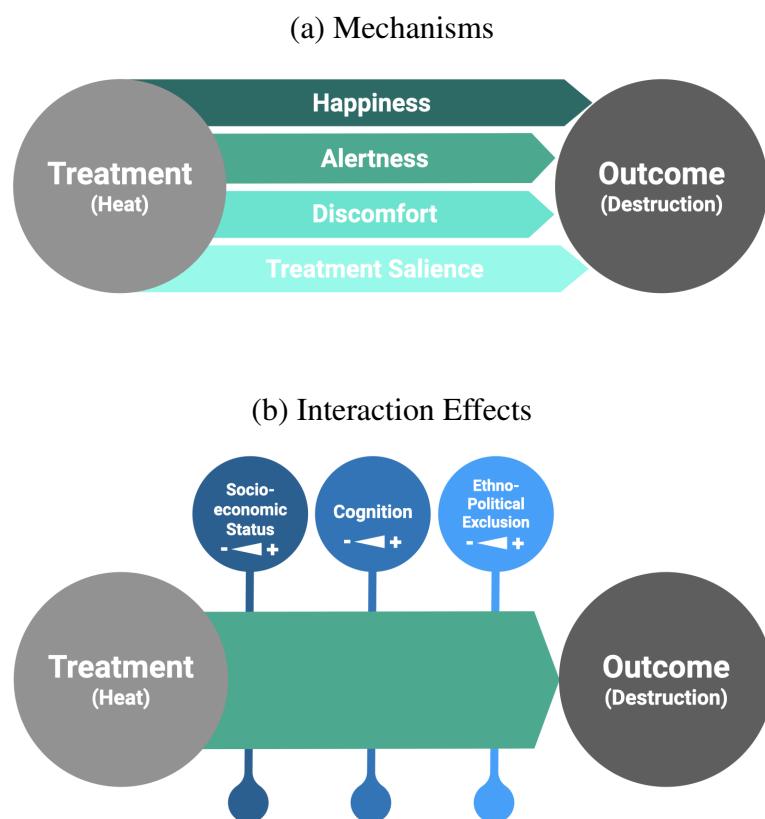
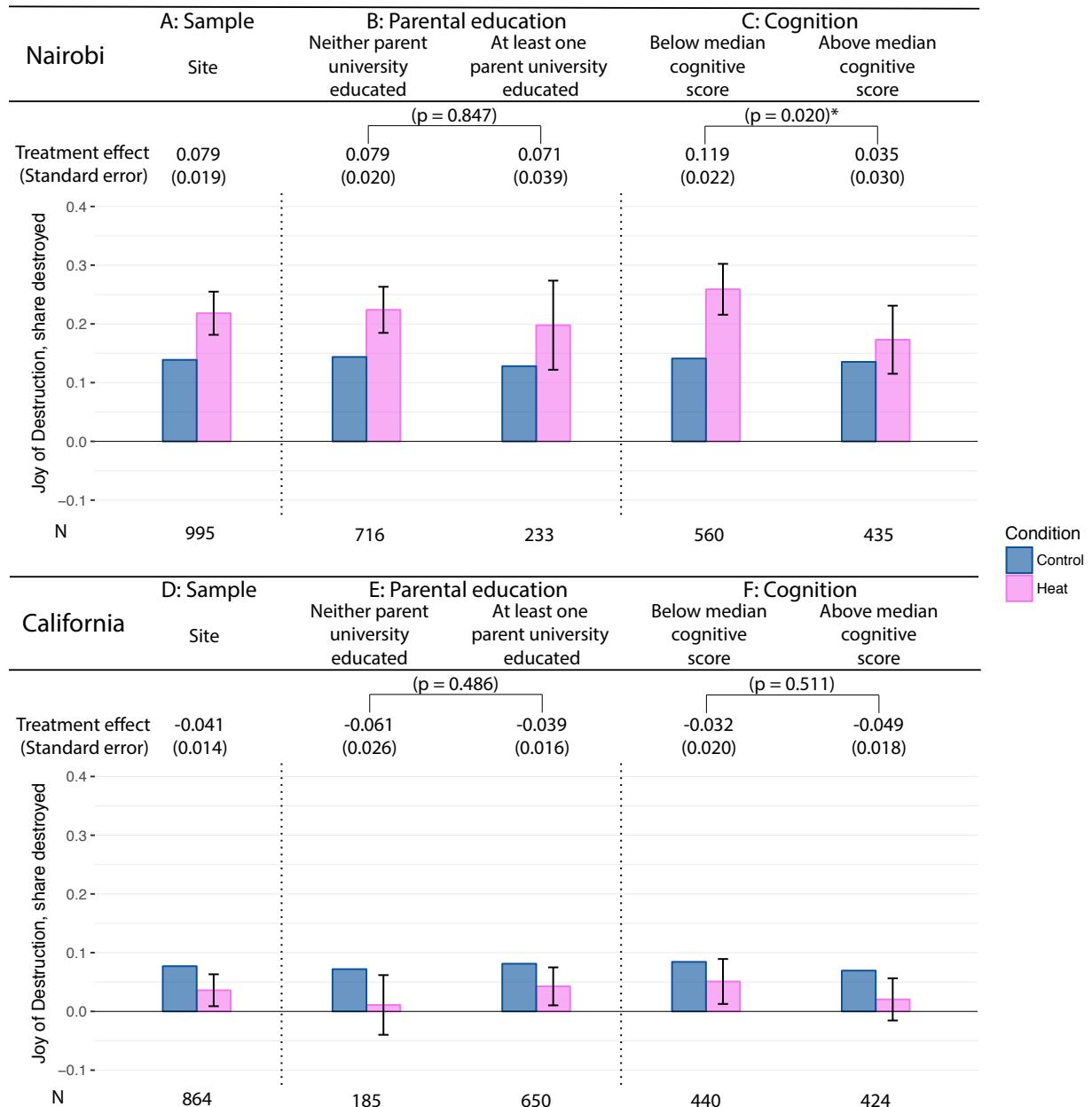
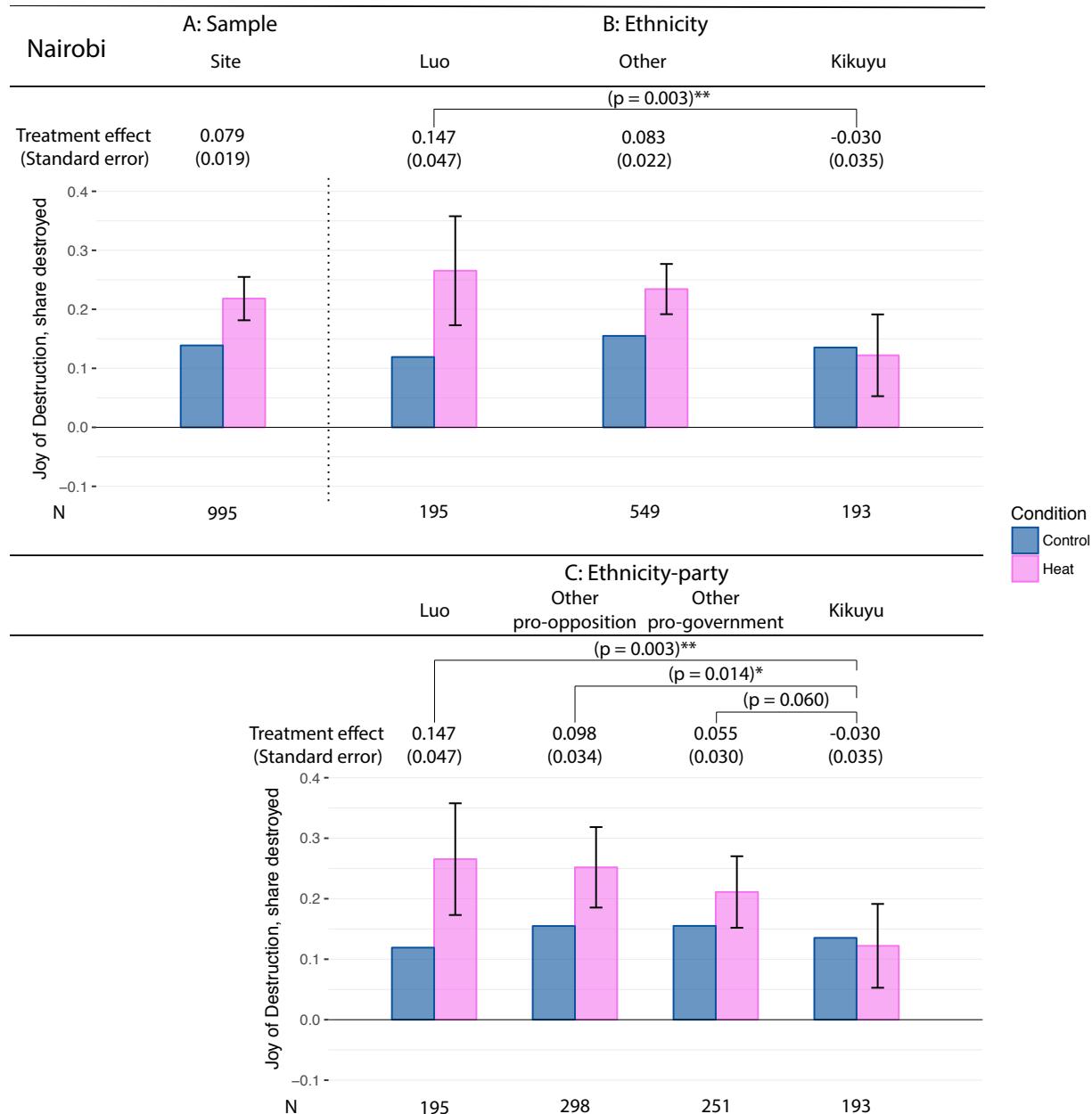


Figure 5: Treatment effects from the Joy of Destruction module in Nairobi and California



Note: Treatment effects and standard errors are presented above the bars. Blue bars represent the average share of vouchers destroyed by control individuals, while pink bars represent the average share of vouchers destroyed by treated individuals, in both the Nairobi sample (Panels A, B, and C) and the California sample (Panels D, E, and F). These sub-group tests were not pre-specified in the pre-analysis plan. Panels A and D provide site-specific average shares of destruction. Panels B and E provide site-specific average shares of destruction by parental university education status (an indicator variable set to one if the participant has at least one university educated parent). Panels C and F provide site-specific average shares of destruction by median “cognition score” status (where “cognition score” is developed by normalizing the sum of individuals’ normalized results from the fluid intelligence task and cognitive reflection test). The standard error bars on the pink bars correspond to the 95% confidence interval around the treatment effect off of the main specification regression for the specific subsample. The overarching lines capture the p-value from the interaction effect from the within-panel regression between the specific group and treatment, where the leftmost group is the reference group. At the bottom of each panel is the number of individuals in that subsample. For comparability with regression results, individuals who did not respond “Female” or “Male” to the gender survey question were dropped (2% of the sample). \* $p < .05$ , \*\* $p < .01$ .

Figure 6: Treatment effects from the Joy of Destruction module in Nairobi across ethnicities and ethno-political groups



Note: Treatment effects and standard errors are presented above the bars. Blue bars represent the average share of vouchers destroyed by control individuals, while pink bars represent the average share of vouchers destroyed by treated individuals. These sub-group tests were not pre-specified in the pre-analysis plan. Panel A provides average shares of destruction in Nairobi. Panel B provides average shares of destruction by self-reported ethnicity in Nairobi. Panel C splits other ethnicities into “other pro-opposition” or “other pro-government.” “Other pro-opposition” refers to those individuals who self-identified with an ethnicity aligned with ethnic Luos, the ethnicity in opposition in the 2017 election, while “other pro-government” refers to those individuals who self-identified with an ethnicity aligned with ethnic Kikuyus, the ethnicity in power in the 2017 election. The standard error bars on the pink bars correspond to the 95% confidence interval around the treatment effect off of the main specification regression for the specific subsample. The overarching lines capture the p-value from the interaction effect (from the within-panel regression) between the specific group and treatment, where the leftmost group is the reference group. At the bottom of each panel is the number of individuals in that subsample. For comparability with regression results, individuals who did not respond “Female” or “Male” to the gender survey question were dropped (2% of the sample). \* $p < .05$ , \*\* $p < .01$ .

# Appendix: Destructive Behavior, Judgment, and Economic Decision-making under Thermal Stress

Ingvild Almås, Maximilian Auffhammer, Tessa Bold,  
Ian Bolliger, Aluma Dembo, Solomon Hsiang,  
Shuhei Kitamura, Edward Miguel, and Robert Pickmans

Additional materials to the article “Destructive Behavior, Judgment, and Economic Decision-making under Thermal Stress” are publicly available on the Open Science Framework (<https://osf.io/9v3fw/>), and include the following:

- The Appendix (contained within);
- The Supplemental Results Appendix and PAP Report (“Supplement”);
- The Experimental Protocol, Design, and Timeline (“EPDT”); and
- The Pre-Analysis Plan, Pre-Analysis Plan Addendum, and Pre-Analysis Plan Amendment (also available on the AEA RCT Registry at <https://www.socialscienceregistry.org/trials/1361> under RCT ID AEARCTR-0001361).

# Contents

<b>Appendices</b>	<b>4</b>
<b>A Nairobi setting</b>	<b>4</b>
A.1 Political context . . . . .	4
A.2 Outside weather . . . . .	5
<b>B California setting</b>	<b>8</b>
B.1 Political context . . . . .	8
B.2 Outside weather . . . . .	8
<b>C Descriptive information</b>	<b>9</b>
<b>D Main results</b>	<b>18</b>
D.1 Main results with primary outcomes . . . . .	18
D.2 Main results with additional outcomes . . . . .	30
<b>E Exploring the effect of heat on destruction</b>	<b>42</b>

## A Nairobi setting

### A.1 Political context

The Kenyan general elections held on August 8th, 2017 to elect the President, as well as members of Parliament and lower levels of government, represent the latest chapter in the longstanding political competition between ethnic Kikuyu and ethnic Luos in Kenya going back to independence in the 1960s (Barkan, 1994). Major political parties and coalitions aligned strongly along ethnic lines, with the governing party aligned with ethnic Kikuyus backing the incumbent Uhuru Kenyatta and the opposition aligned with ethnic Luos backing the opposition leader Raila Odinga. An analysis using a nationally representative sample found that 96 percent of ethnic Kikuyu voters supported Kenyatta, while 99 percent of ethnic Luo voters supported Odinga (Choi et al., 2017).

Previous presidential elections in Kenya had also been marred by allegations of vote rigging. Ethnic violence erupted over the presidential contest in 2007 between the incumbent (and eventual victor) Mwai Kibaki (an ethnic Kikuyu) and Odinga (an ethnic Luo), in which more than 1,200 people were killed. In contrast, the presidential election in 2013 between Kenyatta (an ethnic Kikuyu) and Odinga was more peaceful, although there were still widespread concerns about the integrity of the election (Jorgic and Malalo, 2013). Similar tensions existed around the August 2017 election: On July 27th, 2017, Chris Msando, the head of information, communication and technology at the Independent Electoral and Boundaries Commission (IEBC) was found dead, with marks of murder and torture on his body (Burke, 2017b). Two days later, a man with a machete invaded Deputy President William Ruto's home in unclear circumstances (BBC News, 2017a).

On August 11th, the IEBC announced Kenyatta as the president-elect by a margin of 1.4 million votes (54.4% vs. 44.8%), despite pre-election prediction polls suggesting a far closer election (Awich, 2017). Opposition protestors in Nairobi, Migori, Coast, Homabay, and elsewhere were met with excessive police force, including shootings and beatings. At least 12 people were killed by police in Odinga stronghold Kisumu county and in neighboring Siaya county, and over 100 were badly injured on the night that Kenyatta was declared the winner. The following day, the Kenya National Commission on Human Rights reported that the police had killed at least 24 people nationwide, including one in Kisumu and 17 in Nairobi (Human Rights Watch, 2017).

Evidence from independent experts suggests widespread voter fraud in the recent election (Mebane, 2017). Odinga himself alleged that hackers infiltrated the commission's servers to manipulate the electoral results. He filed a petition with the Supreme Court on August 18th (Al Jazeera, 2017). The Supreme Court opened hearings on Odinga's claims for the first time on August 28th, allowing representatives from both Odinga's and Kenyatta's party to audit the IEBC's electoral voting system (Okuoro, 2017). The IEBC also submitted all result forms to the Supreme Court for scrutiny (Muthoni and Ogembra, 2017). On September 1st, after examining the result forms, in which several suspicious elements were found (thousands of missing paper forms, forms lacking IEBC stamps), the Supreme Court nullified the August 8th election and announced that a new presidential election would take place within 60 days (Muthoni and Ogembra, 2017). The IEBC initially set October 17th as the election date, which was later rescheduled to October 26th.

On October 10th, Odinga told his supporters that he would boycott the second round election, given what he perceived as a lack of reform in the IEBC that would address the problems found in the first round. Meanwhile, the government banned opposition protests in the business districts of three cities (Nairobi, Mombasa, and Kisumu) on October 12th (Obulutsa, 2017). IEBC Chairman Wafula Chebukati admitted partisan division in the commission and expressed skepticism about a fair election on October 18th (Onyango, 2017). By October 27th, a second attempt to repeat voting in four counties in western Kenya was called off due to security concerns (Blomfield, 2017).

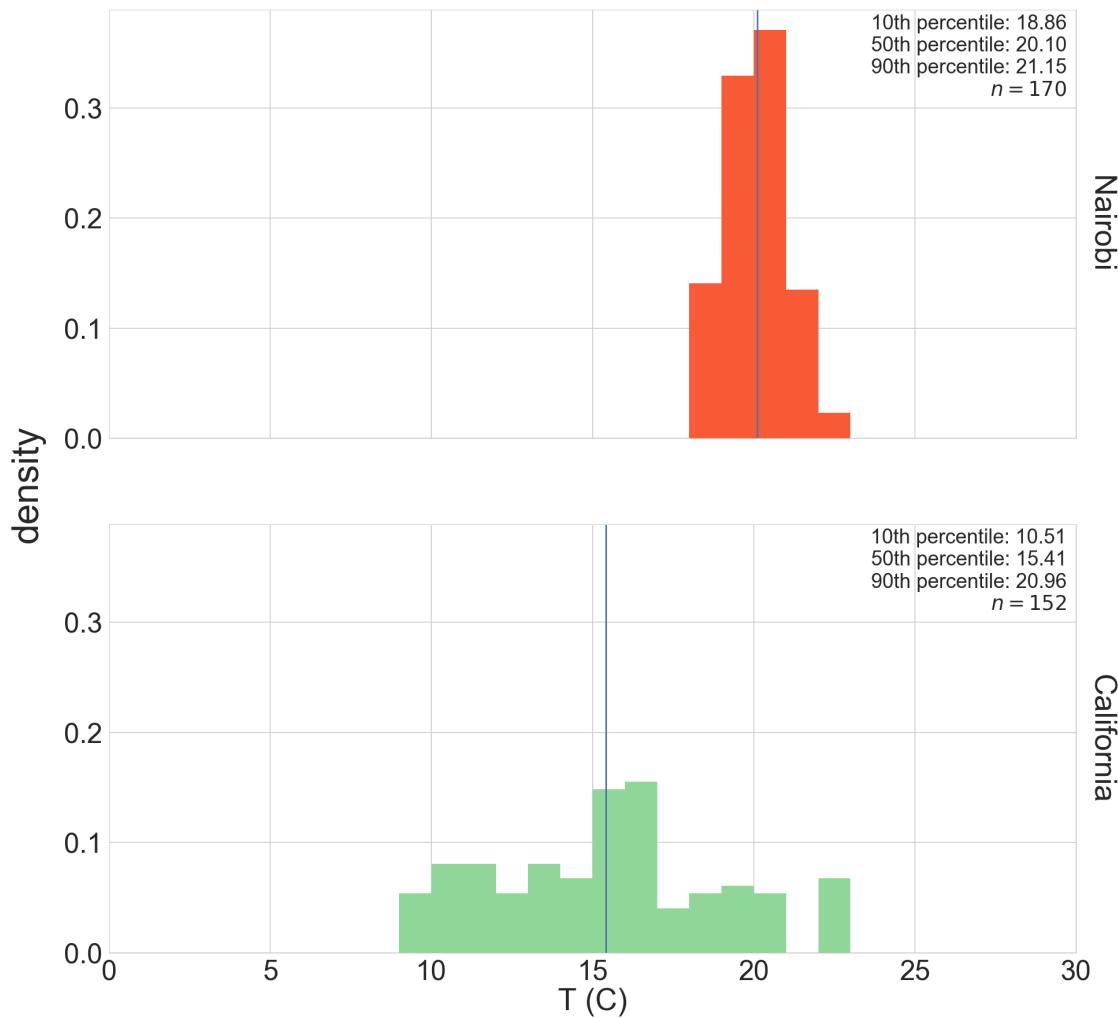
The second round presidential election was held on October 26th, as scheduled, but was widely boycotted in opposition areas. On October 30th the IEBC declared Kenyatta the winner of the election with 98.26% of the vote. However, the voter turnout of 38.84% was much lower than the first round election turnout of nearly 80% (BBC News, 2017b). The Supreme Court upheld Kenyatta's victory on November 20th (Burke, 2017a). Odinga said he considered Kenyatta's government to be illegitimate in a statement issued shortly after the court decision. Political protests continued throughout this period, up to at least February 2018, where on February 2nd, people had protested over the arrest of an opposition lawyer (Africanews, 2018). Given the above events, as well as its recent history of undermined political rights and civil liberties, Freedom House gave Kenya a rating of 48/100 ("partly free") in its 2018 Freedom in the World survey (Freedom House, 2018a).

It was in the midst of this polarized climate that the main experiment in Nairobi took place, starting on September 29th, 2017 and ending on February 8th, 2018. In particular, on October 3rd the University of Nairobi sent students (who were serving as study participants for the experiment) home for several weeks, closing its doors temporarily. The decision to close was due to the student unrest surrounding the elections, with students actively taking part in demonstrations, one of which became violent. See Experimental Protocol, Design, and Timeline Section C for further details on how political events affected recruitment, as well as a timeline (Experimental Protocol, Design, and Timeline Figure C.2) displaying the political events in Kenya during the experimental timeline.

## A.2 Outside weather

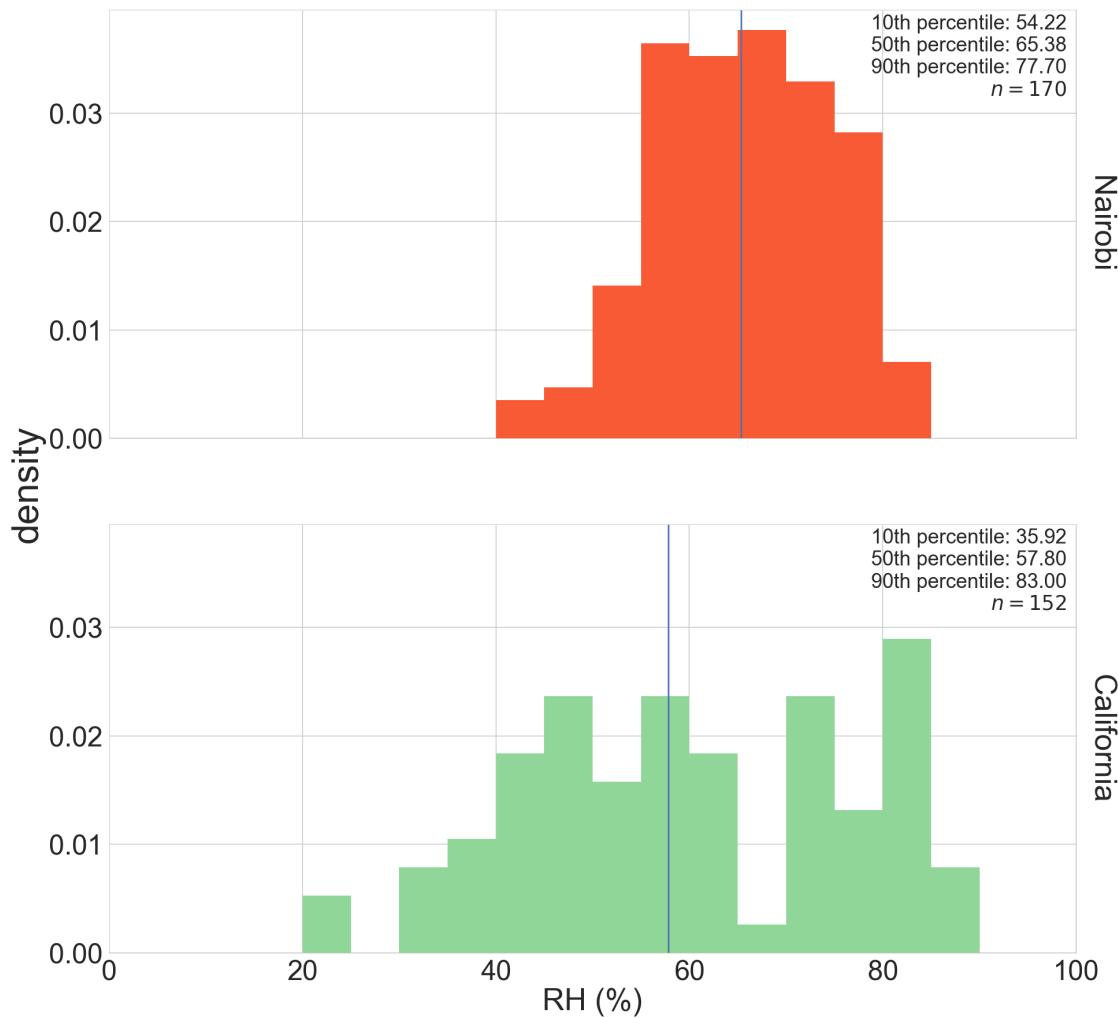
Appendix Figure A.2.1 displays the distributions of daily mean outdoor temperature in Celsius for days on which sessions were run, by site, weighted at the session-condition level (so as to give more weight to those days on which more session-conditions were run). Meanwhile Appendix Figure A.2.2 displays the distribution of daily mean outdoor relative humidity in percentage for days on which sessions were run, by site, weighted at the session-condition level. These means are generated over the 24-hour period of the day that the session was held. Focusing on the dark orange distribution for Nairobi in Appendix Figure A.2.1, daily outside temperature tended to be fairly tight around the median of 20°C. At least 90 percent of recordings in Nairobi fell below the experimental control room target temperature, but this percentage would fall if one were to focus only on temperatures occurring during the day. As for relative humidity (displayed by the dark orange distribution in Appendix Figure A.2.2), the majority of daily mean outside relative humidity recordings within the experimental timeline fall within the 54 to 78 percent range.

Figure A.2.1: Daily mean outdoor temperature (Celsius), by site (session-condition weighted)



Note: The dark orange distribution displays the distribution of daily mean outside temperature recordings (in Celsius) during the course of the experiment in Nairobi, while the light green distribution displays the distribution of daily mean outside temperature recordings (in Celsius) during the course of the experiment in California. These recordings are weighted by session-condition to reflect that some days have more session-conditions run than other days. Sample size values refer to number of session-conditions conducted at each site. Means are generated over the 24 hours of the day that the session was run. These measurements come from weather sensors located outside of the Experimental Social Science Laboratory (Xlab) in Berkeley, California, USA and the Busara Center for Behavioral Economics in Nairobi, Kenya. The solid blue lines indicate the (session-condition weighted) median daily mean outside temperature for each distribution.

Figure A.2.2: Daily mean outdoor relative humidity (%), by site (session-condition weighted)



The dark orange distribution displays the distribution of daily mean outside relative humidity recordings (in percentage) during the course of the experiment in Nairobi, while the light green distribution displays the distribution of daily mean outside relative humidity recordings (in percentage) during the course of the experiment in California. These recordings are weighted by session-condition to reflect that some days have more session-conditions run than other days. Sample size values refer to number of session-conditions conducted at each site. Means are generated over the 24 hours of the day that the session was run. These measurements come from weather sensors located outside of the Experimental Social Science Laboratory (Xlab) in Berkeley, California, USA and the Busara Center for Behavioral Economics in Nairobi, Kenya. The solid blue lines indicates the (session-condition weighted) median daily mean outside relative humidity for each distribution.

## B California setting

### B.1 Political context

In contrast to Kenya, no major national or California elections occurred over the period of time in which the experiment was held in Berkeley. Furthermore, the United States has long been characterized as a stable democracy: while Kenya received a rating of 48/100 (“partly free”) in its 2018 Freedom in the World survey (Freedom House, 2018a), the United States of America received a higher rating of 86/100 (“free”) from Freedom House in its 2018 Freedom in the World survey (Freedom House, 2018b).

This is not to say that there have not been controversies and episodes of political tension during the experimental timeline. One particular event that was also ethnically-charged was the “Unite the Right” rally, held in Charlottesville, Virginia from August 11th to 12th, 2017. White nationalists, neo-Nazis and Ku Klux Klan members gathered, officially to protest the removal of a statue of Confederate General Robert E. Lee, and leading to clashes with counter-protestors that left one person dead and 19 others injured (Heim et al., 2017). In reaction to the Trump administration’s decision to rescind the Deferred Action for Childhood Arrivals (DACA) program that was put in place under the Obama Administration, nationwide protests were held on September 5th, 2017 (Keneally, 2017). On January 20th, 2018, more than a million protestors in cities across the nation participated in the Women’s March, demonstrating against Trump administration policies and advocating for a variety of legislation and policies, including women’s rights, healthcare reform, and the environment (The New York Times, 2018). Ongoing investigations into election interference by foreign entities and concerns over international agreements and relations have also motivated protests and debate in American politics.

Despite the above events, the democratic process in the United States over recent years remains robust, accommodating controversy within agreed-upon procedures. Protests have largely been non-violent and, significantly, have not been targeted against the fairness of democratic institutions.

### B.2 Outside weather

Focusing on the light green distributions for California in Appendix Figures A.2.1 and A.2.2, one can note that daily mean outside temperature varied somewhat over the experimental timeline (particularly in comparison to daily mean outside temperature in Nairobi), but tended to stay within the range of 10 to 20 degrees. Similar to in Nairobi, at least 90 percent of recordings fell below the experimental control room target temperature, but this percentage would fall if one were to focus only on temperatures occurring during the day. As for daily mean outdoor relative humidity, the majority of recordings within the experimental timeline fall between 35 to 83 percent, with the range driven in part by the variability in daily mean outdoor temperature. The distribution for daily mean outdoor relative humidity is wider than that for Nairobi, partly reflecting the fact that the distribution for daily mean outdoor temperature is wider in the California setting than in the Nairobi setting.

## C Descriptive information

Table C.1: Summary statistics

	California			Nairobi		
	Control	Heat	diff	Control	Heat	diff
<b>Panel A: Demographic variables</b>						
Age	19.8 (2.75)	20.1 (2.68)		21.9 (2.73)	22.1 (2.12)	
Female	0.59 (0.49)	0.60 (0.49)		0.35 (0.48)	0.35 (0.48)	
Weight (kg)	64.1 (12.97)	63.6 (12.36)		61.3 (7.84)	63.0 (9.27)	**
Height (cm)	169.2 (11.19)	168.7 (10.58)		166.4 (18.06)	166.9 (17.79)	
Mother's education (university)	0.66 (0.47)	0.69 (0.46)		0.12 (0.32)	0.16 (0.37)	
Father's education (university)	0.72 (0.45)	0.71 (0.45)		0.22 (0.41)	0.22 (0.41)	
Mother employed (self or wage)	0.78 (0.41)	0.76 (0.43)		0.65 (0.48)	0.67 (0.47)	
Father employed (self or wage)	0.95 (0.22)	0.93 (0.26)		0.77 (0.42)	0.76 (0.43)	
Parent income (> \$24k, PPP)	0.88 (0.33)	0.90 (0.30)		0.09 (0.28)	0.08 (0.27)	
In-state status ( $\geq 5$ years)	0.69 (0.46)	0.67 (0.47)				
Major ethnicity (%)				0.67 (0.47)	0.68 (0.47)	
Kikuyu ethnicity (%)				0.24 (0.42)	0.18 (0.38)	*
Luo ethnicity (%)				0.19 (0.39)	0.23 (0.42)	
<b>Panel B: Outdoor measures (session condition level)</b>						
Daily mean temperature (°C)	15.7 (3.8)	15.6 (3.7)		20.0 (0.9)	20.0 (0.9)	
Daily mean relative humidity (%)	59.9 (17.2)	60.1 (17.1)		65.7 (9.2)	65.5 (9.6)	

<b>Panel C: Indoor measures (session condition level)</b>					
Mean temperature (°C)	21.8 (0.6)	30.3 (1.1)	**	21.7 (0.6)	31.2 (1.6)
Mean relative humidity (%)	42.9 (6.8)	28.7 (6.2)	**	58.3 (6.1)	39.8 (7.9)
<b>Panel D: Primary outcomes</b>					
1. Production task	23.6 (3.8)	24.2 (3.9)	*	13.0 (5.4)	13.4 (6.2)
2. Fairness	0.31 (0.23)	0.29 (0.22)		0.32 (0.26)	0.32 (0.25)
3. Risk-taking	408.5 (477.1)	403.6 (450.8)		329.5 (422.5)	329.5 (406.8)
4. Rational choice violation	0.03 (0.16)	0.02 (0.15)		0.03 (0.16)	0.02 (0.15)
5. Patience	0.994 (0.005)	0.993 (0.005)		0.994 (0.013)	0.995 (0.013)
6. Time inconsistency	0.986 (0.174)	0.997 (0.183)		1.055 (0.402)	1.034 (0.345)
7. Trust	0.49 (0.34)	0.47 (0.34)		0.35 (0.28)	0.38 (0.29)
8. Public contribution	627.3 (434.6)	591.9 (452.5)		442.5 (398.8)	457.0 (372.9)
9. Fluid intelligence	0.93 (0.12)	0.94 (0.11)		0.80 (0.21)	0.83 (0.20)
10. Joy of Destruction	0.08 (0.24)	0.04 (0.17)	*	0.14 (0.26)	0.22 (0.29)
11. Cognitive reflection	0.46 (0.32)	0.43 (0.31)		0.21 (0.20)	0.23 (0.20)
12. Charitable donation	246.2 (537.3)	223.0 (493.1)		573.3 (677.1)	561.1 (708.3)
<b>Panel E: Affect</b>					
Happiness	4.61 (1.39)	4.67 (1.29)		5.90 (1.17)	5.72 (1.23)
Alertness	4.27 (1.62)	3.76 (1.62)	**	6.08 (1.32)	5.75 (1.54)
<b>Panel F: Perceptions of heat</b>					
<i>How was room temperature? (%)</i>					

Comfortable	0.79 (0.41)	0.09 (0.29)	**	0.57 (0.50)	0.47 (0.50)	*
Too cold	0.20 (0.40)	0 (0)	**	0.43 (0.50)	0.01 (0.10)	**
Too hot	0.01 (0.11)	0.91 (0.29)	**	0 (0)	0.52 (0.50)	**
<i>Was the experiment about... (%)</i>						
room temperature?	0.09 (0.29)	0.46 (0.50)	**	0.09 (0.29)	0.11 (0.31)	
generosity?	0.74 (0.44)	0.75 (0.44)		0.64 (0.48)	0.59 (0.49)	
gender discrimination?	0.06 (0.23)	0.07 (0.25)		0.03 (0.17)	0.02 (0.13)	

Note: Standard deviations in parentheses. “Prefer not to answer” and “Don’t know” responses are not included in the summary statistics presented above. Thus the responses are conditional on having given an answer that was not one of the aforementioned responses. The information in Panel A comes from the demographic survey held at the end of the experiment and is thus self-reported. University education (for both mother or father) is an indicator variable for the parent having completed university in the Nairobi sample or completing either a Bachelor or Advanced at the university level in the California sample. Employment status (for both mother or father) is an indicator variable for engagement in self-employment or wage-employment. The combined parental income variable is an indicator variable for making above \$24000 a year in the California sample or in Nairobi sample (PPP adjusted). In-state status is an indicator variable for being a resident of California for 5 years or more. Majority ethnicity records whether the respondent self-identified as an ethnicity comprising one of the five largest ethnic groups in Kenya (Kikuyu, Luhya, Luo, Kamba, and Kalenjin). In Panel E, “Happiness” is the value chosen on a scale from 1 to 7 asking how the participant felt at the time of taking the survey question, with 1 being sad and 7 being happy. Meanwhile, “Alertness” is the value chosen on a scale from 1 to 7 asking how the participant felt at the time of taking the survey question, with 1 being tired and 7 being alert. Questions and responses featured in Panel F come from the debriefing questions held at the end of the experiment. Note that participants were able to report that the experiment may have had more than one treatment.

\*  $p < .05$ , \*\*  $p < .01$



Table C.2: Distribution of responses for demographic variables

Age	No.	%	Mother education (Xlab)	No.	%
Response	1848	96	Completed High School	140	16
No response	70	4	Completed University (Bachelor)	318	35
Total	1918	100	Completed University (Advanced)	261	29
Weight	No.	%	Completed Vocational College	59	7
Response	1780	93	Other	78	9
No response	138	7	Prefer not to say	47	5
Total	1918	100	Total	903	100
Height	No.	%	Father education (Xlab)	No.	%
Response	1582	82	Completed High School	137	15
No response	336	18	Completed University (Bachelor)	253	28
Total	1918	100	Completed University (Advanced)	351	39
Gender	No.	%	Completed Vocational College	32	4
Male	1007	53	Other	73	8
Female	871	45	Prefer not to say	57	6
Prefer not to say	40	2	Total	903	100
Total	1918	100	Monthly parental income (Busara)	No.	%
Mother occupation	No.	%	Less than Ksh.10,000	208	20
Unemployed	512	27	Ksh.10,000 - Ksh. 30,000	224	22
Self-employed	527	27	Ksh. 30,000 - Ksh. 60,000	111	11
Wage-employed	729	38	Ksh. 60,000 - Ksh.100,000	74	7
Prefer not to say	150	8	More than Ksh.100,000	56	6
Total	1,918	100	Don't know	237	23
Father occupation	No.	%	Prefer not to say	105	10
Unemployed	251	13	Total	1015	100
Self-employed	612	32	Yearly parental income (Xlab)	No.	%
Wage-employed	803	42	Less than \$24,000	76	8
Prefer not to say	252	13	\$24,000 - \$45,000	82	9
Total	1,918	100	\$45,000 - \$75,000	100	11
Mother education (Busara)	No.	%	\$75,000 - \$121,000	144	16
Completed Primary	253	25	More than \$121,000	289	32
Completed Secondary	275	27	Don't know	130	14
Completed University	129	13	Prefer not to say	82	9
Completed technical college	226	22	Total	903	100
Other	57	6	Residency status (Xlab)	No.	%
Prefer not to say	75	7	No	207	23
Total	1,015	100	Yes, for less than 5 years	77	9
Father education (Busara)	No.	%	Yes, for 5 years or more	599	66
Completed Primary	162	16	Prefer not to say	20	2
Completed Secondary	230	23	Total	903	100
Completed University	198	20	Ethnicity (Busara Center)	No.	%
Completed technical college	253	25	Response	952	94
Other	64	6	No response	63	6
Prefer not to say	108	11	Total	1,015	100
Total	1,015	100			

Note: The above information comes from the demographic survey held at the end of the experiment and is thus self-reported. The two attritors for which we did not have demographic information have not been included in the above table. All demographic variables that can be measured for the pooled sample are balanced across treatment and control for the pooled sample at the 5 percent level.

Table C.3: Post-experiment survey responses

	California			Nairobi		
	Control	Heat	diff	Control	Heat	diff
<b>Perceptions of laboratory environment (prop.)</b>						
<i>How would you describe the comfort of your chair?</i>						
Comfortable	0.86 (0.35)	0.77 (0.42)	*	0.99 (0.12)	0.98 (0.15)	
Uncomfortable	0.14 (0.35)	0.23 (0.42)	*	0.01 (0.12)	0.02 (0.15)	
<i>How would you describe the screen brightness?</i>						
Comfortable	0.82 (0.39)	0.86 (0.35)		0.96 (0.19)	0.92 (0.27)	*
Too bright	0.16 (0.37)	0.13 (0.34)		0.03 (0.18)	0.08 (0.27)	*
Too dim	0.02 (0.14)	0.01 (0.07)		0.01 (0.08)	0.00 (0.06)	
<i>How would you describe the air in the room?</i>						
Comfortable	0.94 (0.24)	0.34 (0.47)	**	0.85 (0.36)	0.60 (0.49)	**
Too dry	0.05 (0.22)	0.25 (0.43)	**	0.05 (0.21)	0.26 (0.44)	**
Too humid	0.01 (0.11)	0.41 (0.49)	**	0.10 (0.31)	0.14 (0.35)	
<i>How would you describe the temperature in the room?</i>						
Comfortable	0.79 (0.41)	0.09 (0.29)	**	0.57 (0.50)	0.47 (0.50)	*
Too cold	0.20 (0.40)	0 (0)	**	0.43 (0.50)	0.01 (0.10)	**
Too hot	0.01 (0.11)	0.91 (0.29)	**	0 (0)	0.52 (0.50)	**
<i>How would you describe the lighting in the room?</i>						
Comfortable	0.86 (0.35)	0.84 (0.36)		0.96 (0.19)	0.96 (0.21)	
Too bright	0.08 (0.27)	0.08 (0.27)		0.04 (0.19)	0.03 (0.17)	
Too dim	0.06 (0.23)	0.07 (0.27)		0.00 (0.06)	0.01 (0.12)	
<i>How would you describe the space in the room?</i>						
Comfortable	0.79 (0.41)	0.71 (0.46)		0.94 (0.25)	0.89 (0.31)	
Too empty	0.04 (0.20)	0.02 (0.13)		0.00 (0.06)	0 (0)	
Too crowded	0.16	0.27	*	0.06	0.11	*

	(0.37)	(0.44)	(0.24)	(0.31)
<i>How would you describe the ambient noise in the room?</i>				
Comfortable	0.83 (0.38)	0.69 (0.46)	**	0.85 (0.36)
Too silent	0.09 (0.28)	0.03 (0.17)	*	0.13 (0.33)
Too noisy	0.08 (0.27)	0.28 (0.45)	**	0.03 (0.17)
<b>What do you think the experiment was about? (prop.)</b>				
Generosity	0.74 (0.44)	0.75 (0.44)		0.64 (0.48)
Screen brightness	0.09 (0.29)	0.07 (0.25)		0.06 (0.23)
Room lighting	0.11 (0.31)	0.08 (0.27)		0.06 (0.25)
Gender discrimination	0.06 (0.23)	0.07 (0.25)		0.03 (0.17)
Room temperature	0.09 (0.29)	0.46 (0.50)	**	0.09 (0.29)
Ambient noise	0.13 (0.34)	0.16 (0.36)		0.05 (0.22)
Room space	0.14 (0.35)	0.15 (0.36)		0.08 (0.27)
None of the above	0.12 (0.33)	0.08 (0.27)		0.17 (0.37)
Other	0.21 (0.41)	0.12 (0.32)	**	0.14 (0.35)

Note: Full set of responses to the debriefing questions, by treatment status and site. Standard deviations in parentheses. Note that only 913 participants were able to provide answers, as the post-experiment survey was implemented halfway into the experiment. Note also that participants could have listed that the experiment was about more than one subject. \*  $p < .05$ , \*\*  $p < .01$ .

Table C.4: Ethnicity-Party Alignment

Luo	Other pro-opposition	Other pro-government	Kikuyu
Luo (196)	Luhya (165) Kamba (119) Coastal (10) Kuria (4) Somali (3)	Kalenjin (96) Kisii (74) Meru (32) Maasai (18) Embu (11) Turkana (9) Samburu (6) Taita (6) Oromo (3) Orma (1) Terik (1)	Kikuyu (198)
<i>N</i> = 196	<i>N</i> = 301	<i>N</i> = 257	<i>N</i> = 198

Note: The classifications above follow regional voting patterns in the 2017 Kenyan elections. The 63 participants who preferred not to state their ethnicity are not classified into the above table. Luhya includes people who identify with either Luhya (96) or Luhya subtribes (e.g., Bukusu (48), Maragoli (12), Marama (5), Isukha (2), Tachoni (2)). Coastal includes people who identify with Coastal tribes (e.g., Digo (2), Giriama (2), Ribe (2), Duruma (1), Mijikenda (1), Pokomo (1), Swahili (1)). Kalenjin includes people who identify with either Kalenjin (34) or Kalenjin subtribes (e.g., Kipsigis (24), Nandi (14), Tugen (12), Marakwet (6), Elgeyo (4), Pokot (2)).

## D Main results

### D.1 Main results with primary outcomes

Table D.1.1: Precision task

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.516 (0.049)* [0.384]	0.695 (0.107) [0.912]	0.590 (0.083) [0.457]	0.451 (0.253) [0.722]
Male	1.974 (0.000)**	2.104 (0.000)**	1.892 (0.000)**	2.049 (0.000)**
Male $\times$ Heat		-0.261 (0.587) [1.000]		
Nairobi	-11.19 (0.000)**	-11.16 (0.000)**		
Nairobi $\times$ Heat		-0.0746 (0.884) [0.866]		
Outcome control mean	17.95	17.95	23.63	12.99
R-squared	0.549	0.549	0.0630	0.0295
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The precision task is also known as the slider task. The outcome in this table is the number of correct sliders made in three minutes. Final earnings from the precision task are based off either being weakly above (high) or below (low) the median within treatment cohort. The median pair is randomly assigned to high or low. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.2: Fairness

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.0167 (0.170) [0.618]	-0.0197 (0.278) [0.912]	-0.0245 (0.114) [0.457]	-0.0101 (0.586) [1.000]
Male	-0.0407 (0.001)**	-0.0349 (0.046)*	-0.0783 (0.000)**	-0.00606 (0.726)
Male $\times$ Heat		-0.0116 (0.613) [1.000]		
Nairobi	0.0337 (0.012)*	0.0251 (0.163)		
Nairobi $\times$ Heat		0.0172 (0.474) [0.789]		
Outcome control mean	0.318	0.318	0.309	0.326
R-squared	0.0102	0.0106	0.0323	0.000519
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. Fairness here refers to the real effort dictator game, where the level of endowment is determined by the number of correct sliders made in the precision task. The outcome in this table is the share of joint earnings (2400 tokens in the high group, 1200 tokens in the low group) that each participant desires to give to the anonymous partner. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.3: Risk-taking

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-1.835 (0.929) [1.000]	1.727 (0.954) [0.912]	-5.553 (0.858) [0.850]	1.926 (0.944) [1.000]
Male	163.6 (0.000)**	173.5 (0.000)**	316.8 (0.000)**	22.10 (0.422)
Male $\times$ Heat		-19.81 (0.656) [1.000]		
Nairobi	-116.3 (0.000)**	-122.9 (0.000)**		
Nairobi $\times$ Heat		13.21 (0.771) [0.861]		
Outcome control mean	366.3	366.3	409.0	329.1
R-squared	0.0396	0.0397	0.112	0.000650
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The outcome in this table is the variance of the coin toss from menu A, in tokens. Note that the expected value is not constant across each coin, so that the outcome does not capture the trade-off between expected value and variance. Note also that under this approach, Coin 7 will be as good as Coin 5, even though Coin 5 strictly dominates Coin 7 with expected utility. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.4: Rational choice violation I

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.00431 (0.561) [1.000]	-0.00303 (0.803) [0.912]	-0.00455 (0.687) [0.768]	-0.00409 (0.678) [1.000]
Male	0.00268 (0.725)	0.00458 (0.701)	0.00540 (0.627)	0.000163 (0.988)
Male $\times$ Heat		-0.00380 (0.815) [1.000]		
Nairobi	0.0000607 (0.993)	-0.000649 (0.955)		
Nairobi $\times$ Heat		0.00143 (0.928) [0.866]		
Outcome control mean	0.0277	0.0277	0.0275	0.0279
R-squared	0.000259	0.000294	0.000499	0.000166
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The outcome in this table is an indicator of transitivity violation using both menus A & B. A transitivity violation comes from choosing two coins in the interior region of the intersection of both menus, where it is not the case that it can be said that one preferring coin A to coin B and then preferring coin B to coin C implies that one prefers coin A to coin C. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.5: Patience

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.00005823 (0.854) [1.000]	-0.00005823 (0.854) [0.912]	-0.00060946 (0.093) [0.457]	0.00096101 (0.279) [0.722]
Heat (Male)		-0.0002442 (0.702) [1.000]		
Heat (Nairobi)		0.00157046 (0.102) [0.679]		
Outcome control mean	0.9938358	0.9938358	0.9937076	0.9939217
Observations	6612	6612	3200	3412

Note: Standard errors clustered at the individual level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. For (1) and (2), “Heat” refers to the difference between the aggregate delta statistics for the treatment and control groups, divided by the square root of the sum of squared standard errors. For (2), “Heat (Male)” refers to the difference between the aggregate delta statistics for male treatment and control groups, subtracting the difference between the aggregate delta statistics for female treatment and control groups, and then dividing by the square root of the sum of squared standard errors. “Heat (Nairobi)” refers to the difference between the aggregate delta statistics for Nairobi treatment and control groups, subtracting the difference between the aggregate delta statistics for California treatment and control groups, and then dividing by the square root of the sum of squared standard errors. (3) and (4) carry out a similar estimation to “Heat” for (1) and (2) but for the California sample and Nairobi sample, respectively. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. Given that the non-linear least squares specification used to measure patience does not incorporate interaction terms, we instead use  $p$ -values from analogous differences in this table for multiple testing adjustments. The outcome in this table is the aggregate  $\delta$  statistic from the non-linear least squares specification featured in Andreoni et al. (2015), carried out at the individual-choice level.  $\delta$  is the daily discount factor between two future days. For comparability with regression results, individuals who did not respond “Female” or “Male” to the gender survey question were dropped (2% of the sample). Before estimation of aggregate parameters, individuals who never altered their decision from a specific corner solution in all convex time budgets were dropped (as they provided insufficient variation for the calculation of utility parameters) as were individuals who exhibited generalized axiom of revealed preference (GARP) violations.

\*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.6: Time inconsistency

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.00048029 (0.961) [1.000]	0.00048029 (0.961) [0.912]	-0.00878209 (0.489) [0.768]	0.0179143 (0.487) [1.000]
Heat (Male)		-0.00824147 (0.673) [1.000]		
Heat (Nairobi)		0.02669638 (0.353) [0.789]		
Outcome control mean	1.001042	1.001042	0.9886241	1.05313
Observations	6612	6612	3200	3412

Note: Standard errors clustered at the individual level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. For (1) and (2), “Heat” refers to the difference between the aggregate beta statistics for the treatment and control groups, divided by the square root of the sum of squared standard errors. For (2), “Heat (Male)” refers to the difference between the aggregate beta statistics for male treatment and control groups, then subtracting the difference between the aggregate beta statistics for female treatment and control groups, and then dividing by the square root of the sum of squared standard errors. “Heat (Nairobi)” refers to the difference between the aggregate beta statistics for Nairobi treatment and control groups, subtracting the difference between the aggregate beta statistics for California treatment and control groups, and then dividing by the square root of the sum of squared standard errors. (3) and (4) carry out a similar estimation to “Heat” for (1) and (2) but for the California sample and Nairobi sample, respectively. Note that the effects presented above are multiplied by -1, so that a positive difference reflects more time inconsistency. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. Given that the non-linear least squares specification used to measure patience does not incorporate interaction terms, we instead use  $p$ -values from analogous differences in this table for multiple testing adjustments. The outcome in this table is the aggregate  $\beta$  statistic from the non-linear least squares specification featured in Andreoni et al. (2015), carried out at the individual-choice level.  $\beta$  measures present bias, and values less than 1 denote time inconsistency. For comparability with regression results, individuals who did not respond “Female” or “Male” to the gender survey question were dropped (2% of the sample). Before estimation of aggregate parameters, individuals who never altered their decision from a specific corner solution in all convex time budgets were dropped (as they provided insufficient variation for the calculation of utility parameters) as were individuals who exhibited generalized axiom of revealed preference (GARP) violations. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.7: Trust

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.00780 (0.617) [1.000]	-0.0356 (0.157) [0.912]	-0.0215 (0.377) [0.759]	0.0333 (0.096) [0.471]
Male	0.0699 (0.000)**	0.0523 (0.016)*	0.0582 (0.020)*	0.0806 (0.000)**
Male $\times$ Heat		0.0350 (0.235) [1.000]		
Nairobi	-0.137 (0.000)**	-0.160 (0.000)**		
Nairobi $\times$ Heat		0.0460 (0.164) [0.679]		
Outcome control mean	0.417	0.417	0.495	0.348
R-squared	0.0465	0.0490	0.00804	0.0215
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The outcome in this table is the share of endowed tokens (out of 600) entrusted to the other person in the first round of the trust game. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.8: Public contribution

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-8.923 (0.647) [1.000]	-36.93 (0.260) [0.912]	-36.88 (0.213) [0.520]	15.32 (0.555) [1.000]
Male	38.11 (0.067)	37.83 (0.157)	9.059 (0.762)	64.82 (0.025)*
Male $\times$ Heat		0.454 (0.990) [1.000]		
Nairobi	-170.5 (0.000)**	-196.4 (0.000)**		
Nairobi $\times$ Heat		51.93 (0.191) [0.679]		
Outcome control mean	529.7	529.7	629.6	442.5
R-squared	0.0383	0.0392	0.00184	0.00674
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. Public contribution here refers to the public goods game. The outcome in this table is the amount of tokens (out of 1200) put into the shared fund. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.9: Fluid intelligence

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.0150 (0.061) [0.384]	-0.0102 (0.371) [0.912]	0.000858 (0.919) [0.850]	0.0273 (0.037)* [0.256]
Male	0.0176 (0.045)*	0.00387 (0.757)	0.0207 (0.012)*	0.0146 (0.334)
Male $\times$ Heat		0.0273 (0.093) [1.000]		
Nairobi	-0.127 (0.000)**	-0.137 (0.000)**		
Nairobi $\times$ Heat		0.0197 (0.220) [0.679]		
Outcome control mean	0.862	0.862	0.935	0.799
R-squared	0.121	0.124	0.00787	0.00567
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. Fluid intelligence is measured through Raven's Progressive Matrices. The outcome in this table is the share of six matrices answered correctly. Each puzzle answered correctly yields an Airtime Voucher worth 50 KSh (or an Amazon Gift Card worth 1 dollar in the California sample), which provides the earnings to be used for the next module. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.10: Joy of Destruction

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.0234 (0.069) [0.384]	-0.0200 (0.167) [0.912]	-0.0409 (0.004)** [0.053]	0.0792 (0.000)** [0.001]++
Male	0.0369 (0.001)**	0.0626 (0.000)**	0.0252 (0.106)	0.0474 (0.004)**
Male × Heat			-0.0519 (0.016)* [0.238]	
Nairobi	0.113 (0.000)**	0.0463 (0.008)**		
Nairobi × Heat			0.133 (0.000)** [0.001]++	
Outcome control mean	0.110	0.110	0.0771	0.139
R-squared	0.0643	0.0805	0.0138	0.0267
Observations	1859	1859	864	995

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The outcome in this table is the share of the anonymous partner's earned Airtime Vouchers (Amazon Gift Cards in the California sample) destroyed by the participant. The earned Airtime Vouchers and Amazon Gift Cards resulted from the number of Raven's Progressive Matrices answered correctly in the previous module, where one puzzle answered correctly yielded one Airtime Voucher or one Amazon Gift Card. Airtime Vouchers were worth 50 KSh each, while Amazon Gift Cards were worth 1 dollar each. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.11: Cognitive reflection

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.00228 (0.837) [1.000]	-0.0268 (0.168) [0.912]	-0.0228 (0.217) [0.520]	0.0159 (0.218) [0.722]
Male	0.0845 (0.000)**	0.0800 (0.000)**	0.177 (0.000)**	-0.000810 (0.954)
Male $\times$ Heat		0.00894 (0.717) [1.000]		
Nairobi	-0.245 (0.000)**	-0.263 (0.000)**		
Nairobi $\times$ Heat		0.0369 (0.128) [0.679]		
Outcome control mean	0.325	0.325	0.455	0.212
R-squared	0.178	0.179	0.0773	0.00162
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The outcome in this table is the share of questions (out of 5) from the Cognitive Reflection Test answered correctly. \*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

Table D.1.12: Charitable donation

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-4.443 (0.887) [1.000]	-54.72 (0.215) [0.912]	-12.67 (0.679) [0.768]	0.412 (0.994) [1.000]
Male	7.254 (0.820)	8.537 (0.845)	-58.73 (0.104)	73.31 (0.156)
Male $\times$ Heat			-1.045 (0.986) [1.000]	
Nairobi	323.8 (0.000)**	302.4 (0.000)**		
Nairobi $\times$ Heat			44.54 (0.468) [0.789]	
Matched with ingroup charity	-9.405 (0.771)	-66.75 (0.148)	31.79 (0.409)	-76.37 (0.186)
Matched with ingroup charity $\times$ Heat			121.0 (0.090)	
Earnings in tokens	0.00356 (0.734)	0.00377 (0.719)	-0.0182 (0.117)	0.0295 (0.100)
Outcome control mean	410.4	410.4	245.0	561.8
R-squared	0.0657	0.0672	0.00695	0.00685
Observations	1806	1806	861	945

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Multiple testing adjusted False Discovery Rate (FDR)  $q$ -value significance level in square brackets. Within Section D.2, for each specification (1)-(4), multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with Heat, across primary outcomes. Additionally, within Section D.2, in (2) multiple hypothesis testing adjustments are performed on the set of  $p$ -values associated with the interaction between Heat and Male, across primary outcomes, as well as on the set of  $p$ -values associated with the interaction between Heat and Nairobi, across primary outcomes. The outcome in this table is the amount of tokens earned in the experiment that is donated to the randomly selected charity. In Nairobi, “Matched with ingroup charity” is an indicator taking on a value of one if a participant is matched to a charity associated with her ethnicity, and 0 otherwise. In California, “Matched with ingroup charity” is an indicator taking on a value of one if a participant has resided in the San Francisco Bay Area for five years or more and is matched with a charity in the San Francisco Bay Area. “Earnings in tokens” captures the amount of tokens earned in the experiment.

\*  $p < .05$ , \*\*  $p < .01$ ; +  $q < .05$ , ++  $q < .01$

## D.2 Main results with additional outcomes

Table D.2.1: Rational choice violation II

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.0119 (0.313)	-0.0153 (0.459)	-0.00701 (0.660)	-0.0163 (0.344)
Male	0.000669 (0.955)	-0.00977 (0.593)	-0.0311 (0.055)	0.0301 (0.073)
Male $\times$ Heat		0.0209 (0.402)		
Nairobi	0.0471 (0.000)**	0.0544 (0.004)**		
Nairobi $\times$ Heat		-0.0147 (0.521)		
Outcome control mean	0.0896	0.0896	0.0618	0.1138
R-squared	0.00771	0.00812	0.00446	0.00288
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is an indicator for having made a first order stochastic dominance (F OSD) violation by having selected coin 7 in menu A in the risk preference module. \*  $p < .05$ , \*\*  $p < .01$

Table D.2.2: Rational choice violation III

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.00628 (0.458)	-0.0131 (0.102)	-0.00000483 (0.999)	0.0117 (0.437)
Male	0.00603 (0.487)	-0.0102 (0.366)	-0.00106 (0.879)	0.0126 (0.418)
Male $\times$ Heat		0.0325 (0.032)*		
Nairobi	0.0491 (0.000)**	0.0473 (0.000)**		
Nairobi $\times$ Heat		0.00357 (0.824)		
Outcome control mean	0.0330	0.0330	0.0092	0.0539
R-squared	0.0188	0.0208	0.0000296	0.00125
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is an indicator for having made a violation of the Generalized Axiom of Revealed Preference (GARP) between menu A and B or between menu C and D in the time preference module. \*  $p < .05$ , \*\*  $p < .01$

Table D.2.3: Patience, non-parametric, menu A

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.0318 (0.673)	0.0515 (0.664)	0.0341 (0.744)	0.0300 (0.782)
Male	-0.113 (0.119)	-0.0907 (0.394)	-0.0553 (0.595)	-0.166 (0.103)
Male × Heat		-0.0439 (0.783)		
Nairobi	-1.127 (0.000)**	-1.130 (0.000)**		
Nairobi × Heat		0.00717 (0.963)		
Outcome control mean	4.471	4.471	5.087	3.934
R-squared	0.117	0.117	0.000398	0.00257
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the choice from Menu A, where choice 1 (with value 1) equals 840 tokens today and 0 tokens in 3 weeks, and choice 6 (with value 6) equals 0 tokens today and 1200 tokens in 3 weeks (the trade-off is fixed between 1 token today or 1.43 tokens in 3 weeks). \*  $p < .05$ , \*\*  $p < .01$

Table D.2.4: Patience, non-parametric, menu B

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.116 (0.151)	-0.163 (0.259)	-0.128 (0.294)	-0.105 (0.329)
Male	-0.411 (0.000)**	-0.455 (0.000)**	-0.456 (0.000)**	-0.369 (0.001)**
Male $\times$ Heat		0.0884 (0.589)		
Nairobi	1.393 (0.000)**	1.393 (0.000)**		
Nairobi $\times$ Heat		0.000451 (0.998)		
Outcome control mean	2.803	2.803	2.119	3.399
R-squared	0.131	0.131	0.0164	0.0117
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the choice from Menu B, where choice 1 (with value 1) equals 1020 tokens today and 0 tokens in 3 weeks, and choice 6 (with value 6) equals 0 tokens today and 1020 tokens in 3 weeks (the trade-off is fixed between 1 token today or 1 token in 3 weeks). \*  $p < .05$ , \*\*  $p < .01$

Table D.2.5: Patience, non-parametric, menu C

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.126 (0.120)	-0.0929 (0.509)	-0.176 (0.172)	-0.0823 (0.422)
Male	-0.201 (0.014)*	-0.0964 (0.415)	-0.0231 (0.835)	-0.365 (0.002)**
Male $\times$ Heat		-0.209 (0.250)		
Nairobi	-1.305 (0.000)**	-1.378 (0.000)**		
Nairobi $\times$ Heat		0.147 (0.397)		
Outcome control mean	4.290	4.290	5.039	3.637
R-squared	0.143	0.144	0.00256	0.0122
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the choice from Menu C, where choice 1 (with value 1) equals 840 tokens in 3 weeks and 0 tokens in 7 weeks, and choice 6 (with value 6) equals 0 tokens in 3 weeks and 1200 tokens in 7 weeks (the trade-off is fixed between 1 token in 3 weeks or 1.43 tokens in 7 weeks). \*  $p < .05$ , \*\*  $p < .01$

Table D.2.6: Patience, non-parametric, menu D

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.0168 (0.818)	-0.124 (0.279)	-0.118 (0.230)	0.0709 (0.508)
Male		-0.196 (0.007)**	-0.204 (0.049)*	-0.179 (0.063)
Male $\times$ Heat			0.0165 (0.910)	
Nairobi	1.492 (0.000)**	1.400 (0.000)**		
Nairobi $\times$ Heat			0.184 (0.199)	
Outcome control mean	2.519	2.519	1.799	3.148
R-squared	0.169	0.170	0.00468	0.00408
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the choice from Menu D, where choice 1 equals 1020 tokens in 3 weeks and 0 tokens in 7 weeks, and choice 6 equals 0 tokens in 3 weeks and 1020 tokens in 7 weeks (the trade-off is fixed between 1 token in 3 weeks or 1 token in 7 weeks). \*  $p < .05$ , \*\*  $p < .01$

Table D.2.7: Trustworthiness

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	0.00855 (0.407)	0.0293 (0.054)	0.00878 (0.507)	0.00584 (0.706)
Male	-0.0258 (0.022)*	-0.00243 (0.878)	-0.0336 (0.039)*	-0.0226 (0.152)
Male × Heat		-0.0469 (0.034)*		
Nairobi	0.100 (0.000)**	0.0965 (0.000)**		
Nairobi × Heat		0.00842 (0.688)		
sharesentto	0.115 (0.000)**	0.115 (0.000)**	0.201 (0.000)**	0.00892 (0.780)
sharesent	0.271 (0.000)**	0.273 (0.000)**	0.188 (0.000)**	0.383 (0.000)**
Outcome outcome mean	0.245	0.245	0.227	0.262
R-squared	0.150	0.152	0.155	0.176
Observations	1580	1580	731	849

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the share of tokens (received and multiplied by 3) given back to the anonymous partner in the trust game. sharesentto captures the share of 600 tokens that the recipient received from the anonymous partner in the first round. sharesent captures how much the share of 600 tokens sent that the individual sent to her anonymous partner in the first round. \*  $p < .05$ , \*\*  $p < .01$

Table D.2.8: Trust measure

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.0288 (0.819)	0.0906 (0.626)	0.105 (0.534)	-0.147 (0.426)
Male	-0.142 (0.280)	-0.162 (0.383)	-0.504 (0.005)**	0.193 (0.304)
Male × Heat		0.0405 (0.863)		
Nairobi	0.388 (0.004)**	0.520 (0.006)**		
Nairobi × Heat		-0.264 (0.311)		
Outcome control mean	6.220	6.220	5.963	6.443
R-squared	0.00475	0.00530	0.00971	0.00167
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the number chosen from a 0-to-10 scale that asked how well does the following statement described one as a person: as long as I am not convinced otherwise, I assume that people have only the best intentions (where 0 means does not describe one at all, and a 10 means describes one perfectly). \*  $p < .05$ , \*\*  $p < .01$

Table D.2.9: Correct beliefs about others' contributions

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.000439 (0.982)	-0.0303 (0.371)	0.00687 (0.822)	-0.00682 (0.797)
Male	0.00693 (0.687)	-0.0393 (0.116)	0.00145 (0.952)	0.0120 (0.626)
Male × Heat		0.0924 (0.011)*		
Nairobi	0.00303 (0.884)	0.0213 (0.470)		
Nairobi × Heat		-0.0368 (0.371)		
Outcome control mean	0.212	0.212	0.206	0.218
R-squared	0.000101	0.00315	0.0000742	0.000264
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is an indicator for having guessed correctly about one other person's contribution in the group towards the fund in the public goods module. \*  $p < .05$ , \*\*  $p < .01$

Table D.2.10: Time spent on Cognitive Reflection Test

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	1.188 (0.338)	1.879 (0.402)	2.043 (0.354)	0.429 (0.743)
Male	-2.885 (0.017)*	-3.126 (0.051)	-7.421 (0.000)**	1.250 (0.363)
Male × Heat		0.487 (0.832)		
Nairobi	8.389 (0.000)**	9.273 (0.000)**		
Nairobi × Heat		-1.770 (0.515)		
Outcome control mean	167.8	167.8	163.3	171.8
R-squared	0.0287	0.0290	0.0172	0.00114
Observations	1866	1866	862	1004

Note: Standard errors clustered at the session level. Per-comparison *p*-values in parentheses. The outcome in this table is the total amount of time spent on the Cognitive Reflection Test, where participants were allowed 3 minutes in total to complete 5 questions. \* *p* < .05, \*\* *p* < .01

Table D.2.11: Happiness

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.0791 (0.208)	0.0426 (0.657)	0.0490 (0.579)	-0.190 (0.031)*
Male		0.174 (0.005)**	0.167 (0.073)	0.208 (0.025)*
Male × Heat			0.0155 (0.906)	
Nairobi	1.127 (0.000)**	1.248 (0.000)**		
Nairobi × Heat			-0.243 (0.073)	
Outcome control mean	5.313	5.313	4.625	5.914
R-squared	0.180	0.182	0.00618	0.00952
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being sad and 7 being happy. \*  $p < .05$ , \*\*  $p < .01$

Table D.2.12: Alertness

	(1) Pooled	(2) Pooled	(3) California	(4) Nairobi
Heat	-0.420 (0.000)**	-0.576 (0.000)**	-0.530 (0.000)**	-0.322 (0.001)**
Male	0.389 (0.000)**	0.333 (0.001)**	0.729 (0.000)**	0.0743 (0.439)
Male × Heat		0.112 (0.446)		
Nairobi	1.817 (0.000)**	1.726 (0.000)**		
Nairobi × Heat		0.181 (0.230)		
Outcome control mean	5.247	5.247	4.281	6.090
R-squared	0.303	0.305	0.0739	0.0132
Observations	1878	1878	874	1004

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being tired and 7 being alert. \*  $p < .05$ , \*\*  $p < .01$

## E Exploring the effect of heat on destruction

Table E.1: Heterogeneous treatment effects, Nairobi

	(1) Destruction	(2) Destruction	(3) Destruction	(4) Destruction
Heat	0.0804 (0.000)**	0.118 (0.000)**	0.148 (0.002)**	0.0743 (0.000)**
Parent university education	-0.0142 (0.649)			-0.0122 (0.708)
Parent university education $\times$ Heat	-0.0122 (0.763)			0.00888 (0.831)
Above median cognition		-0.00575 (0.803)		-0.00136 (0.954)
Above median cognition $\times$ Heat		-0.0801 (0.027)*		-0.0711 (0.057)
Other eth			0.0366 (0.253)	0.0291 (0.387)
Other eth $\times$ Heat			-0.0697 (0.152)	-0.0561 (0.256)
Kikuyu			0.0173 (0.627)	0.0109 (0.770)
Kikuyu $\times$ Heat			-0.161 (0.005)**	-0.135 (0.022)*
Outcome control mean	0.139	0.139	0.143	0.145
R-squared	0.0201	0.0321	0.0344	0.0399
Observations	959	995	944	895

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the share of the anonymous partner's earned Airtime Vouchers (Amazon Gift Cards in the California sample) destroyed by the participant. The earned Airtime Vouchers and Amazon Gift Cards resulted from the number of Raven's Progressive Matrices answered correctly, where one puzzle answered correctly yielded one Airtime Voucher or one Amazon Gift Card. Airtime Vouchers were worth 50 KSh each, while Amazon Gift Cards were worth 1 dollar each. "Parent university education" is an indicator for whether the participant identifies at least one parent as having university education. "Above median cognition" is an indicator for being above the site-specific cognition score, where cognition score is developed by normalizing the sum of individuals' normalized results from the fluid intelligence and cognitive reflection tasks. "Other eth" is an indicator variable for the individual self-identifying as an ethnicity that is not Kikuyu or Luo. "Kikuyu" is an indicator variable for the individual self-identifying as Kikuyu. The full regression model in the last column uses sample demeaned regressors. Thus, the treatment effect on Heat is for the Luo population. \*  $p < .05$ , \*\*  $p < .01$

Table E.2: Heterogeneous treatment effects, California

	(1) Destruction	(2) Destruction	(3) Destruction
Heat	-0.0565 (0.035)*	-0.0334 (0.095)	-0.0435 (0.003)**
Parent university education	0.00805 (0.781)		0.0116 (0.698)
Parent university education $\times$ Heat	0.0221 (0.482)		0.0275 (0.386)
Above median cognition		-0.0149 (0.542)	-0.0156 (0.541)
Above median cognition $\times$ Heat		-0.0156 (0.547)	-0.0250 (0.352)
Outcome control mean	0.077	0.077	0.079
R-squared	0.0110	0.0136	0.0188
Observations	854	864	835

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the share of the anonymous partner's earned Airtime Vouchers (Amazon Gift Cards in the California sample) destroyed by the participant. The earned Airtime Vouchers and Amazon Gift Cards resulted from the number of Raven's Progressive Matrices answered correctly, where one puzzle answered correctly yielded one Airtime Voucher or one Amazon Gift Card. Airtime Vouchers were worth 50 KSh each, while Amazon Gift Cards were worth 1 dollar each. "Parent university education" is an indicator for whether the participant identifies at least one parent as having university education. "Above median cognition" is an indicator for being above the site-specific cognition score, where cognition score is developed by normalizing the sum of individuals' normalized results from the fluid intelligence and cognitive reflection tasks. The full regression model in the last column uses sample demeaned regressors. \*  $p < .05$ , \*\*  $p < .01$

Table E.3: First stage results, Nairobi

	(1) Happiness	(2) Alertness	(3) Discomfort	(4) Temperature salience
Heat	-0.199 (0.110)	-0.280 (0.031)*	0.108 (0.035)*	0.0152 (0.521)
Outcome control mean	5.917	6.083	0.422	0.090
R-squared	0.00697	0.00964	0.0117	0.000652
Observations	571	571	571	571

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Happiness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being sad and 7 being happy. Alertness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being tired and 7 being alert. “Discomfort” is an indicator for whether the participant indicated another choice aside from “comfortable” in the post-experiment survey. “Temperature salience” is an indicator for whether the participant indicated temperature in the post-experiment survey as something the experiment was studying. \*  $p < .05$ , \*\*  $p < .01$

Table E.4: Mechanisms, Nairobi

	(1) Destruction	(2) Destruction	(3) Destruction	(4) Destruction	(5) Destruction
Heat	0.161 (0.247)	-0.00531 (0.961)	0.111 (0.004)**	0.0930 (0.002)**	0.0902 (0.001)**
Happiness		-0.0106 (0.526)			-0.00599 (0.678)
Happiness × Heat		-0.0122 (0.600)			-0.0237 (0.335)
Alertness			-0.00778 (0.642)		-0.00662 (0.673)
Alertness × Heat			0.0167 (0.361)		0.0229 (0.200)
Discomfort				0.0281 (0.403)	0.0225 (0.510)
Discomfort × Heat				-0.0383 (0.479)	-0.0437 (0.442)
Temperature salience					0.0876 (0.127) 0.0854 (0.142)
Temperature salience × Heat					-0.00414 (0.963) 0.00215 (0.982)
Outcome control mean	0.153	0.153	0.153	0.153	0.153
R-squared	0.0310	0.0272	0.0268	0.0330	0.0433
Observations	571	571	571	571	571

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the share of the anonymous partner's earned Airtime Vouchers (Amazon Gift Cards in the California sample) destroyed by the participant. The earned Airtime Vouchers and Amazon Gift Cards resulted from the number of Raven's Progressive Matrices answered correctly, where one puzzle answered correctly yielded one Airtime Voucher or one Amazon Gift Card. Airtime Vouchers were worth 50 KSh each, while Amazon Gift Cards were worth 1 dollar each. Happiness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being sad and 7 being happy. Alertness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being tired and 7 being alert. "Discomfort" is an indicator for whether the participant indicated another choice aside from "comfortable" in the post-experiment survey. "Temperature salience" is an indicator for whether the participant indicated temperature in the post-experiment survey as something the experiment was studying. \*  $p < .05$ , \*\*  $p < .01$

Table E.5: First stage results, California

	(1)	(2)	(3)	(4)
	Happiness	Alertness	Discomfort	Temperature salience
Heat	0.134 (0.372)	-0.499 (0.004)**	0.698 (0.000)**	0.361 (0.000)**
Outcome control mean	4.506	4.381	0.213	0.094
R-squared		0.00252	0.0259	0.498
Observations		338	338	338

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. Happiness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being sad and 7 being happy. Alertness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being tired and 7 being alert. “Discomfort” is an indicator for whether the participant indicated another choice aside from “comfortable” in the post-experiment survey. “Temperature salience” is an indicator for whether the participant indicated temperature in the post-experiment survey as something the experiment was studying. \*  $p < .05$ , \*\*  $p < .01$

Table E.6: Mechanisms, California

	(1) Destruction	(2) Destruction	(3) Destruction	(4) Destruction	(5) Destruction
Heat	-0.155 (0.053)	-0.0288 (0.627)	-0.0600 (0.118)	-0.0157 (0.592)	-0.0148 (0.541)
Happiness		-0.0161 (0.140)			-0.0212 (0.092)
Happiness $\times$ Heat		0.0275 (0.069)			0.0308 (0.072)
Alertness			0.00775 (0.540)		0.0117 (0.401)
Alertness $\times$ Heat			0.000798 (0.959)		-0.00491 (0.774)
Discomfort				-0.0648 (0.017)*	-0.0843 (0.012)*
Discomfort $\times$ Heat				0.0831 (0.074)	0.110 (0.039)*
Temperature salience					0.0616 (0.533)
Temperature salience $\times$ Heat					0.0856 (0.393)
Outcome control mean	0.078	0.078	0.078	0.078	0.078
R-squared	0.0127	0.00829	0.0125	0.00913	0.0336
Observations	338	338	338	338	338

Note: Standard errors clustered at the session level. Per-comparison  $p$ -values in parentheses. The outcome in this table is the share of the anonymous partner's earned Airtime Vouchers (Amazon Gift Cards in the California sample) destroyed by the participant. The earned Airtime Vouchers and Amazon Gift Cards resulted from the number of Raven's Progressive Matrices answered correctly, where one puzzle answered correctly yielded one Airtime Voucher or one Amazon Gift Card. Airtime Vouchers were worth 50 KSh each, while Amazon Gift Cards were worth 1 dollar each. Happiness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being sad and 7 being happy. Alertness refers to the number chosen from a 1-7 scale that asked how one felt towards the end of the experiment, with 1 being tired and 7 being alert. "Discomfort" is an indicator for whether the participant indicated another choice aside from "comfortable" in the post-experiment survey. "Temperature salience" is an indicator for whether the participant indicated temperature in the post-experiment survey as something the experiment was studying. \*  $p < .05$ , \*\*  $p < .01$

## References

- Africanews (2018, February). Protests hit Nairobi slum after arrest of opposition official. [http://www.africanews.com/2018/02/02/protest-hit-nairobi-slum-after-arrest-of-opposition-official//.](http://www.africanews.com/2018/02/02/protest-hit-nairobi-slum-after-arrest-of-opposition-official/)
- Al Jazeera (2017, August). Protests over election fraud claim turn deadly in Kenya. [https://www.aljazeera.com/news/2017/08/kenya-police-protesters-clash-poll-fraud-claim-170809081850902.html.](https://www.aljazeera.com/news/2017/08/kenya-police-protesters-clash-poll-fraud-claim-170809081850902.html)
- Awich, L. (2017, July). Pollsters differ on Uhuru, Raila support. [https://www.the-star.co.ke/news/2017/07/24/pollsters-differ-on-uhuru-raila-support\\_c1602604.](https://www.the-star.co.ke/news/2017/07/24/pollsters-differ-on-uhuru-raila-support_c1602604)
- Barkan, J. D. (1994). *Beyond Capitalism vs. Socialism in Kenya & Tanzania*. Lynne Rienner.
- BBC News (2017a, July). Kenya deputy president ruto's home entered by knifeman. [https://www.bbc.com/news/world-africa-40764446.](https://www.bbc.com/news/world-africa-40764446)
- BBC News (2017b, October). Kenya election: Turnout under 34% amid opposition boycott. [https://www.bbc.com/news/world-africa-41773919.](https://www.bbc.com/news/world-africa-41773919)
- Blomfield, A. (2017, October). Kenya calls off repeat presidential rerun in restive western regions amid fears of 'pogroms'. [https://www.telegraph.co.uk/news/2017/10/27/kenya-calls-second-attempt-re-run-presidential-vote-amid-fears/.](https://www.telegraph.co.uk/news/2017/10/27/kenya-calls-second-attempt-re-run-presidential-vote-amid-fears/)
- Burke, J. (2017a, November). Kenya court upholds President Kenyatta's election victory. [https://www.theguardian.com/world/2017/nov/20/kenya-court-upholds-president-kenyatta-election-victory.](https://www.theguardian.com/world/2017/nov/20/kenya-court-upholds-president-kenyatta-election-victory)
- Burke, J. (2017b, July). Kenyan election official 'tortured and murdered' as fears of violence grow. [https://www.theguardian.com/world/2017/jul/31/kenyan-election-official-christopher-msando-dead-before-national-vote.](https://www.theguardian.com/world/2017/jul/31/kenyan-election-official-christopher-msando-dead-before-national-vote)
- Choi, D. D., L. R. Arriola, and V. Rateng (2017, August). The key to Kenya's close election: the power of partisanship. [https://www.washingtonpost.com/news/monkey-cage/wp/2017/08/07/heres-how-partisanship-not-the-economy-is-shaping-kenyas-close-election/?utm\\_term=.a04a71788ed3.](https://www.washingtonpost.com/news/monkey-cage/wp/2017/08/07/heres-how-partisanship-not-the-economy-is-shaping-kenyas-close-election/?utm_term=.a04a71788ed3)
- Freedom House (2018a). Freedom in the World 2018: Kenya Profile. [https://freedomhouse.org/report/freedom-world/2018/kenya.](https://freedomhouse.org/report/freedom-world/2018/kenya)
- Freedom House (2018b). Freedom in the World 2018: United States Profile. [https://freedomhouse.org/report/freedom-world/2018/united-states.](https://freedomhouse.org/report/freedom-world/2018/united-states)
- Heim, J., E. Silverman, T. R. Shapiro, and E. Brown (2017). One dead as car strikes crowds amid protests of white nationalist gathering in Charlottesville; two police die in helicopter crash. [https://www.washingtonpost.com/local/fights-in-advance-of-saturday-protest-in-charlottesville/2017/08/12/155fb636-7f13-11e7-83c7-5bd5460f0d7e\\_story.html?utm\\_term=.bd12eb2b41fb.](https://www.washingtonpost.com/local/fights-in-advance-of-saturday-protest-in-charlottesville/2017/08/12/155fb636-7f13-11e7-83c7-5bd5460f0d7e_story.html?utm_term=.bd12eb2b41fb)
- Human Rights Watch (2017, August). Kenya: Post-Election Killings, Abuse. [https://www.hrw.org/news/2017/08/27/kenya-post-election-killings-abuse.](https://www.hrw.org/news/2017/08/27/kenya-post-election-killings-abuse)
- Jorgic, D. and H. Malalo (2013, March). Kenya's Odinga challenges election defeat in top court. [https://www.reuters.com/article/us-kenya-elections-kenyas-odinga-challenges-election-defeat-in-top-court-idUSBRE92F03U20130316.](https://www.reuters.com/article/us-kenya-elections-kenyas-odinga-challenges-election-defeat-in-top-court-idUSBRE92F03U20130316)

- Keneally, M. (2017, September). DACA announcement sparks protests nationwide, dozens arrested at Trump Tower. <https://abcnews.go.com/Politics/arrests-made-daca-protest-york/story?id=49625957>.
- Mebane, Jr., W. R. (2017). Anomalies and frauds(?) in the kenya 2017 presidential election. Working paper.
- Muthoni, K. and P. Ogembu (2017, August). Presidential petition: What scrutiny of key IEBC forms revealed. <https://www.standardmedia.co.ke/article/2001253102/what-scrutiny-of-key-iebc-forms-revealed>.
- Obulutsa, G. (2017, October). Kenya bans city-center protests as vote tension mounts. <https://www.reuters.com/article/us-kenya-election/kenya-bans-city-center-protests-as-vote-tension-mounts-idUSKBN1CH173>.
- Okuoro, S. (2017, August). Supreme Court allows NASA to audit IEBC election materials. <https://www.standardmedia.co.ke/article/2001252815/supreme-court-allows-nasa-to-audit-iebc-election-materials>.
- Onyango, P. (2017, October). Wafula Chebukati sets conditions for him to remain as IEBC chair. <https://www.standardmedia.co.ke/article/2001257758/wafula-chebukati-sets-conditions-for-him-to-remain-as-iebc-chair>.
- The New York Times (2018, January). <https://www.nytimes.com/2018/01/20/us/womens-march.html>.  
<https://www.nytimes.com/2018/01/20/us/womens-march.html>.