IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17870

Heat and Team Production: Experimental Evidence from Bangladesh

Teevrat Garg
Maulik Jagnani
Elizabeth Lyons

APRIL 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17870

Heat and Team Production: Experimental Evidence from Bangladesh

Teevrat Garg
University of California, San Diego and IZA

Maulik Jagnani
Tufts University

Elizabeth Lyons
University of California, San Diego

APRIL 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Heat and Team Production: Experimental Evidence from Bangladesh*

Heat's impact on economic growth and aggregate productivity is well-established, but while individual impairments are well-understood as mechanisms, the role of interpersonal dynamics remain unexplored despite the growing prevalence of team-based occupations. In our experiment, programmers were randomly assigned to work individually or in pairs under warm (29°C) or control (24°C) conditions. We found that heat had no effect on individual performance but impaired team performance—not through decreased effort but likely through impaired collaboration. This negative impact was strongest in heterogeneous teams, suggesting heat exacerbates coordination challenges, confirmed by poorer partner evaluations in warm conditions.

JEL Classification: J24, Q54, Q56

Keywords: team production, heat stress, labor productivity

Corresponding author:
Teevrat Garg
University of California
San Diego
La Jolla
CA 92093
USA

E-mail: teevrat@ucsd.edu

* The authors gratefully acknowledge excellent implementation assistance from ARCED Foundation, and in particular from Mehrab Ali, Sadia Chowdhury, Pritha Tasmin, and Ashraf Uddin Mian, and research assistance from Sean Walsh, and for funding support provided by the Academic Senate, Center for Global Transformation, and the Policy Design and Evaluation Lab all at University of California San Diego, and University of Colorado Denver. We are grateful to Jacob LaRiviere, Craig McIntosh, Ariel Ortiz-Bobea, Julie Schaffner, Steve Block, and participants at numerous conferences and seminars for valuable feedback.

1 Introduction

The impact of heat on economic growth and aggregate productivity is well-documented (Burke, Hsiang and Miguel, 2015; Dell, Jones and Olken, 2012; Heal and Park, 2016). However, we know relatively little about the mechanisms of impact. In particular, while prior research has primarily emphasized individual cognitive and physical impairments as drivers of these productivity losses,[1] the role of interpersonal dynamics remains largely unexplored.

On the one hand, studies show that heat increases aggression and impatience (Almas et al., 2020; Carias et al., 2024), potentially hindering interpersonal interaction.[2] On the other hand, heat-induced productivity losses via individual cognitive or physical impairments could be offset by other team members through, for example, sharing tasks, identifying mistakes, or offering encouragement.

Given the increasing importance of occupations that rely on team-based production—such as managers, programmers, physicians, lawyers, and economists—it is crucial to understand whether heat disrupts coordination and communication within teams in ways that differ from its effects on individual performance (Deming, 2017). Furthermore, if the proportion of team-based occupations continues to rise, and heat affects teams differently than individuals, projections about worker output under various climate change scenarios must consider this distinction (Rode et al., 2022).

However, there are no causal estimates of heat's impact on team production versus individual production, likely due to estimation challenges. First, existing data used to estimate the heat-productivity relationship rarely specify whether tasks involved interpersonal interaction or were performed individually. Second, the possibility that workers may select into teams based on the day's temperature can complicate the causal identification of heat's effect

[1]A growing body of work has shown that heat stress reduces productivity of individual workers across many occupational roles (see Lai et al., 2023, for a review).

[2]Numerous studies in psychology and economics, utilizing both experimental and quasi-experimental methods, have found that higher temperatures increase the likelihood of individuals exhibiting aggressive or violent behavior towards others. See Hsiang, Burke and Miguel, 2013 and Burke, Hsiang and Miguel, 2015, for a review.

on team relative to individual productivity. This includes its potential influence on whether workers choose to collaborate and whom they choose to work with to produce output.

This paper addresses these estimation challenges via a field experiment, presenting the first causal estimates of heat's impact on team production versus individual production. We conduct a randomized control trial in Dhaka, Bangladesh, hiring computer programmers to work on a series of incentivized coding tasks within 4 hours. Workers are randomized into two cross-cutting treatments: (i) warm rooms (indoor dry bulb temperature of 29°C) vs. control rooms (24°C), and (ii) individual programming vs. pair programming, where two programmers collaborated side by side at one computer.[3]

We report three principal findings. First, workers randomly assigned to individual coding in warm rooms perform no worse than those in control rooms. However, teams in warm rooms perform significantly worse than both individuals in warm rooms and teams in control rooms. Teams in control rooms were more than twice as likely to make any progress on the task compared to those in warm rooms. Second, we fail to find evidence that this decline in team performance is due to reduced effort: we find no effect of heat on minutes worked or on characters typed per minute. Third, the negative effects of heat on teams in warm rooms are more pronounced for mixed-gender teams and teams with differences in academic seniority. This suggests that heat exacerbates coordination barriers in heterogeneous teams. Our post-experiment survey supports these findings. Teams in heterogeneous pairs in warm rooms report lower partner assessments and a desire to change partners for future tasks.

This paper contributes to a growing literature in environmental economics on the effects of heat on worker productivity. Previous studies have shown that heat stress reduces productivity in both low- and high-skill workers across lower- and higher-income settings (see Lai et al., 2023, for a review). However, we are the first to separately identify the causal effects

[3]Pair programming is a common collaborative coding method where two programmers work together at one computer. The roles within the pair are split into 'driver' and 'navigator'. The driver handles the actual typing of the code, while the navigator acts as an active observer, monitoring the code that's written. Together, they engage in all aspects of software development, including design, coding, and debugging, maintaining constant communication and often switching roles during the session (Chong et al., 2005).

of heat on team production versus individual production, highlighting the role of interpersonal dynamics as a key transmission channel for temperature-related aggregate productivity losses. In line with a global meta-analysis (Hancock, Ross and Szalma, 2007), the best prior evidence from South Asia indicates that the negative effects of heat on worker productivity begin at either 29.5°C or 33°C (Adhvaryu, Kala and Nyshadham, 2020; Somanathan et al., 2021).[4] Similarly, using data from over 9,000 Demographic and Health Surveys interviewers in 46 developing countries, LoPalo (2023) finds that the negative effects of heat on worker productivity begin when temperatures exceed a dry (wet) bulb temperature of 35°C (29°C). We show that while heat may not adversely affect individual production below these thresholds, it can negatively impact team production at lower temperatures. Our findings suggest these effects stem not from reduced individual effort or efficiency, but rather from heat's impact on interpersonal dynamics essential to collaborative cognitive tasks.

Our findings are also crucial due to the growing importance of team production (Jones, 2009; Wuchty, Jones and Uzzi, 2007). For instance, in 2017, 78 percent of U.S. employment was in occupations where group work was reported to be either "very" or "extremely" important (O*NET, 2020). Furthermore, cognitive occupations that rely on interpersonal interaction are increasingly occurring in regions with high temperatures but with extremely limited deployment of commercial climate control technologies. There has been a significant shift in the geographic distribution of these jobs from being concentrated in North America and Europe to increasingly being located in regions like South Asia and Southeast Asia, and to a lesser extent, Africa (Blom, Lan and Adil, 2014; Thursby and Thursby, 2006). However, the demand for commercial-scale air-conditioning units in both Africa and South Asia in 2018 was about 5% of that in the United States; the corresponding number is slightly higher in Southeast Asia at about 10% (JRAIA, 2019).

[4] Two other meta-analyses also support these findings: Seppänen, Fisk and Lei-Gomez (2006) reports a 2% drop in office productivity (e.g., text processing, simple calculations, and customer call handling times) when temperatures exceed a wet bulb temperature of 25°C. Similarly, Hsiang (2010) finds that task efficiency decreases by roughly 1%-2% for each degree above this threshold. Notably, a 25°C wet bulb temperature equates to an ambient dry bulb temperature of 31°C at 65% humidity.

Lastly, our study contributes to a large organizational behavior, psychology, and economics literature by exploring how team heterogeneity affects performance. Research spanning 50 years in organizational behavior and psychology suggests that social differences – such as race, gender, or age – typically undermine group functionality, often due to challenges in social integration, communication, and increased conflict (see Mannix and Neale, 2005 for a review). Similarly, in economics, studies have shown that age, nationality, and ethnic diversity tend to negatively impact team productivity (e.g., Hjort, 2014; Lyons, 2017; see Hansen, Owan and Pan, 2006 for a review). For example, investigations into age diversity have revealed a negative correlation with the frequency of technical communication within teams. Furthermore, gender diversity has been associated with more conflicts, decreased performance in team-based cognitive tasks, and a reduction in creative outputs. Our results complement this literature by suggesting a novel mediator for these relationships: heat stress. Specifically, our findings indicate that heat exacerbates coordination barriers in heterogeneous teams.

2 Experimental Design

To assess how heat affects team versus individual production, we organized a field experiment with undergraduate computer science students in Dhaka, Bangladesh. We randomly assigned participating students to either teams of two or to work independently. We also randomized their working conditions to either a control environment (24°C) or a warmer one (29°C). Crucially, participants were unaware that the study focused on the impact of heat on performance. On each day of the study, all participants either worked in pairs or individually, ensuring they also did not know about the variation in team composition. This determination was not revealed to participants until they showed up for the study. This experimental design allowed us to specifically examine the effects of heat on individual and team production and to compare these impacts.

2.1 Experimental Setting

Context and Participants. We conducted the experiment in Dhaka, Bangladesh, at the offices of our implementation partner, the Aureolin Research, Consultancy, and Expertise Development (ARCED) Foundation.

The study spanned 25 days over eight weeks in October and November 2022, coinciding with the onset of winter. Despite being the cooler season, Dhaka maintains high temperatures; during the study period, the average daily maximum was 31.2°C, the minimum was 23.4°C, and the overall average was 27.3°C—higher than the annual averages of 30.8°C, 22°C, and 26°C, respectively.

We selected our sample from the population of internship-eligible second, third, and fourth-year computer science students at major universities in Dhaka. The Information Technology (IT) sector in Bangladesh is a significant contributor to the economy, estimated to be worth $1.4 billion in FY 2022 and supporting about 300,000 jobs (Abdullah, 2023). Thus, assessing IT worker productivity as temperatures rise is crucial to understanding the country's ongoing economic development. Additionally, the large, highly skilled pool of IT students, driven by substantial career opportunities post-graduation, makes this group particularly relevant for our study.

Participant Recruitment. To recruit participants for the study, ARCED advertised in local university computer science departments using flyers and emails sent to faculty and staff. Interested students were invited to sign up by completing a survey accessed through a QR code or a URL. The survey collected their names, phone numbers, and emails, which ARCED used to schedule their sessions. Additionally, it asked for their university affiliation, credit hours accumulated, semesters completed, and previous experience with Java, the programming language used in the task.

After the sign-up period, we randomized the order in which prospective participants were contacted for scheduling. Three days before each task day, we called students in this randomized order to confirm their availability.[5] We continued calling until we confirmed twenty available participants for each session, then proceeded to schedule the next batch of participants for the following test day. Despite these arrangements, there were consistently multiple participants who failed to show up on the day of the task; the average session included 6 participants, balanced across our treatment arms as discussed below.

Task Description. In collaboration with computer science graduate students in the US and Bangladesh, we developed a Java-based programming task for our study participants. The task required participants to enhance an existing script by adding five features to implement an application programming interface (API). Given Java's object-oriented nature, even programmers without specific Java experience but familiar with object-oriented programming could successfully complete the task within the allotted four hours.[6] Detailed task instructions are provided in Appendix Section B. Each feature was designed to be added independently and in any order, allowing participants to complete between zero and five features.

Participants received a flat fee of 1100 BDT (approximately 13 USD) for attending a session, with the opportunity to earn up to an additional 45% of this base fee in bonuses. Specifically, for each successfully implemented feature, participants received a bonus of 9% of the base pay.[7]

The design of this task serves two important research purposes. Firstly, it provides a clear performance-based outcome—namely, the number of features successfully added. Secondly, our interface enables the measurement of inputs, such as the number of characters typed and work hours, which is significant given the hypothesized effects of heat on worker effort.

[5]A three-day scheduling window was chosen based on pilot studies indicating that participants' schedules frequently change and commitments beyond three days are often forgotten.

[6]The task was designed by a computer scientist familiar with our participants' expected programming abilities. We validated the task through pilot testing with participants of similar skill levels to our target sample.

[7]Participants were told that the could exit the task at any time without the loss of the fixed payment.

2.2 Experimental Treatments

We implemented two experimental treatments in our study, each randomized differently: one at the day level and the other at the room level.

Before the study began, we randomized whether tasks would be completed in teams or individually each day. This was supported by assigning participants to study days based on the randomized order in which they were called, ensuring random allocation to either team or individual work conditions.[8] Figure A.1 illustrates that the room layouts for team and independent work days are similar, ensuring that layout variations do not affect performance differences between team and independent workers.

On each session day, we randomized which of the two identically laid-out rooms would be set to warm (29°C) or control (24°C) temperatures, and participants were also randomly assigned to one of these rooms.[9]

Lastly, we randomly assigned participants to specific seat numbers in each room. Therefore, on team days, two participants were randomly paired at each assigned seat, ensuring that team formations were also randomized.

Temperature Treatment. We utilized air conditioning to maintain specific temperatures in both the warm and control rooms, setting the warmer room at 29°C and the control room at 24°C. Notably, 90% of participants began the experiment between 12 pm and 3 pm—the day's hottest segment—while only 10% started between 9 am and 12 pm. Consequently, the outdoor temperature upon entry was close to the average daily maximum of 31.2°C, ranging from 27.7 to 34.1°C, necessitating air conditioning for both environments.

Our temperature settings were chosen to ensure a significant differential between the two rooms while remaining within ethically acceptable and commonly recommended limits. The

[8]Participants knew they were part of a study but were not informed about the specific treatment arms.

[9]Both rooms were equipped with two ceiling fans, which remained operational throughout the experiment. Ceiling fans are commonly used in Bangladeshi households and workplaces. Unlike air conditioners, which actually lower the room temperature, ceiling fans create a 'wind chill effect' by circulating air. This breeze makes individuals feel cooler without changing the air temperature. It's important to note that ceiling fans are designed to cool people, not rooms.

control room temperature of 24°C aligns with recommendations from the US Occupational Safety and Health Administration (OSHA), which advises employers to maintain office temperatures between 20°C and 24.4°C. Similarly, the warm room temperature of 29°C conforms to guidelines from a US federal program that recommends keeping home temperatures between 25.5°C and 29.4°C.[10] Temperature levels were regulated by built-in monitors in the air conditioners, and additional temperature monitors verified that each room maintained the intended temperature.

Team Treatment. We assigned study participants to work settings either individually or in teams of two as pair programmers. Pair programming, a prevalent industry practice particularly among entry-level programmers, involves two programmers working together at a single computer. In this setup, one person acts as the 'driver,' typing the code, while the other serves as the 'navigator,' who provides the overall structure of the code and reviews the driver's output. A meta-analysis of experimental studies on the effectiveness of pair programming indicates that it is particularly beneficial for achieving accuracy in highly complex programming tasks and also offers a time advantage on simpler tasks. By collaborating, programmers complete tasks and achieve goals that might be challenging or impossible to accomplish individually (Hannay et al., 2009). In our experiment, we advised teams to adhere to industry norms by having drivers and navigators switch roles every half hour.

2.3 Analysis Plan

To estimate the impact of heat on team production relative to individual production, we estimate the following equation:

$$Y_{ird} = \beta_0 + \beta_1 Warm_r + \beta_2 Team_d + \beta_3 Warm_r \cdot Team_d + \epsilon_{rd}$$

[10] `https://www.cnn.com/2019/08/20/health/thermostat-recommendations-energy-star-trnd/index.html`

where Y_{ird} represents the outcome of interest for participant or team i in room r on day d. $Warm_r$ is a binary indicator that equals one if the participant was assigned to the warm room (29°C), and $Team_d$ is a binary indicator that equals one if the participant was part of a team of two rather than working individually. We cluster standard errors at the room-day level to account for randomization at this level.

β_1 quantifies the effect of being in a warm room on individual outcomes, indicating whether warmer conditions (relative to control conditions at 24°C) lead to an increase or decrease in performance for individuals. β_2 measures the impact of working in a team compared to working individually, isolating the influence of collaboration on performance in control conditions. β_3 captures how warm conditions affect team performance compared to individual performance, highlighting variations in the heat's impact depending on whether participants are working in teams or alone.

Our primary outcomes of interest are first, whether participants successfully added at least one feature to the code, and second, the total number of features participants added. We favor using indicator variables that capture whether participants successfully added any features over the continuous measure of the total number of features added. This preference is due to the distribution where nearly half of the participants added no features, and only four managed to add all five (Figure A.2). A linear probability model is well-suited for estimating our treatment effects given this distribution. For completeness, we also present estimated treatment effects for adding two, three, four, and five features to the code.

To explore the mechanisms by which heat influences worker output, we examine treatment effects on various measures of input or effort. These include the total characters typed, the total time participants spent in the experiment and actively engaged at the computer,[11] and whether they took any breaks.[12] Additionally, we assess impacts on subjective survey-based

[11]Total minutes active in task measures the duration from the first to the last recorded action. Total minutes in the experiment measures the duration from check-in to check-out (including time spent on post-experiment survey), as recorded by our implementation staff.

[12]A break is defined as any instance where a participant temporarily left the experimental room (e.g., to use the restroom or get water). Participants were instructed to raise their hand when they needed a break, allowing field staff to record each occurrence.

measures that capture participants' mood after the task, their experience with the task, and their assessment of their partner.

Sample. Our sample consisted of 232 individuals, with 98 assigned to work in pairs and 134 working individually, yielding 183 data points for output and input measures. Post-experiment surveys were conducted individually with all 232 participants.

Aligned with our recruitment strategy, which targeted internship-eligible second, third, and fourth-year computer, science students, our participants were notably experienced: they averaged 100 credits and were in their 9th semester.[13] Given that the task required completing a Java-based programming assignment, nearly all participants were proficient in Java. Additionally, 20% of our sample was female.

All participants spent at least 84 minutes in the experiment and 26 minutes actively working on the task. The average person spent over three and a half hours in the experiment and roughly two hours actively working. Even participants unsuccessful in adding any features spent an average of 110 minutes actively working during the 180-minutes spent in the experiment. This substantial time investment across all participants, despite approximately half of teams not completing any features, reflects the intentional accessibility yet challenging nature of our programming assignment.[14]

First Stage. We successfully maintained the designated temperatures in our study settings, with the control room at 24°C and the warm room 5°C higher at 29°C (Table 1, Panel A, Column 1). Furthermore, these temperature conditions were independent of the participants' grouping, ensuring that both individual and team tasks were conducted under

[13] Computer science at major universities in Dhaka is structured as a 4-year degree program, divided into 3 semesters per year, totaling 12 semesters.

[14] Analysis of testing behavior reveals that participants who successfully added features ran significantly more code functionality tests (36 on average) compared to those who completed none (9 on average). This suggests that successful participants may have employed more iterative coding approaches, avoiding the 'fixation' on unsuccessful solution paths that can hinder problem-solving (e.g. Smith and Blankenship, 1991). However, our experiment was not designed to test the reasons for success or failure beyond our treatments, so we cannot say exactly why almost half the teams did not complete any features.

identical thermal conditions in their respective environments.

Humidity. Indoor humidity levels were somewhat higher in the warmer rooms, regardless of the team treatment configuration: 65% RH versus 72% RH (Table 1, Panel A, Column 2). This rise in humidity is due to warmer air's capacity to hold more moisture,[15] and is amplified by the dehumidifying effect of air conditioners.[16] As such, the treatment effects observed in our study should be interpreted as the combined impact of both higher temperatures and increased humidity.[17] It is important to note, however, that experimental results suggest that relative humidity significantly impacts human physiological responses only when the temperature exceeds 30°C (Li et al., 2018), which is higher than the temperatures in our warm room.

Balance. To validate our randomization process, we demonstrate that indoor date-room characteristics such as room air quality and the number of participants per room show no correlation with treatment assignments (Table 1, Panel A, Columns 3-5).[18] Treatment distribution was consistent over the 8-week duration of the study and across different days of the week (Table 1, Panel A, Columns 6-7). Furthermore, external conditions such as outdoor temperature, humidity, and precipitation were evenly balanced across treatment groups, underscoring the effectiveness of our randomization in controlling for these variables (Table 1, Panel B, Columns 1-5). Finally, treatments are balanced across several participant characteristics: the number of college credits accumulated, current semester of study, prior knowledge of Java, or gender (Table 1, Panel C, Columns 1-4).

[15] `https://www.weather.gov/lmk/humidity`

[16] `https://www.epa.gov/mold/what-are-main-ways-control-moisture-your-home`

[17] While average humidity generally rises with temperature, in drier climates than Dhaka, the relationship between humidity and temperature may be less distinct than what was observed in our setting.

[18] Indoor PM2.5 and PM10 levels during our study were extremely poor, exceeding 60 $\mu g/m^3$. On average, there were 6 participants in each room.

3 Effect of Heat on Team Production Relative to Individual Production

Table 2, Panel A presents the estimated effects of heat on the output of independent workers and teams. 'Any Feature (0/1)' is our main outcome of interest, indicating whether at least one feature was added to the script.[19] We report three findings:

First, we find no negative impact of heat on the performance of independent workers. If anything, individuals in the warm room tend to perform slightly better than their counterparts in the control room. This is not unusual in that, as previously discussed, almost all prior work finds the negative effects of heat on productivity at thresholds well past the 29°C which is the temperature in our warm rooms. Second, teams are nearly twice as likely as independent workers to add any features to the script, consistent with the aforementioned meta-analysis of experimental studies on the effectiveness of pair programming. Third, teams perform significantly worse in warmer rooms, with high temperatures negating the performance advantage of team production. Specifically, the likelihood that teams will add any features in warmer rooms drops by 44 percentage points, making teams in control rooms more than twice as likely to add features compared to those in warm rooms. Importantly, we can statistically reject the hypothesis that the impact of heat on individual production is similar to its impact on team production (p-value < 0.05).

These treatment effects persist when the outcome variable reflects whether at least two features—the sample median—were not added (Table A.1). However, we observed no significant effects of room temperature regardless of team/individual structure on the addition of a higher number of features. This pattern we observe on the addition of at least one or two features persists when considering total features completed as a count variable, albeit

[19] As we note in Online Appendix C, we pre-specified each of the outcome variables we report in the paper with two exceptions. The main modification is in Table 2, Panel A where we introduce "Any Feature Completed (0/1)" as a transformation of the pre-specified variable "Total Features Completed". Similarly, in Table 2, Panel B, we use "Any Break (0/1)" because breaks follow a similar distribution. These additions were made ex-post given the realized distribution of these outcome variables. Using the pre-specified variables provides qualitatively similar but understandably noisier estimates.

statistically imprecisely.

4 Mechanisms

The absence of adverse effects of heat on individual productivity suggests that heat may be reducing team performance through mechanisms other than the directly affecting individual performance, for example, by increasing coordination costs or the likelihood of coordination failure. Specifically, heat can raise the relative cost of effort in teams by making coordination more challenging. High temperatures can slow down communication and response times, impair decision-making (Carias et al., 2024), and lead to delays or increased errors. This breakdown in coordination necessitates more time and effort to achieve the same output levels, thus raising the overall cost of coordination. Moreover, heat might diminish the likelihood that a given amount of effort leads to feature completion by complicating the realization of complementarities. For example, high temperatures can impair cognitive functions and affect mood, such as increasing aggression (Almas et al., 2020), which reduces the effectiveness of brainstorming and collaborative problem-solving. This hinders the ability of teams to synergize their diverse skills, ultimately impacting team productivity.

In this section, we explore the mechanisms underlying the adverse effect of heat on team production. We do this in two ways. First, we examine data on inputs and survey data. Second we consider heterogeneity in team composition.

4.1 Input and Survey Data

Inputs. We test whether teams in warm rooms exert different levels of effort compared to individuals or teams in control rooms by analyzing task inputs (Table 2, Panel B). We find no significant difference in total characters typed (Column 1) or per-minute number of characters typed (Column 2), nor in the total minutes in the experiment (Column 3) or active on the task (Column 4) by teams in warm rooms.[20] However, teams in warm rooms were significantly more likely to take breaks during the work session than those in control rooms, which may suggest costly adaptation strategies for coping with the heat (Column 5).[21] Overall, this evidence indicates that at least some of the productivity decline amongst teams in warm rooms can be explained by teams taking more breaks.

Survey-Based Outcomes. While we do not directly observe coordination failures, we utilized responses from our post-experiment survey to evaluate whether heat affected workers' experiences during the task. As presented in Table A.2, teams in warm rooms did not perceive the task as more difficult or less engaging.[22] Furthermore, these teams were no more likely to prefer teamwork or exhibit significantly different levels of positive or negative affect than those in control rooms.[23] These null effects suggest that respondents in warm rooms—-whether working individually or in teams—did not find the tasks markedly harder. Additionally, we found no evidence that teams in warm rooms had worse opinions of their teammates compared to those in control rooms.[24]

In the next section, we will explore whether team composition heterogeneity masks the

[20]These are imperfect measures of effort. For instance, an efficient worker may engage in careful planning before writing code, resulting in fewer total characters typed while still achieving the desired output efficiently. The lack of any discernible differences across individuals and teams suggests that physical inputs were no different but does not characterize all effort, for example, precision of coding which could be cognitively demanding.

[21]For instance, through a field experiment in Indonesia, Masuda et al. (2021) show that worker productivity was 8% lower in deforested relative to forested settings, where wet bulb globe temperatures were 3°C higher. They demonstrate that productivity losses are driven by behavioral adaptations in the form of increased number of work breaks, and provide evidence that suggests breaks are in part driven by awareness of heat effects on work.

[22]Workers in teams were asked to rate the task as difficult or engaging on a scale of 1-10. In the analysis presented, we convert their score into a binary variable equal to 1 if the score was greater than 5, and 0 otherwise. This transformation was made to capture a general "favorable" versus "unfavorable" impression, although our results are qualitatively similar when using a continuous measure of difficulty or engagement.

[23]To capture preference for teamwork, we asked participants if they would prefer to work with a partner for a similar task in the future—a simple yes or no question. Negative affects include anger, anxiety, grumpiness, irritation, and worry. Positive affects include energy and happiness.

[24]Workers in teams were asked to rate their partners on a scale of 1-10. In the analysis presented, we convert their score into a binary variable equal to 1 if the score was greater than 5, and 0 otherwise. This transformation was made to capture a general "favorable" versus "unfavorable" impression, although our results are qualitatively similar when using a continuous measure of partner score. For our other outcome measure for partner assessment, we asked team members if they would prefer a different partner for a similar task in the future—a simple yes or no question.

null effects observed in these survey-based measures.

4.2 Heterogeneity in Team Composition

Our findings reveal that higher temperatures (29°C) reduce team output but not individual output. This average impact may be at least partially explained by teams in warm rooms taking more breaks. Furthermore, these effects are not reflected in post-experiment survey responses regarding the task or perceptions of teammates. We now explore whether heat has a differential effect on teams that differ on observable dimensions. We consider two sources of heterogeneity — gender and academic seniority.[25]

Output. In Table 3, Panel A, we focus our analysis on team production, specifically examining the heterogeneous effects of heat on mixed-gender teams. We find that heat disproportionately affects teams with members of opposite genders. Specifically, mixed-gender teams in warm rooms are 55% less likely to complete any feature compared to same-gender teams in cool rooms (Column 1). Further analysis in Table 4, Panel A, reports the differential effects of heat on teams with varying levels of academic seniority. We define academic seniority based on the current semester of study, identifying teams as having dissimilar experience if there is an above median gap in academic seniority (two semesters) between team members. While the effects are economically significant, they are also somewhat imprecise. Nonetheless, the evidence consistently suggests that heterogeneous teams suffer more from reduced output under heat stress.

In Table 5, Panel A, we demonstrate that teams with the greatest differences among members experience the most significant negative impacts of heat stress on output. Specifically, teams composed of members who differ in both gender and academic seniority exhibit substantial declines in productivity.[26]

[25] As we discuss in Online Appendix C, we did not pre-specify these heterogeneity tests.

[26] Figure A.3 illustrates the distribution of teams that are homogeneous in gender and experience (dissimilar = 0), differ only in experience (dissimilar semester), differ only in gender (dissimilar gender), or differ in both (dissimilar = 2).

Importantly, there is no discernible difference in how heat affects individual output based on gender or academic seniority (Table A.3). This suggests that the poorer performance of heterogeneous teams compared to homogeneous ones under heat stress is not linked directly to the individual characteristics of team members. Instead, it may be related to how these characteristics impact team dynamics and interactions when under heat stress.

Inputs. Given that the effects of heat on teams are particularly pronounced among heterogeneous teams, randomly determined by differences such as gender or academic seniority, we examined whether task inputs and post-experiment survey responses vary significantly among these groups. Table 3, Panel B provides some evidence that heterogeneous teams exert less effort than homogeneous teams. Specifically, mixed-gender teams type significantly fewer characters per minute. However, the effort of teams with dissimilar academic seniority remains unaffected by heat (Table 4, Panel B).

Interestingly, mixed-gender teams are less likely to take breaks than same-gender teams under heat stress. Similarly, teams with dissimilar academic seniority also show a reduced likelihood of taking breaks, though this finding is less statistically certain. Furthermore, for team members who differ on both dimensions, being in a warm room decreases the likelihood of taking breaks by 58 percentage points, although this estimate is also not statistically significant (Table 5, Panel B). These findings suggest that the previously discussed average effects on breaks were, in fact, productivity-enhancing and did not account for the reduction in output. That is, while homogeneous teams take breaks to mitigate the impacts of heat on team production, coordination failures in heterogeneous teams hinder such mitigative break-taking.

Survey-Based Outcomes. Importantly, Table 3, Panel C shows that gender-diverse team members in warm rooms are 37 percentage points less likely to rate their partners favorably and 31 percentage points more likely to prefer a different partner. Similarly, workers in teams with experience gaps are 37 percentage points less likely to view their partners favorably and 36 percentage points more likely to seek different partners (Table 4, Panel C).

These effects are more pronounced in teams that are diverse in both gender and experience (Table 5, Panel C).[27] For team members who differ on both dimensions, being in a warm room leads to a 63 percentage point decrease in the likelihood of favorably rating partners, which represents a 92% reduction from the control mean. Additionally, such team members in warm rooms are 63 percentage points more likely to desire a different teammate, compared to a control mean of 36%.

5 Conclusion

Our results show the negative effects of heat on economic production can manifest at lower temperatures than previously reported, particularly when we consider settings like team production and alternative mechanisms such as coordination failures. While individual productivity remains steady across varying temperatures, teams—especially those with members from different genders or academic seniority—face greater challenges in warmer conditions. Homogeneous teams appear better able to adapt their workflow in hotter settings, thus minimizing productivity losses. However, heat specifically disrupts effective collaboration in heterogeneous teams, leading to decreased productivity and less favorable perceptions among team members.

These findings reveal a striking implication for work allocation policies within firms. Our results show that once ambient temperatures reach 29°C, the productivity advantage of teamwork is effectively eliminated, with per-worker productivity in two-person teams falling to approximately half that of individuals working alone. This suggests organizations might benefit from temperature-responsive work allocation policies—assigning more individual tasks during warm periods and reserving collaborative work for cooler environments or times of

[27]We fail to find consistent evidence that members of heterogeneous teams in warm rooms were more or less likely to perceive the task as difficult or engaging, exhibit different levels of positive or negative affect, or show a preference for teamwork (Tables A.4-A.6).

day. Such adaptive management could be particularly valuable in climate-vulnerable regions or in facilities with limited cooling resources, where targeted temperature control might be prioritized for collaborative spaces.

References

Abdullah, Mamun. 2023. "Bangladesh becoming IT hub of South Asia." *Dhaka Tribune.*

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2020. "The Light and the Heat: Productivity Co-Benefits of Energy-Saving Technology." *The Review of Economics and Statistics*, 102(4): 779–792.

Almas, Ingvild, Maximilian Auffhammer, Tessa Bold, Ian Bolliger, Aluma Dembo, Solomon M. Hsiang, Shuhei Kitamura, Edward Miguel, and Robert Pickmans. 2020. "Destructive Behavior, Judgment, and Economic Decision-Making Under Thermal Stress." *National Bureau of Economic Research Working Paper Series.*

Blom, Andreas, George Lan, and Mariam Adil. 2014. *A decade of development in sub-Saharan African science, technology, engineering and mathematics research.* World Bank Group.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. "Global non-linear effect of temperature on economic production." *Nature*, 527(7577): 235–239.

Carias, Michelle E., David W. Johnston, Rachel Knott, and Rohan Sweeney. 2024. "Temperature's toll on decision-making." *The Economic Journal.*

Chong, Jan, Robert Plummer, Larry J. Leifer, Scott R. Klemmer, Ozgur Eris, and George Toye. 2005. "Pair Programming: When and Why it Works." *Proceedings of the Psychology of Programming Interest Group (PPIG).*

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken. 2012. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century." *American Economic Journal: Macroeconomics*, 4(3): 66–95.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics*, 132(4): 1593–1640.

Hancock, PA, JM Ross, and JL Szalma. 2007. "A meta-analysis of performance response under thermal stressors." *Human Factors*, 49(5): 851–877.

Hannay, Jo E., Tore Dybå, Erik Arisholm, and Dag I.K. Sjøberg. 2009. "The effectiveness of pair programming: A meta-analysis." *Information and Software Technology*, 51(7): 1110–1122.

Hansen, Zeynep, Hideo Owan, and Jie Pan. 2006. "The Impact of Group Diversity on Performance and Knowledge Spillover – An Experiment in a College Classroom." *National Bureau of Economic Research Working Paper Series.*

Heal, Geoffrey, and Jisung Park. 2016. "Temperature Stress and the Direct Impact of Climate Change: A Review of an Emerging Literature." *Review of Environmental Economics and Policy*, 10(2): 347–362.

Hjort, Jonas. 2014. "Ethnic Divisions and Production in Firms." *The Quarterly Journal of Economics*, 129(4): 1899–1946.

Hsiang, Solomon M. 2010. "Temperatures and Cyclones Strongly Associated with Economic Production in the Caribbean and Central America." *Proceedings of the National Academy of Sciences of the United States of America*, 107: 15367–15372.

Hsiang, Solomon M., Marshall Burke, and Edward Miguel. 2013. "Quantifying the Influence of Climate on Human Conflict." *Science*, 341(6151).

Jones, Benjamin F. 2009. "The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?" *The Review of Economic Studies*, 76(1): 283–317.

JRAIA. 2019. "World Air Conditioner Demand by Region." Japan Refrigeration and Air Conditioning Industry Association.

Lai, Wangyang, Yun Qiu, Qu Tang, Chen Xi, and Peng Zhang. 2023. "The Effects of Temperature on Labor Productivity." *Annual Review of Resource Economics*, 15: 213–232.

Li, Baizhan, Chenqiu Du, Meilan Tan, Hong Liu, Emmanuel Essah, and Runming Yao. 2018. "A modified method of evaluating the impact of air humidity on human acceptable air temperatures in hot-humid environments." *Energy and Buildings*, 158: 393–405.

LoPalo, Melissa. 2023. "Temperature, Worker Productivity, and Adaptation: Evidence from Survey Data Production." *American Economic Journal: Applied Economics*, 15(1): 192–229.

Lyons, Elizabeth. 2017. "Team Production in International Labor Markets: Experimental Evidence from the Field." *American Economic Journal: Applied Economics*, 9(3): 70–104.

Mannix, Elizabeth, and Margaret A. Neale. 2005. "What Differences Make a Difference? The Promise and Reality of Diverse Teams in Organizations." *Psychological Science in the Public Interest*, 6(2): 31–55.

Masuda, Yuta J., Teevrat Garg, Ike Anggraeni, Kristie Ebi, Jennifer Krenz, Edward T. Game, Nicholas H. Wolff, and June T. Spector. 2021. "Warming from tropical deforestation reduces worker productivity in rural communities." *Nature Communications*, 12: 1601.

O*NET. 2020. "Work context: work with work group or team."

Rode, Ashwin, Rachel E. Baker, Tamma Carleton, Anthony D'Agostino, Michael Delgado, Timothy Foreman, Diana R. Gergel, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren, Amir Jina, Robert E. Kopp, Steven B. Malevich, Kelly McCusker, Ishan Nath, Matthew Pecenco, James Rising, and Jiacan Yuan. 2022. "Labor Disutility in a Warmer World: The Impact of Climate Change on the Global Workforce." *Mimeo.*

Seppänen, Olli, William J. Fisk, and Quanhong Lei-Gomez. 2006. "Effect of temperature on task performance in office environment." *5th International Conference on Cold Climate Heating, Ventilating and Air Conditioning.*

Smith, Steven M, and Steven E Blankenship. 1991. "Incubation and the persistence of fixation in problem solving." *The American journal of psychology*, 61–87.

Somanathan, E., Rohini Somanathan, Anant Sudarshan, and Meenu Tewari. 2021. "The Impact of Temperature on Productivity and Labor Supply: Evidence from Indian Manufacturing." *Journal of Political Economy*, 129(6): 1797–1827.

Thursby, Marie, and Jerry Thursby. 2006. *Here or There?: A Survey of Factors in Multinational R&D Location – Report to the Government-University-Industry Research Roundtable.* National Academy of Sciences and National Academy of Engineering and Institute of Medicine.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in Production of Knowledge." *Science*, 316: 1036–1039.

Tables and Figures

Table 1: Differences in Indoor Temperature Between Treatment Arms and Balance on Date-Room-Level, Date-Level, and Individual-Level Observables

Panel A: Indoor Date-Room-Level Data							
	(1) Temperature (C) β / SE	(2) Relative Humidity β / SE	(3) PM2.5 ($\mu g/m^3$) β / SE	(4) PM10 ($\mu g/m^3$) β / SE	(5) # Participants β / SE	(6) Week of Year β / SE	(7) Day of Week β / SE
Warm (0/1)	4.98***	7.05***	-6.61	-7.35	0.99	-0.35	-0.18
	(0.18)	(2.00)	(11.76)	(12.42)	(1.66)	(1.06)	(1.06)
Team (0/1)	0.00	2.87	5.61	6.18	0.89	1.00	0.48
	(0.15)	(2.43)	(10.52)	(11.44)	(1.21)	(1.09)	(0.96)
Warm X Team (0/1)	-0.41	-0.70	0.75	0.88	-1.12	0.08	0.46
	(0.33)	(3.75)	(16.28)	(17.26)	(2.12)	(1.60)	(1.55)
Control Mean	24.20	64.86	62.50	65.68	6.15	44.23	3.33
Observations	183	183	183	183	183	183	183
R^2	0.968	0.319	0.024	0.026	0.022	0.045	0.020
Panel B: Outdoor Date-Level Data							
	(1) Maximum Temperature (C) β / SE	(2) Minimum Temperature (C) β / SE	(3) Average Temperature (C) β / SE	(4) Relative Humidity β / SE	(5) Precipitation (mm) β / SE		
Warm (0/1)	0.76	0.64	0.63	-1.75	-9.26		
	(1.03)	(1.43)	(1.11)	(2.62)	(15.10)		
Team (0/1)	-0.53	-1.09	-0.59	-2.12	-14.56		
	(0.90)	(1.41)	(1.01)	(2.61)	(14.05)		
Warm X Team (0/1)	-1.15	-0.20	-0.52	3.84	9.31		
	(1.30)	(2.19)	(1.50)	(3.58)	(15.10)		
Control Mean	31.16	23.47	27.28	67.46	14.60		
Observations	183	183	183	183	183		
R^2	0.075	0.040	0.041	0.022	0.032		
Panel C: Individual-Level Data							
	(1) # Credits β / SE	(2) Current Semester β / SE	(3) Knows Java (0/1) β / SE	(4) Female (0/1) β / SE			
Warm (0/1)	-2.86	-0.07	-0.03	-0.08			
	(5.25)	(0.49)	(0.02)	(0.06)			
Team (0/1)	3.29	0.35	-0.00	0.02			
	(5.08)	(0.39)	(0.00)	(0.07)			
Warm X Team (0/1)	-1.41	-0.39	0.03	0.09			
	(7.80)	(0.70)	(0.02)	(0.11)			
Control Mean	100.72	8.95	1.00	0.20			
Observations	232	232	232	232			
R^2	0.008	0.008	0.025	0.012			

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, team production, and team production in warm room for (i) indoor date-room-level data in Panel A, (ii) outdoor date-level data in Panel B, and (iii) individual-level data in Panel C. Control mean indicates mean values for individual production in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table 2: Impact of Heat on Output and Inputs in Team and Individual Production

Panel A: Output

	(1) Any Feature (0/1) β / SE	(2) Total Features β / SE
Warm (0/1)	0.17*	0.17
	(0.09)	(0.21)
Team (0/1)	0.37***	0.81***
	(0.10)	(0.26)
Warm X Team (0/1)	-0.44**	-0.61
	(0.20)	(0.61)
Control Mean	0.44	1.27
Observations	183	183
R^2	0.066	0.035

Panel B: Inputs

	(1) Total Characters β / SE	(2) Characters per Minute β / SE	(3) Time in Experiment (Minutes) β / SE	(4) Time Active in Task (Minutes) β / SE	(5) Any Break (0/1) β / SE
Warm (0/1)	-2484.62	-16.98	8.43	4.64	0.07
	(2098.78)	(11.28)	(13.59)	(15.52)	(0.11)
Team (0/1)	-269.87	-5.62	17.71	16.79	-0.16**
	(2620.69)	(11.96)	(17.90)	(17.79)	(0.07)
Warm X Team (0/1)	3096.71	19.66	-7.88	-0.93	0.36**
	(3388.45)	(15.76)	(25.92)	(25.92)	(0.16)
Control Mean	8254.47	59.38	202.20	125.91	0.27
Observations	183	183	183	183	183
R^2	0.011	0.014	0.018	0.016	0.064

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, team production, and team production in warm room for (i) output in Panel A and (ii) inputs in Panel B. Control mean indicates mean values for individual production in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table 3: Impact of Heat on Output, Inputs, and Subjective Survey Responses in Same Gender and Mixed Gender Teams

Panel A: Output

	(1) Any Feature (0/1) β / SE	(2) Total Features β / SE
Warm (0/1)	-0.08	0.11
	(0.19)	(0.64)
Different Gender (0/1)	0.09	-0.11
	(0.15)	(0.40)
Warm X Different Gender (0/1)	-0.55**	-1.48**
	(0.20)	(0.69)
Control Mean	0.79	2.11
Observations	49	49
R^2	0.194	0.145

Panel B: Inputs

	(1) Total Characters β / SE	(2) Characters per Minute β / SE	(3) Time in Experiment (Minutes) β / SE	(4) Time Active in Task (Minutes) β / SE	(5) Any Break (0/1) β / SE
Warm (0/1)	2234.81	13.25	2.80	9.43	0.59***
	(3160.73)	(12.84)	(25.82)	(24.99)	(0.10)
Different Gender (0/1)	34.86	-2.03	19.76	25.39	0.20
	(1361.04)	(9.99)	(19.31)	(19.98)	(0.20)
Warm X Different Gender (0/1)	-4468.93	-28.69**	-9.86	-20.43	-0.47*
	(2716.43)	(13.33)	(36.51)	(33.16)	(0.27)
Control Mean	7974.26	54.37	214.05	135.18	0.05
Observations	49	49	49	49	49
R^2	0.051	0.110	0.019	0.030	0.276

Panel C: Survey

	(1) Partner Score > 5 (0/1) β / SE	(2) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	0.12	0.06
	(0.10)	(0.12)
Different Gender (0/1)	0.18	-0.18**
	(0.15)	(0.08)
Warm X Different Gender (0/1)	-0.37*	0.31*
	(0.20)	(0.16)
Control Mean	0.63	0.37
Observations	98	98
R^2	0.034	0.051

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, mixed gender teams, and mix gender teams in warm room for (i) output in Panel A, (ii) inputs in Panel B, and (iii) subjective survey responses in Panel C. Control mean indicates mean values for same gender teams in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table 4: Impact of Heat on Output, Inputs, and Subjective Survey Responses in Similar Academic Seniority and Dissimilar Academic Seniority Teams

Panel A: Output

	(1) Any Feature (0/1) β / SE	(2) Total Features β / SE
Warm (0/1)	-0.15	-0.08
	(0.22)	(0.65)
Dissimilar Semester (0/1)	-0.06	-0.30
	(0.16)	(0.43)
Warm X Dissimilar Semester (0/1)	-0.30	-0.96
	(0.28)	(0.80)
Control Mean	0.85	2.23
Observations	49	49
R^2	0.153	0.109

Panel B: Inputs

	(1) Total Characters β / SE	(2) Characters per Minute β / SE	(3) Time in Experiment (Minutes) β / SE	(4) Time Active in Task (Minutes) β / SE	(5) Any Break (0/1) β / SE
Warm (0/1)	124.69	3.83	-3.26	0.39	0.54***
	(4261.18)	(19.01)	(20.78)	(24.16)	(0.14)
Dissimilar Semester (0/1)	-1984.25	-9.92	19.49	15.12	-0.08
	(2888.49)	(13.69)	(24.76)	(25.34)	(0.10)
Warm X Dissimilar Semester (0/1)	660.65	-5.46	14.53	12.15	-0.28
	(3740.42)	(18.50)	(31.89)	(28.71)	(0.28)
Control Mean	9013.46	58.91	209.80	134.86	0.15
Observations	49	49	49	49	49
R^2	0.020	0.045	0.061	0.044	0.290

Panel C: Survey

	(1) Partner Score > 5 (0/1) β / SE	(2) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	0.15	0.00
	(0.12)	(0.13)
Dissimilar Semester (0/1)	0.06	-0.13
	(0.15)	(0.10)
Warm X Dissimilar Semester (0/1)	-0.37**	0.36**
	(0.17)	(0.17)
Control Mean	0.65	0.38
Observations	98	98
R^2	0.050	0.061

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, dissimilar academic seniority teams, and dissimilar academic seniority teams in warm room for (i) output in Panel A, (ii) inputs in Panel B, and (iii) subjective survey responses in Panel C. Control mean indicates mean values for similar academic seniority teams in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table 5: Impact of Heat on Output, Inputs, and Subjective Survey Responses in Heterogeneous and Homogeneous Teams

Panel A: Output

	(1) Any Feature (0/1) β / SE	(2) Total Features β / SE
Warm (0/1)	-0.21	-0.15
	(0.21)	(0.73)
Dissimilar = 1 (0/1)	-0.31*	-0.95**
	(0.16)	(0.39)
Dissimilar = 2 (0/1)	0.09	-0.12
	(0.07)	(0.36)
Warm X Dissimilar = 1 (0/1)	0.32	0.51
	(0.25)	(0.79)
Warm X Dissimilar = 2 (0/1)	-0.79***	-2.18**
	(0.21)	(0.80)
Control Mean	0.91	2.45
Observations	49	49
R^2	0.343	0.245

Panel B: Inputs

	(1) Total Characters β / SE	(2) Characters per Minute β / SE	(3) Time in Experiment (Minutes) β / SE	(4) Time Active in Task (Minutes) β / SE	(5) Any Break (0/1) β / SE
Warm (0/1)	546.77	12.23	-25.33	-13.75	0.61***
	(4921.95)	(21.23)	(26.85)	(27.45)	(0.13)
Dissimilar = 1 (0/1)	-2583.33	-2.14	-17.28	-23.63	0.01
	(3208.68)	(15.88)	(24.23)	(15.06)	(0.16)
Dissimilar = 2 (0/1)	-1180.56	-10.45	38.62	40.58	0.08
	(2562.26)	(11.67)	(25.67)	(24.43)	(0.21)
Warm X Dissimilar = 1 (0/1)	2363.26	-12.69	92.45***	67.69***	-0.14
	(4960.93)	(24.85)	(28.74)	(22.08)	(0.24)
Warm X Dissimilar = 2 (0/1)	-3588.14	-24.52	-20.41	-24.26	-0.58
	(4238.67)	(19.15)	(43.92)	(37.85)	(0.33)
Control Mean	9203.73	56.88	217.73	142.43	0.09
Observations	49	49	49	49	49
R^2	0.060	0.104	0.248	0.183	0.303

Panel C: Survey

	(1) Partner Score > 5 (0/1) β / SE	(2) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	0.12	0.09
	(0.14)	(0.14)
Dissimilar = 1 (0/1)	-0.13	0.04
	(0.14)	(0.10)
Dissimilar = 2 (0/1)	0.23	-0.28***
	(0.17)	(0.09)
Warm X Dissimilar = 1 (0/1)	0.05	-0.20
	(0.17)	(0.14)
Warm X Dissimilar = 2 (0/1)	-0.63***	0.63***
	(0.18)	(0.17)
Control Mean	0.68	0.36
Observations	98	98
R^2	0.099	0.131

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, heterogeneous teams, and heterogeneous teams in warm room for (i) output in Panel A, (ii) inputs in Panel B, and (iii) subjective survey responses in Panel C. Control mean indicates mean values for homogeneous teams in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

A Online Appendix: Tables and Figures

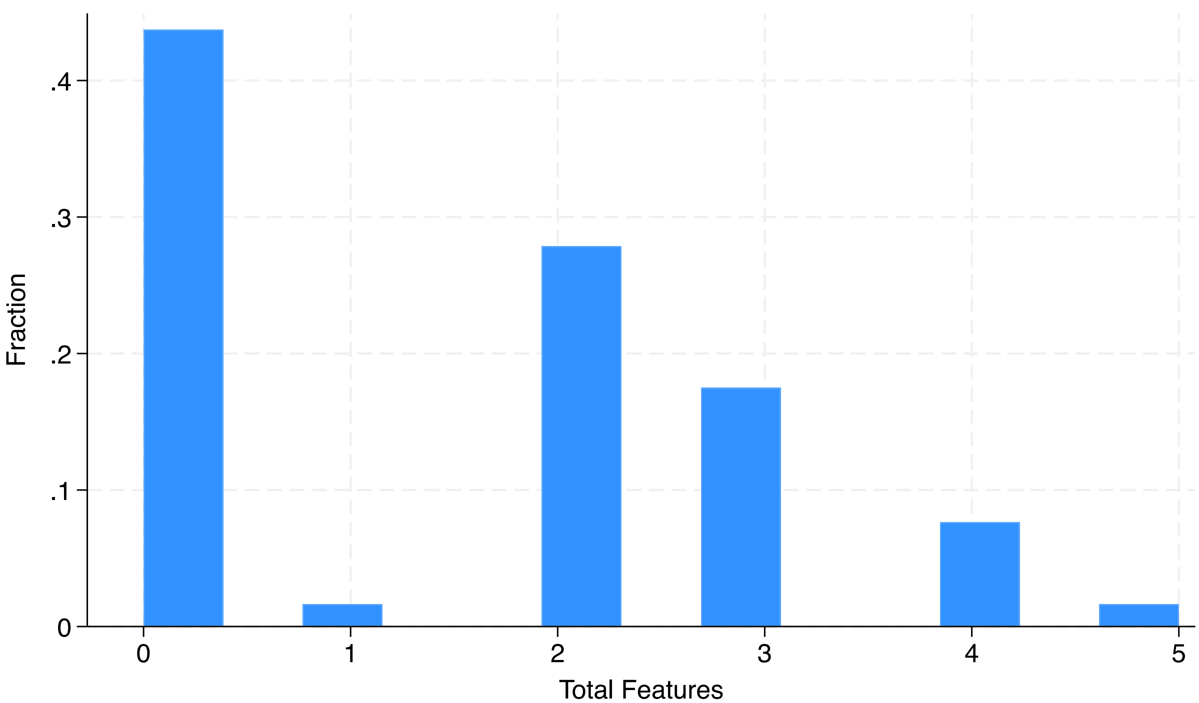
Figure A.1: Room Layout for Team and Individual Production

(a) Team layout

(b) Independent work layout

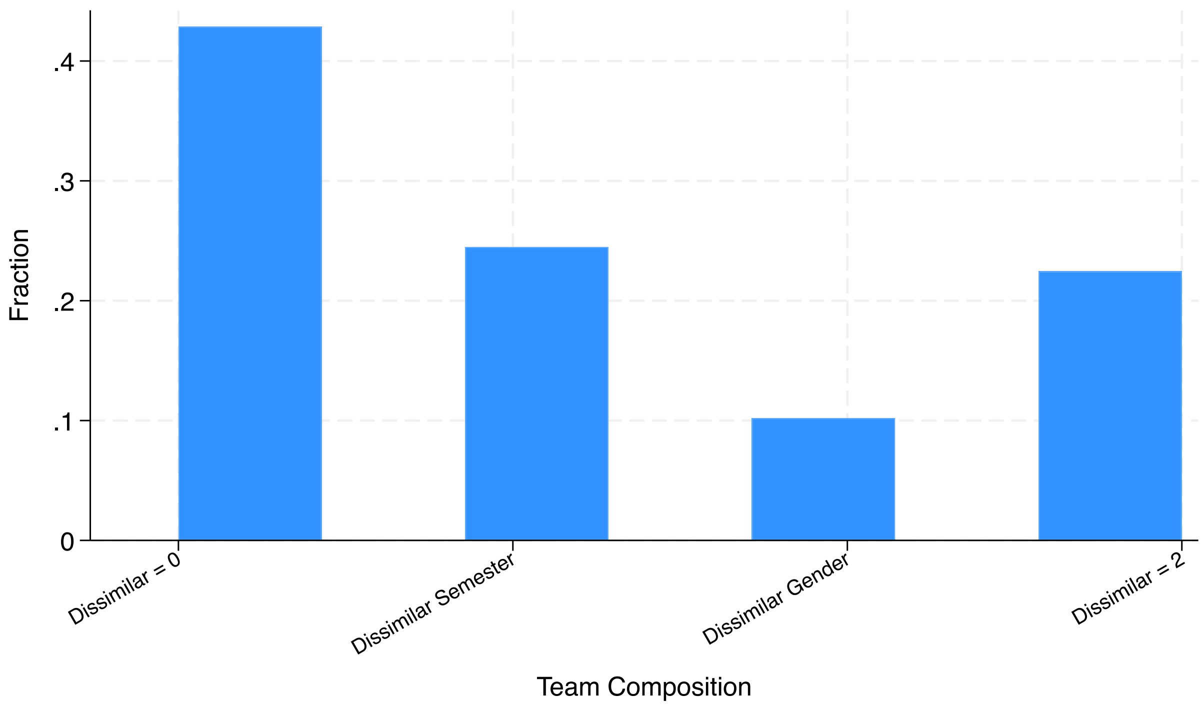
Notes: This figure presents room layout for team (Panel (a)) and individual production (Panel (b)) across both warm and control rooms.

Figure A.2: Sample Distribution of Job Output (Total Features Added)



Notes: This figure presents the distribution of total features added by teams and individuals in our sample.

Figure A.3: Sample Distribution of Team Composition



Notes: This figure presents the distribution of teams that are homogeneous in gender and experience (dissimilar = 0), differ only in experience (dissimilar semester), differ only in gender (dissimilar gender), or differ in both (dissimilar = 2).

Table A.1: Impact of Heat on Output in Team and Individual Production

	(1) Any Features β / SE	(2) At Least 2 Features β / SE	(3) At Least 3 Features β / SE	(4) At Least 4 Features β / SE	(5) All Features β / SE	(6) Total Features β / SE
Warm (0/1)	0.17*	0.15*	-0.08	-0.07	0.00	0.17
	(0.09)	(0.08)	(0.08)	(0.04)	(0.02)	(0.21)
Team (0/1)	0.37***	0.34***	0.07	0.04	-0.01	0.81***
	(0.10)	(0.10)	(0.08)	(0.08)	(0.01)	(0.26)
Warm X Team (0/1)	-0.44**	-0.43**	0.16	0.06	0.04	-0.61
	(0.20)	(0.20)	(0.18)	(0.11)	(0.05)	(0.61)
Control Mean	0.44	0.44	0.27	0.11	0.01	1.27
Observations	183	183	183	183	183	183
R^2	0.066	0.054	0.026	0.022	0.009	0.035

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, team production, and team production in warm room for output as measured by features added. Control mean indicates mean values for individual production in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table A.2: Impact of Heat on Subjective Survey Responses from Team Members and Individuals

	(1) Task Difficult (0/1) β / SE	(2) Task Engaging (0/1) β / SE	(3) Prefer Teamwork (0/1) β / SE	(4) Negative Affect (sd) β / SE	(5) Positive Affect (sd) β / SE	(6) Partner Score > 5 (0/1) β / SE	(7) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	-0.06	-0.11	-0.03	0.25	-0.25	-0.00	0.16
	(0.08)	(0.06)	(0.09)	(0.27)	(0.27)	(0.08)	(0.12)
Team (0/1)	-0.07	0.07	0.17**	-0.21	0.21		
	(0.10)	(0.06)	(0.08)	(0.16)	(0.16)		
Warm X Team (0/1)	0.07	0.00	0.02	0.07	-0.07		
	(0.15)	(0.10)	(0.12)	(0.31)	(0.31)		
Control Mean	0.55	0.67	0.57	-0.04	0.04	0.69	0.31
Observations	232	232	232	232	232	98	98
R^2	0.003	0.018	0.033	0.028	0.028	0.000	0.028

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, team production, and team production in warm room on subjective survey responses from team members and individuals. Control mean indicates mean values for individual production in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table A.3: Impact of Heat on Output in Individual Production for Male vs. Female and Higher Academic Seniority vs. Lower Academic Seniority

	(1) Any Features β / SE	(2) Total Features β / SE	(3) Any Features β / SE	(4) Total Features β / SE	(5) Any Features β / SE	(6) Total Features β / SE
Warm (0/1)	0.17*	0.17	0.15	0.13	0.15	0.20
	(0.09)	(0.21)	(0.12)	(0.42)	(0.12)	(0.43)
Female (0/1)	-0.22	-0.58			-0.19	-0.35
	(0.16)	(0.49)			(0.21)	(0.85)
Warm X Female (0/1)	-0.15	-0.40			-0.15	-0.86
	(0.21)	(0.56)			(0.34)	(0.97)
High Semester (0/1)			-0.07	-0.59	-0.06	-0.51
			(0.15)	(0.44)	(0.17)	(0.51)
Warm X High Semester (0/1)			0.04	0.04	0.04	-0.08
			(0.21)	(0.59)	(0.21)	(0.63)
Female X High Semester (0/1)					-0.05	-0.38
					(0.24)	(0.93)
Warm X Female X High Semester (0/1)					-0.01	0.80
					(0.46)	(1.17)
Control Mean	0.48	1.38	0.48	1.61	0.47	1.38
Observations	134	134	134	134	134	134
R^2	0.070	0.040	0.032	0.043	0.074	0.081

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, female, higher academic seniority, female in warm room, higher academic seniority in warm room, and higher academic seniority female in warm room on output. Control mean indicates mean values for individual production in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table A.4: Impact of Heat on Subjective Survey Responses in Same Gender and Mixed Gender Teams

	(1) Task Difficult (0/1) β / SE	(2) Task Engaging (0/1) β / SE	(3) Prefer Teamwork (0/1) β / SE	(4) Negative Affect (sd) β / SE	(5) Positive Affect (sd) β / SE
Warm (0/1)	0.10	0.00	0.01	0.33	-0.33
	(0.16)	(0.10)	(0.12)	(0.25)	(0.25)
Different Gender (0/1)	0.03	0.10	0.01	0.23	-0.23
	(0.15)	(0.16)	(0.15)	(0.22)	(0.22)
Warm X Different Gender (0/1)	-0.22	-0.32	-0.08	-0.08	0.08
	(0.16)	(0.20)	(0.19)	(0.54)	(0.54)
Control Mean	0.47	0.71	0.74	-0.31	0.31
Observations	98	98	98	98	98
R^2	0.017	0.041	0.002	0.035	0.035

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, mixed gender teams, and mix gender teams in warm room for subjective survey responses. Control mean indicates mean values for same gender teams in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table A.5: Impact of Heat on Subjective Survey Responses in Similar Academic Seniority and Dissimilar Academic Seniority Teams

	(1) Task Difficult (0/1) β / SE	(2) Task Engaging (0/1) β / SE	(3) Prefer Teamwork (0/1) β / SE	(4) Negative Affect (sd) β / SE	(5) Positive Affect (sd) β / SE
Warm (0/1)	-0.08	-0.08	-0.08	0.39	-0.39
	(0.21)	(0.10)	(0.10)	(0.30)	(0.30)
Dissimilar Semester (0/1)	-0.04	0.09	-0.13*	0.11	-0.11
	(0.17)	(0.08)	(0.07)	(0.30)	(0.30)
Warm X Dissimilar Semester (0/1)	0.22	-0.04	0.12	-0.14	0.14
	(0.25)	(0.20)	(0.18)	(0.49)	(0.49)
Control Mean	0.50	0.69	0.81	-0.30	0.30
Observations	98	98	98	98	98
R^2	0.016	0.020	0.012	0.028	0.028

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, dissimilar academic seniority teams, and dissimilar academic seniority teams in warm room for subjective survey responses. Control mean indicates mean values for similar academic seniority teams in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table A.6: Impact of Heat on Subjective Survey Responses in Heterogeneous and Homogeneous Teams

	(1) Task Difficult (0/1) β / SE	(2) Task Engaging (0/1) β / SE	(3) Prefer Teamwork (0/1) β / SE	(4) Negative Affect (sd) β / SE	(5) Positive Affect (sd) β / SE
Warm (0/1)	-0.00	-0.12	-0.11	0.54	-0.54
	(0.23)	(0.11)	(0.10)	(0.33)	(0.33)
Dissimilar = 1 (0/1)	-0.05	-0.22**	-0.31**	0.54*	-0.54*
	(0.19)	(0.09)	(0.13)	(0.28)	(0.28)
Dissimilar = 2 (0/1)	-0.00	0.23***	-0.03	0.14	-0.14
	(0.22)	(0.04)	(0.12)	(0.27)	(0.27)
Warm X Dissimilar = 1 (0/1)	0.05	0.29	0.28	-0.59	0.59
	(0.26)	(0.23)	(0.22)	(0.60)	(0.60)
Warm X Dissimilar = 2 (0/1)	0.00	-0.38*	-0.02	-0.04	0.04
	(0.26)	(0.19)	(0.18)	(0.62)	(0.62)
Control Mean	0.50	0.77	0.86	-0.48	0.48
Observations	98	98	98	98	98
R^2	0.002	0.101	0.062	0.063	0.063

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, heterogeneous teams, and heterogeneous teams in warm room for subjective survey responses. Control mean indicates mean values for homogeneous teams in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively.

Table A.7: Multiple Hypothesis Test Correction: Impact of Heat on Inputs in Team and Individual Production

	(1) Total Characters	(2) Characters per Minute	(3) Time in Experiment (Minutes)	(4) Time Active in Task (Minutes)	(5) Any Break (0/1)
	β / SE	β / SE	β / SE	β / SE	β / SE
Warm (0/1)	-2484.62	-16.98	8.43	4.64	0.07
	(2098.78)	(11.28)	(13.59)	(15.52)	(0.11)
Team (0/1)	-269.87	-5.62	17.71	16.79	-0.16**
	(2620.69)	(11.96)	(17.90)	(17.79)	(0.07)
Warm X Team (0/1)	3096.71	19.66	-7.88	-0.93	0.36**
	(3388.45)	(15.76)	(25.92)	(25.92)	(0.16)
Control Mean	8254.47	59.38	202.20	125.91	0.27
Observations	183	183	183	183	183
R^2	0.011	0.014	0.018	0.016	0.064

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of warm room, team production, and team production in warm room for inputs. Control mean indicates mean values for individual production in control rooms. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent level denoted by *, **, and ***, respectively. Anderson's sharpened q-values are computed within the family of input outcomes and 10, 5, and 1 percent statistical significance levels are denoted by †, ††, †††, respectively.

Table A.8: Multiple Hypothesis Test Correction: Impact of Heat on Inputs in Heterogenous and Homogenous Teams

Panel A: Gender	(1) Total Characters	(2) Characters per Minute	(3) Time in Experiment (Minutes)	(4) Time Active in Task (Minutes)	(5) Any Break (0/1)
	β / SE	β / SE	β / SE	β / SE	β / SE
Warm (0/1)	2234.81	13.25	2.80	9.43	0.59***†††
	(3160.73)	(12.84)	(25.82)	(24.99)	(0.10)
Different Gender (0/1)	34.86	-2.03	19.76	25.39	0.20
	(1361.04)	(9.99)	(19.31)	(19.98)	(0.20)
Warm X Different Gender (0/1)	-4468.93	-28.69**	-9.86	-20.43	-0.47*
	(2716.43)	(13.33)	(36.51)	(33.16)	(0.27)
Control Mean	7974.26	54.37	214.05	135.18	0.05
Observations	49	49	49	49	49
R^2	0.051	0.110	0.019	0.030	0.276

Panel B: Academic Seniority	(1) Total Characters	(2) Characters per Minute	(3) Time in Experiment (Minutes)	(4) Time Active in Task (Minutes)	(5) Any Break (0/1)
	β / SE	β / SE	β / SE	β / SE	β / SE
Warm (0/1)	124.69	3.83	-3.26	0.39	0.54***††
	(4261.18)	(19.01)	(20.78)	(24.16)	(0.14)
Dissimilar Semester (0/1)	-1984.25	-9.92	19.49	15.12	-0.08
	(2888.49)	(13.69)	(24.76)	(25.34)	(0.10)
Warm X Dissimilar Semester (0/1)	660.65	-5.46	14.53	12.15	-0.28
	(3740.42)	(18.50)	(31.89)	(28.71)	(0.28)
Control Mean	9013.46	58.91	209.80	134.86	0.15
Observations	49	49	49	49	49
R^2	0.020	0.045	0.061	0.044	0.290

Panel C: Gender and Academic Seniority	(1) Total Characters	(2) Characters per Minute	(3) Time in Experiment (Minutes)	(4) Time Active in Task (Minutes)	(5) Any Break (0/1)
	β / SE	β / SE	β / SE	β / SE	β / SE
Warm (0/1)	546.77	12.23	-25.33	-13.75	0.61***
	(4921.95)	(21.23)	(26.85)	(27.45)	(0.13)
Dissimilar = 1 (0/1)	-2583.33	-2.14	-17.28	-23.63	0.01
	(3208.68)	(15.88)	(24.23)	(15.06)	(0.16)
Dissimilar = 2 (0/1)	-1180.56	-10.45	38.62	40.58	0.08
	(2562.26)	(11.67)	(25.67)	(24.43)	(0.21)
Warm X Dissimilar = 1 (0/1)	2363.26	-12.69	92.45***	67.69***	-0.14
	(4960.93)	(24.85)	(28.74)	(22.08)	(0.24)
Warm X Dissimilar = 2 (0/1)	-3588.14	-24.52	-20.41	-24.26	-0.58
	(4238.67)	(19.15)	(43.92)	(37.85)	(0.33)
Control Mean	9203.73	56.88	217.73	142.43	0.09
Observations	49	49	49	49	49
R^2	0.060	0.104	0.248	0.183	0.303

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of (i) warm room, mixed gender teams, and mix gender teams in warm room in Panel A, (ii) warm room, dissimilar academic seniority teams, and dissimilar academic seniority teams in warm room in Panel B, and (iii) warm room, heterogeneous teams, and heterogeneous teams in warm room in Panel C. Control mean indicates mean values for (i) same gender teams in control rooms in Panel A, (ii) similar academic seniority teams in control rooms in Panel B, and (iii) homogeneous teams in control room in Panel C. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively. Anderson's sharpened q-values are computed within the family of all input outcomes and 10, 5, and 1 percent statistical significance levels are denoted by †, ††, †††, respectively.

Table A.9: Multiple Hypothesis Test Correction: Impact of Heat on Subjective Survey Responses in Heterogeneous and Homogeneous Teams

Panel A: Gender		
	(1) Partner Score > 5 (0/1) β / SE	(2) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	0.12	0.06
	(0.10)	(0.12)
Different Gender (0/1)	0.18	-0.18**
	(0.15)	(0.08)
Warm X Different Gender (0/1)	-0.37*	0.31*
	(0.20)	(0.16)
Control Mean	0.63	0.37
Observations	98	98
R^2	0.034	0.051

Panel B: Academic Seniority		
	(1) Partner Score > 5 (0/1) β / SE	(2) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	0.15	0.00
	(0.12)	(0.13)
Dissimilar Semester (0/1)	0.06	-0.13
	(0.15)	(0.10)
Warm X Dissimilar Semester (0/1)	-0.37**	0.36**
	(0.17)	(0.17)
Control Mean	0.65	0.38
Observations	98	98
R^2	0.050	0.061

Panel C: Gender and Academic Seniority		
	(1) Partner Score > 5 (0/1) β / SE	(2) Prefer Different Teammate (0/1) β / SE
Warm (0/1)	0.12	0.09
	(0.14)	(0.14)
Dissimilar = 1 (0/1)	-0.13	0.04
	(0.14)	(0.10)
Dissimilar = 2 (0/1)	0.23	-0.28***†
	(0.17)	(0.09)
Warm X Dissimilar = 1 (0/1)	0.05	-0.20
	(0.17)	(0.14)
Warm X Dissimilar = 2 (0/1)	-0.63***††	0.63***††
	(0.18)	(0.17)
Control Mean	0.68	0.36
Observations	98	98
R^2	0.099	0.131

Notes: This table presents the regression estimates, following the main estimating equation, of the effect of (i) warm room, mixed gender teams, and mix gender teams in warm room in Panel A, (ii) warm room, dissimilar academic seniority teams, and dissimilar academic seniority teams in warm room in Panel B, and (iii) warm room, heterogeneous teams, and heterogeneous teams in warm room in Panel C. Control mean indicates mean values for (i) same gender teams in control rooms in Panel A, (ii) similar academic seniority teams in control rooms in Panel B, and (iii) homogeneous teams in control room in Panel C. Standard errors are clustered at the room-date level. Statistical significance at the 10, 5, and 1 percent levels denoted by *, **, and ***, respectively. Anderson's sharpened q-values are computed within the family of all survey outcomes and 10, 5, and 1 percent statistical significance levels are denoted by †, ††, †††, respectively.

B Online Appendix: Experiment Script

Introduction

Good Morning, my name is [Enumerator Name] and I will be facilitating today's session.

Welcome to our office of ARCED Foundation and thank you for signing up to participate in our survey.

I will be here all day, if you have any questions throughout the day regarding the task or otherwise, please do not hesitate to ask me.

Now I will tell you a little bit about our study and what your schedule will be today, but before I move on, does anyone have any questions?

[Enumerator will wait to see if participants have any questions and will answer all questions before moving on.]

Experiment

We are conducting a study to understand how high-skilled computer programmers work on programming tasks. Professors Elizabeth Lyons and Teevrat Garg of the University of California San Diego and Maulik Jagnani, University of Colorado Denver are leading the research.

You have been selected to participate because you responded to the advertisement for this task that was posted at your University, have a computer science background or education, and are 18 years of age or older.

Today we will conduct a 4-hours session in which you will be presented with a programming task. There will then be a short post-task survey. You will earn a minimum of 1100 BDT for the session with the possibility of earning up to an additional 500 BDT depending on how much of the task you are able to complete in the session. You will earn 100 BDT for every feature you complete. There are 5 features in total. You will receive your payment at the end of the session.

Do you have any questions so far?

[Enumerator will wait until all participant questions are answered before moving on.]

Survey ID

You have all been given a Survey ID, which will be your unique identifier ID for the rest of the day. Please use this throughout the day, wherever you are asked for a "Survey ID", or "Username".

Task

Please wait while we enable your task. Within the next few minutes, you will begin your task.

To start your task, please click on the “Task URL” on your browser bookmark.

[Enumerator will walk around and ensure all participants are at the task page.]

First, you will have to sign up by providing a username, a password and a Location. Your “username” will be the “Survey ID” that has been assigned to you. Your “password” can be any combination of your choosing. For location, please put your ROOM Number.

Please create a new user profile using the above instructions now.

[Enumerator will walk around to check that all participants have set up a new account using their unique Survey ID as the name. Enumerator will wait until all participants have made an account before moving on.]

Now you will be transferred to the task page where all your tasks are given. There are 5 features in total.

Task Instructions
[Enumerator will read out all specific instructions outlined at the beginning of the task.]

You will implement an API in Java for an advanced calculator. This calculator application programming interface (API) supports operations on a variety of objects:

1. Long type
2. Double type
3. Complex numbers
4. Vectors
5. Logic

The (1-4) operations include:

1. adding two objects
2. adding a list of objects
3. subtracting two objects
4. multiplying two objects
5. multiplying a list of objects
6. dividing two objects
7. string representation of objects

To assist you, a basic skeleton program is given to you below. Please complete the classes and methods implementation and feel free to add any additional methods if needed. Note that you do not need to necessarily implement a main() method since you will implement an API. **Please do not delete any class.**

After clicking on the Submit button, your program will be compiled and tested, and results will be shown to you. If all the tests pass, then you can be assured that your program works according to the requirements.

NOTE: The Java task is very similar to other object-oriented language tasks and having familiarity with object-oriented programming, even without Java-specific knowledge, is sufficient for solving this task.

All the instructions that I have just read out are also at the beginning of your task and you can refer to them anytime.

Please raise your hand during the task if you have any questions/concerns, need a break or would like to get some more water.

[Enumerator will ensure first 2 lines of the task are correct:
Line 1: import java.lang.Long;
Line 2: import java.lang.Double;]

You may now begin the task.

It is now _____ pm. Your task period is now over. You have now run out of time to complete your task. Please submit whatever you have completed. You will now be asked to complete a post survey form.

End of Task
You have now completed the task. I would now like to ask you to wait upto 15 minutes so that we can calculate your participation fee, based on how many tasks you have completed. Meanwhile, please fill up the post-task survey.

If you have any questions at this stage, please feel free to ask.

[Enumerator calculates all tasks completed for each student and provides the participation fee accordingly. Students are then escorted out of the premises.]

C Online Appendix: Explanation and Summary of Deviations from Pre-Registration

Our pre-registration was for the initial experimental sessions conducted in late-2019 at the offices of Innovation for Poverty Action in Dhaka. Unfortunately, these sessions produced unusable data due to numerous unexplainable discrepancies. Specifically, we could not determine whether these discrepancies were the result of non-obvious errors in session management or recording. As a result, we were unable to use the data from these sessions. However, these sessions were informative as they demonstrated the number of participants we could recruit within our target time frame. This insight forced us to reconsider our target sample size and led us to eliminate our incentive design treatment.

As we were finalizing our experiment redesign and tightening our session protocols, the COVID-19 pandemic struck Bangladesh. Subsequently, Innovation for Poverty Action (IPA) Bangladesh was found non-compliant with government regulations and had to cease operations. Consequently, we were unable to conduct any further sessions for the next two years. Our experiments required in-person attendance to control room temperature effectively, and we opted not to require masks, fearing they might affect both room temperature and teamwork dynamics. Therefore, we postponed further sessions until it was safe to host participants in person without masks and until we could secure a new implementation partner. Unfortunately, due to the upheaval and disappointment caused by the pandemic and the closure of IPA Bangladesh, we did not update our pre-registration to reflect these changes in our experimental design.

Once it became safe to conduct sessions again in Dhaka, in 2022, we partnered with a new research implementation organization, ARCED Foundation, to carry out our experiment. During the process of onboarding ARCED Foundation and preparing for the experiment, we inadvertently failed to update our pre-registration. However, it's important to note that we obtained IRB approval for the changes in our experiment design and research implementation partner. Lastly, we have since updated our pre-registration to reflect the redesign of the experiment after the sessions were completed.

Our 2022 study deviates from our 2019 experimental pre-registration in following important ways:

1. We did not include the incentive design treatment described in the pre-registration in our study. The incentive design treatment was intended to vary the bonus structure workers faced to test if increasing bonuses over time would reduce any negative effects of heat on performance. We opted not to include this treatment because we did not believe we would have the necessary statistical power to analyze eight experimental conditions.

2. The primary outcomes we analyze are effort and performance, as outlined in our pre-registration. However, given the distribution of output (our performance measure), we primarily focus on indicator variables to capture whether or not a certain number of features were added to the code. In addition, given the similarity of findings across effort measures, we focus primarily on characters typed in total and per minute rather

than characters typed in addition to mouse clicks or scrolls. Furthermore, given the distribution of output and effort, we do not analyze total features added per characters typed. Lastly, we focus on whether individuals took a break rather than # of breaks because very few participants took more than one break.

3. We did not pre-specify heterogeneity tests that consider the effects of heat on mixed gender teams and teams with different academic seniority. These are additional tests we introduce to provide supporting evidence on mechanisms.