Stress, Ethnicity, and Prosocial Behavior

Johannes Haushofer
Stockholm University, Research Institute of Industrial Economics, Max Planck Institute for Collective Goods, National Bureau of Economic Research, and Bureau for Research in the Economic Analysis of Development

Sara Lowes
University of California San Diego, National Bureau of Economic Research, and Bureau for Research in the Economic Analysis of Development

Abednego Musau
University Medical Centre Groningen

David Ndetei
University of Nairobi and Africa Mental Health Research and Training Foundation

Nathan Nunn
University of British Columbia, National Bureau of Economic Research, Bureau for Research in the Economic Analysis of Development, and Canadian Institute for Advanced Research

Moritz Poll
Brown University and Busara Center for Behavioral Economics

Nancy Qian
Northwestern University, Fudan University, National Bureau of Economic Research, and Bureau for Research in the Economic Analysis of Development

While observational evidence suggests that people behave more prosocially toward members of their own ethnic group, many laboratory studies fail to find this effect. One possible explanation is that coethnic

We thank the participants for their time; the Busara Center for Behavioral Economics, Nairobi, for implementing the study; and Clemens Kirschbaum, PhD, Technische Universität

Electronically published April 28, 2023

Journal of Political Economy Microeconomics, volume 1, number 2, May 2023.

https://doi.org/10.1086/722367

preference only emerges during times of stress. To test this hypothesis, we pharmacologically increase levels of the stress hormone cortisol, after which participants complete laboratory experiments with coethnics and non-coethnics. We find mixed evidence that increased cortisol decreases prosocial behavior. Coethnic preferences do not vary with cortisol. However, in contrast to previous studies, we find strong and robust evidence of coethnic preference.

I. Introduction

Observational studies have documented a relationship between ethnic identity and a wide range of social, economic, and political outcomes in many settings, including sub-Saharan Africa (e.g., Easterly and Levine 1997; Franck and Rainer 2012; Hjort 2014; Rasul and Rogger 2015; Alesina, Michalopoulos, and Papaioannou 2016; Marx, Stoker, and Suri 2019). To test for a causal role of ethnicity, laboratory experiments in sub-Saharan Africa have randomly paired participants with coethnic or non-coethnic partners and studied behavior in economic exchange games. Surprisingly, these studies often fail to find evidence of a coethnic preference in behavior (Habyarimana et al. 2007, 2009; Berge et al. 2020; Blum, Hazlett, and Posner 2021). The difference between observational findings and lab-based findings presents an important puzzle in our understanding of how ethnicity affects social outcomes and how we interpret and address ethnic tensions.

This study attempts to make progress on this issue by examining a potential explanation for the incongruence between experimental and observational studies. We consider the possibility that stress may strengthen coethnic preference. In contrast to laboratory studies, which occur in a relatively low-stress environment, observational studies are more likely to reflect decisions made when individuals are under stress. Under peaceful and low-stress conditions, coethnicity may be less salient; individuals may follow norms that dictate treating non-coethnic individuals and coethnic individuals similarly. However, when stress is high, coethnic preferences may rise to the surface.

In this study, we examine the effects of stress on coethnic preference. Our study includes 1,784 participants who, in a laboratory setting in Nairobi, Kenya, play behavioral games with other players. Nairobi is a natural

Dresden, Germany, for analyzing the salivary cortisol samples. This study was approved by the Princeton Institutional Review Board (IRB; protocol no. 7200), as well as the Kenya Medical Research Institute (protocol no. 494) and the Kenyan Pharmacy and Poisons Board (PPB/ECCT/15/12/04/2016(37)). The Harvard, Northwestern, and Bocconi University IRBs and the National Bureau of Economic Research IRB ceded review to Princeton after initial approval. It was preregistered at https://www.socialscienceregistry.org/trials/3937. We are grateful for funding provided by the National Science Foundation. We thank Joris Mueller for stellar research assistance. This paper was edited by John A. List.

setting for studying coethnic preferences. Neighborhoods are often organized along ethnic lines and political competition takes place mainly along ethnic lines, with recent elections experiencing prolonged post-election violence (Ndegwa 1997; Waki 2008; Hjort 2014; Macchiavello and Morjaria 2015; Jakiela and Ozier 2019).

As in previous studies, we randomly manipulate whether the other player belongs to the same or a different ethnic group. The games are a standard one-shot dictator game, choose your dictator game, and both stages of the trust game. These measure altruism, how altruistic the participant perceives another player to be and their altruism toward that player, trust, and trustworthiness, respectively. We also administer a social proximity survey in which respondents are presented with different player profiles and asked how close the respondent feels to them, how much they trust them, and how likely it is that they would be friends with them.

Our experiment makes an important departure from existing studies. To study the effects of stress, we experimentally manipulate the cortisol levels of half of the participants before they play the games. Real-world stressors induce several types of physiological responses such as an increase in the hormones cortisol and adrenaline and an increase in heart rate. Thus, stress is generally characterized by the presence of these markers (Kirschbaum and Hellhammer 1989; de Kloet 2004; de Kloet, Joëls, and Holsboer 2005). The most common strategy used to study stress is to expose subjects to stressful situations to induce physiological responses and then examine subsequent behavior. This strategy faces several important challenges. First, the treatment can trigger responses other than stress. Second, the subject is aware of the treatment, which causes concern about experimenter demand effects. Finally, the experimenter must identify a treatment that is unpleasant enough to elevate stress levels, while still being ethical.

In this study, we implement an alternative strategy that has been implemented in the neurobiological stress literature but is less common in the economics literature (Schwabe et al. 2010; Henckens et al. 2011; Kandasamy et al. 2014). Rather than exposing subjects to a stressful activity to induce the physiological responses of stress, we manipulate the physiological response itself and ask how this physiological response affects coethnic preference. For half the sample, we manipulate participants' cortisol levels by orally administering hydrocortisone pills, the pharmacological precursor of cortisol, which is converted to cortisol upon ingestion. The other half of participants receive an identical-looking placebo pill.

There are several advantages of focusing on cortisol in the laboratory setting. The first advantage is that cortisol can be cleanly manipulated; that is, we can isolate the effects of the manipulation of cortisol. This allows us to infer that any observed behavioral effects are due to this particular marker of stress. In contrast, other stress induction methods, such as social or physical stressors, generate additional physiological responses, including increases in heart rate and levels of other hormones such as adrenaline and noradrenaline. The second advantage is that the hydrocortisone treatment reliably elevates cortisol levels throughout the study session, which mitigates concerns that treatment intensity will fluctuate over the course of the experiment.

Participants were randomly paired with another participant for each round of the behavioral games. Although the identity of the other player was always unknown, participants were provided with the other player's age group (young, middle-aged, old), gender, and mother tongue, which was either Luo, Kikuyu, Luhya, or Kamba, and served as a direct indicator of the other player's ethnicity.

The study generates several findings. First, the administration of hydrocortisone was effective at increasing cortisol levels, and the magnitude of this increase was substantially larger than other stress induction methods.

Second, we find some indication, although limited, that hydrocortisone causes participants to behave less prosocially. The hydrocortisone treatment decreases giving in the dictator game by 4.4% of the sample mean, although this result varies in precision across specifications. In the trust game, it decreases the amount of money allocated to the other player by 5.6% of the sample mean. There is no effect on the amount of money returned by player 2 in the trust game or in the responses in the social proximity survey. When we estimate an average effect across all outcome measures, we find a negative but imprecise effect of hydrocortisone.

Third, we find strong and consistent evidence of coethnic preference. Participants allocate more money to coethnic partners than to non-coethnic partners in the dictator game (5.2% of the sample mean) and are 7.2 percentage points more likely to choose a coethnic dictator in the choose your dictator game. In the first (sending) stage of the trust game, participants allocate more money to coethnic partners relative to non-coethnic partners (7.5% of the sample mean), although coethnicity does not affect the amount player 2 returns.[1] Finally, we find that participants report feeling closer to and more positively toward coethnics in the social proximity survey. When we consider an average effect across all outcomes, we estimate a precise and large coethnic preference.

Fourth, we find no evidence that coethnic preference is stronger under hydrocortisone. The estimated interaction of hydrocortisone and coethnicity—that is, the differential effect of coethnicity under the hydrocortisone treatment—is small in magnitude, imprecisely estimated,

[1] As we will discuss, this appears to be due to the fact that we elicited participants' actions using the strategy method, which asks participants how much they would give in every possible scenario that they could face. The evidence indicates that this led to a lack of attention and focus in decision-making.

and varies in sign across games. When we estimate an average effect across all measures, we actually find that hydrocortisone decreases coethnic preference, although the estimate is imprecise and small in magnitude.

As in all experimental studies, experimenter demand effects are an important concern. This is even more true in our study, which communicates the other player's ethnicity directly with information on mother tongue. We use several strategies to address concerns about experimenter demand effects. First, we convey information about the other players in a natural manner, telling participants that "We are not able to tell you the exact identity of the other participants in the tasks. Instead, we are only able to provide you with some basic information about them. In particular, we will provide you with information about his/her age group, gender, and mother tongue. The other participants have the same information about you." The bundle of characteristics that we report—a person's age category (young, middle-aged, or old), gender (female or male), and mother tongue—would be some of the first things one would learn about another person if meeting in a real-world situation. Communicating ethnic identity in this manner shrouds our interest in ethnicity, while still conveying the information precisely to participants. Focus groups conducted during the pilot indicate that the experiment's interest in ethnicity was not obvious to participants.

Second, we elicit social preferences using surveys, which we expect to be more susceptible to demand effects, in addition to the main incentivized behavioral tasks, which we expect to be less susceptible. Thus, comparing the estimated effects using different measures allows us to assess the likely severity of demand effects.

Finally, we directly measure demand effects using a recently developed method that actively reveals the expectations of the experimenter and measures how much behavior changes in response (De Quidt, Haushofer, and Roth 2018). Participants play an additional single round of the dictator game at the end of the experimental session, before which half of participants are presented with the following statement: "We expect that people in your group will give more than they otherwise would." We find that this statement has virtually no effect on the amount given in the dictator game. This is true for the full sample and for each of the treatment subsamples: the hydrocortisone group, the placebo group, those paired with a coethnic, and those paired with a noncoethnic. This suggests that it is unlikely that demand effects are driving our results.

Our findings make several contributions to the existing literature. Cortisol is important to study for the following reasons. Cortisol is one of the key neurobiological markers of stress (Kirschbaum and Hellhammer 1989; de Kloet 2004; de Kloet, Joëls, and Holsboer 2005). There is substantial evidence that stress increases cortisol levels (Hines and Brown

1936; Kirschbaum, Pirke, and Hellhammer 1993; Minkley et al. 2014). Thus, in contexts where individuals have increased stress, they will also typically have increased cortisol levels. Second, the existing literature on the effects of interventions on stress focuses on cortisol as an indicator of stress (Coates and Herbert 2008; Haushofer and Shapiro 2016; Haushofer et al. 2020). Thus, our understanding of the effects of interventions on stress is often proxied by the effects on cortisol levels. Finally, increased cortisol levels have led to changes in behavior, such as changes in time and risk preferences (Kandasamy et al. 2014; Riis-Vestergaard et al. 2018); other paradigms for studying stress have also affected these same behaviors (Porcelli and Delgado 2009; Delaney, Fink, and Harmon 2014).

We contribute to the evidence on the effects of stress on prosocial behavior, which is limited and somewhat mixed. In a study with 80 participants, von Dawans et al. (2012) find that exposure to a social stressor (the Trier social stress test, TSST) increases sharing in a dictator game and the levels of trust and trustworthiness in the trust game. With a sample of 78 participants, Margittai et al. (2015) find an increase in altruism in a dictator game for in-group but not out-group members after the TSST. In contrast, Vinkers et al. (2013) find the opposite effect of the same stressor in the dictator game among a sample of 72 participants. Our finding of a negative effect of hydrocortisone on prosocial behavior is consistent with the findings of Vinkers et al. (2013) but conflicts with those of von Dawans et al. (2012) and Margittai et al. (2015). We see two potential explanations for these discrepancies. First, hydrocortisone, which only affects cortisol levels, may have different behavioral effects than the TSST, which has broad physiological and psychological consequences. Additionally, our sample is 22 times as large as those used in these previous studies. With the small samples in the previous studies, it is perhaps not surprising that there are conflicting results in the existing literature.

Second, our findings speak to the apparent discrepancy between studies that estimate the importance of ethnicity using observational data in sub-Saharan Africa (e.g., Easterly and Levine 1997; Alesina, Baquir, and Easterly 1999; Alesina et al. 2003, 2021; Franck and Rainer 2012; Michalopoulos 2012; Hjort 2014; Yanagizawa-Drott 2014; Burgess et al. 2015; Depetris-Chauvin 2015; Rasul and Rogger 2015; Alesina, Michalopoulos, and Papaioannou 2016; Marx, Stoker, and Suri 2019) relative to those that use experimental methods (Habyarimana et al. 2007, 2009; Berge et al. 2020; Blum, Hazlett, and Posner 2021). They also contribute to a better understanding of whether findings from laboratory studies generalize to the real world (Levitt and List 2007). Our findings of a strong coethnic preference raise the possibility that how player identity is communicated may affect whether coethnic preference is found in laboratory studies in sub-Saharan Africa. This possibility is consistent with the fact that, within the literature, coethnic preference appears to be less commonly found in behavioral games, where ethnicity is often indirectly reported, than in implicit association tests (IATs) where ethnicity, by the nature of the IAT, is directly reported (e.g., Lowes et al. 2015; Berge et al. 2020). It is also consistent with the fact that in-group preference has been found in laboratory studies outside of the African continent (Fershtman and Gneezy 2001; Hewstone, Rubin, and Willis 2002; Chen and Li 2009; Balliet, Wu, and De Dreu 2014).

Our findings also contribute to existing studies that try to understand why lab-based evidence of coethnic preference has been so elusive. Most closely related is Blum, Hazlett, and Posner (2021), who investigate the importance of specific experimental protocols for identifying coethnic preference. In the study, the authors use nearly identical protocols and the same behavioral lab to confirm earlier findings from Berge et al. (2020), which show no evidence of a coethnic preference in the dictator game or a public goods game. They then show that they are able to find evidence of coethnic preferences when they use misattribution task measures. These tasks, which were recently developed in the social psychology literature, measure more automatic associations and evaluations. Their findings, like ours, provide evidence for coethnic preference within the African context and suggest that previous findings of no coethnic preference are due to the sensitivity of the estimates to the details of the experimental protocols.

Finally, our findings are related to the existing observational evidence suggesting that conflict may induce stronger coethnic preference. For example, the meta-analysis of Bauer et al. (2016) finds that war exposure may increase prosocial behavior toward in-group members but not toward out-group members. Hjort (2014) finds evidence that Kenyan flower workers discriminate in favor of a coethnic coworker relative to a noncoethnic coworker prior to an election; this coethnic preference appears even stronger following the outbreak of election violence. Blouin and Mukand (2019) find evidence of coethnic preference in Rwanda following the genocide and show that the government's efforts at postgenocide nation building through radio programming has been successful at reducing coethnic preference. To the extent that one consequence of conflict is heightened stress, these prior findings suggest that stress might affect in-group preferences. However, our finding that hydrocortisone does not lead to greater coethnic preference indicates that stress as reflected by increased cortisol levels is unlikely to be the main driving force behind these results. Thus, the prior findings are likely due to other effects of conflict.

This paper proceeds as follows. Section II describes the experimental design. Section III discusses the estimating equations. Sections IV and

V present our results and prespecified robustness checks. Section VI concludes.

II. Design

The full study design and analysis were prespecified in Haushofer et al. (2019), and all prespecified analysis is reported in the main body of the paper or in the appendix. Analysis that was not prespecified will be noted in the paper. We also point out a number of minor design changes that occurred after the submission of the preanalysis plan in table A1.

A. Sampling Strategy

The study takes place at the Busara Center for Behavioral Economics, a behavioral science laboratory in Nairobi, Kenya. Busara maintains a database of registered participants that is intended to be representative of residents of the informal settlements (commonly referred to as "slums") of Nairobi. Individuals are recruited directly by Busara's field officers who go to the relevant neighborhoods of Nairobi. Busara has been continually recruiting since 2012. Previous analysis shows that Busara's participant pool is similar to the population of Nairobi and of Kenya based on observable demographics such as gender, age, and ethnicity (Haushofer et al. 2014).

We restrict our study to individuals who had not previously participated in lab experiments. Participants need to have access to the mobile money system M-Pesa because it is used to send payments. We require participants to be between 18 and 40 years of age and exclude participants who are pregnant or breast-feeding. All participants undergo thorough health screening prior to participating. More details on the extensive participant safety measures, including a health screening protocol, can be found in appendix D.

We also require participants to belong to one of the four largest ethnic groups in the area: Luo, Kikuyu, Luhya, and Kamba. Recruitment took place in three of Nairobi's major informal settlements: Kibera, Kawangware, and Viwandani. A map of the settlements is shown in figure A1. Figure A2 presents the neighborhood and ethnic composition of each lab session.

An important factor that affects whether our final participant pool is representative of the broader population is the attrition rate (List 2020). There are a number of reasons why someone who was originally contacted for the study may not participate in the end. They may decline the invitation or accept the invitation but fail to arrive for their appointment time. They may also have been screened out during a thorough health exam, although they still received the show-up and transport fee in that case.

A summary of the different steps of the process as well as the nature of attrition is provided in table A2. Each entry reports the mean and standard deviation for a characteristic of a sample. The table shows how the characteristics (i.e., gender, ethnic identity, age, and education) change in each step as we move from the population of Kenya to the final sample in our paper.

We find that our final sample is slightly more likely to be male than the Kenyan and Nairobi populations, and this is primarily due to men being more likely to sign up to participate. The final group of participants is also younger. For most steps of the process, younger individuals were more likely to agree to continue. They were also less likely to be screened out during the medical exam.[2] In terms of ethnic groups, we find that our final sample is representative in its share of Kikuyu, but slightly overrepresents Luo, Luhya, and Kamba. The overrepresentation is expected since we only included four ethnic groups in our study. Note that the Kenya National Bureau of Statistics, from which these data are obtained, does not separately report mother tongue for Nairobi. We also find that more-educated individuals were more likely to participate in the study, driven mostly by higher secondary attainment. When considering the external validity of our estimates, these differences between our sample and the broader population should be taken into account.

B. Stress Treatment

Humans respond to stressful situations by activating two main stress hormone pathways: the sympathetic-adrenal-medullary (SAM) axis, which releases catecholamines (adrenaline and noradrenaline, also called epinephrine and norepinephrine), and the hypothalamic-pituitary-adrenal (HPA) axis, which releases cortisol. Because of its close link to stress, cortisol is often called a “stress hormone” (e.g., Lupien et al. 2007).

Studying stress using the oral administration of hydrocortisone to increase cortisol levels is common in the neurobiological stress literature (Schwabe et al. 2010; Henckens et al. 2011; Kandasamy et al. 2014). Another common strategy is to induce moderate levels of stress by having participants engage in uncomfortable activities. While there are trade-offs between the two strategies, there are a number of advantages to using hydrocortisone. First, raising cortisol levels by administering hydrocortisone is less risky and uncomfortable for participants than the implementation

[2] Because younger individuals tended to have more education, individuals with more education, particularly secondary education, were more likely to participate in the study.

of uncomfortable tasks that are meant to induce stress; hydrocortisone does not generate the feeling of stress, while other protocols (such as the cold pressor task, in which respondents are asked to put their hand in cold water) do (Henckens et al. 2011). Thus, by administering hydrocortisone, we are able to induce the neurobiological stress response without making participants uncomfortable.

Second, since participants are unable to detect whether or not they receive hydrocortisone (e.g., Riis-Vestergaard et al. 2018), a fact that we confirm in our setting, its administration does not suffer from experimenter demand effects to the same degree as other laboratory stressors.

Third, the physiological effect of hydrocortisone does not depend on any cultural variables. For example, the Trier social stress test (TSST) has been shown to be ineffective in Kenya, possibly because public speaking has different social significance in the Kenyan context relative to a Western context (Haushofer, Jang, and Lynham 2015). Similarly, even physical stressors such as the cold pressor task, which in principle should have similar effects across cultures because of their comparable physiological effects, can be construed differently. For example, it may be viewed as a welcome challenge in which one can prove oneself rather than a stressful experience, and thus it may have different effects between collectivist and individualist cultures. This task has only shown very transitory effects on stress levels in Kenya (Haushofer, Jang, and Lynham 2015). In contrast, hydrocortisone has effects on cortisol levels in our Kenyan sample similar to those it has in Western samples.

Finally, social and physical stressors generate many physiological effects, including the release of other stress hormones such as adrenaline and noradrenaline. In contrast, hydrocortisone administration leads to a specific increase in cortisol only, without other physiological effects, and thus allows us to make an unambiguous attribution of any behavioral effects to a single physiological mechanism.[3]

Participants are randomly assigned to be in either a treatment group that receives 20 mg of oral hydrocortisone or a control group that receives an identical-looking placebo. The procedure is double blind: neither the participants nor the administering staff know to which treatment a participant is assigned. This avoids biasing the participants themselves or the way that staff interacts with them. Saliva samples are collected at six predetermined points during the experiment: immediately before administration of the pill and immediately before each stage of each game and the social proximity survey. These samples are later assayed for salivary cortisol

[3] Targeting adrenaline and noradrenaline, in contrast, is more likely to produce side effects. For example, the drug yohimbine, which is commonly used, affects the sympathetic nervous system, increasing heart rate and blood pressure.

and allow us to observe whether the hydrocortisone treatment successfully increases participants' salivary cortisol, and if so, for how long.

C. Coethnicity Treatment

To manipulate coethnicity, we provide information on the other player's mother tongue, which is either Luo, Kikuyu, Luhya, or Kamba. In general, this is strongly associated with ethnicity in Kenya and among the four ethnic groups in our sample, it is synonymous with ethnicity. Participants are also given information on the other player's age group (young, middle-aged, old) and gender (female or male). Our protocols are similar to previous studies that also report information of players directly to the participants (Lowes et al. 2015; Berge et al. 2020; Blum, Hazlett, and Posner 2021). However, in contrast to some previous studies (Berge et al. 2020), but consistent with others (Lowes et al. 2015; Blum, Hazlett, and Posner 2021), we report the ethnic identity of the other player directly.

D. Measuring Prosocial Preferences

After administration of the hydrocortisone and placebo pills, each participant completes four activities that measure prosocial attitudes and behaviors: the dictator game (DG), where they allocate money between themselves and the other player; the choose your dictator game (CYD), where they choose which of two individuals they would like to be the decision maker in a dictator game in which they are the recipient; the standard trust game, both as player 1 (TG1), where they choose how much of their endowment to send to the other player, and as player 2 (TG2), where they choose how much of each possible increased amount to send back to player 2; and a social proximity survey, where they are provided information about the attributes of a specific individual (e.g., old, female, Kikuyu) and asked three questions: how close they feel to this person, how likely they are to be their friend, and how much they trust someone like them. In order not to skew participants' behavior by connotations with words such as "dictator" or "trust," we refer to the games during the experimental sessions as the "allocation task," "choose the person task," and "two-stage allocation task." Experimental sessions are randomly assigned to two orderings of games: (i) DG, CYD, TG1, TG2; (ii) TG1, TG2, DG, CYD. The social proximity survey is always completed after all of the behavioral games.

Following this, participants are asked to make an incentivized guess of their treatment status.[4] Finally, they participate in a single round of the

[4] We ask participants to guess which of the pills they got, and if they guess right, they receive an extra 50 Kenyan shillings (KES 50), approximately USD 0.50, in their payout.

dictator game, which was part of the module measuring experimenter demand effects. A full timeline of the activities is provided in appendix A.4.

For each of the four main games, the participant plays six iterations, each time with a different player or pair of players. For each game, one of the six iterations is randomly selected to calculate the participant's payoff for that activity. Participants are thus paid for one decision from each game. Participants are informed that they are playing with real people who have taken part in the study, but who are not currently in the room with them. To determine payoffs, participants in the study are randomly matched to players (and their choices) from a pool of previous participants, which grew as the study progressed. All activities are completed on touch screen computers; a short training is given to all participants on the use of touch screen computers prior to the experiment.

For each experimental task, participants are randomly paired with another player. Participants are informed that they will not be told the exact identity of the players that they are paired with, but that they will be given information on their age group, gender, and mother tongue. They are also informed that the other player will receive the same information about the participant.[5] Participants are assigned to play with a man or a woman with equal probability, and with someone from each of the age categories with equal probability. For four of the six iterations of the dictator game and the trust game, participants are paired with one individual from each of the four ethnic groups. For the remaining two iterations, they are paired with a coethnic partner for one iteration, and a noncoethnic partner—who is randomly selected from the other three ethnic groups—for the other iteration.

We ensured that a particular partner type (e.g., young, female, Luo) is not repeated within a game for a participant. For the CYD game, the six iterations of the game correspond to the six possible pairings of the four ethnicities in the study in random order; for example, the participants are presented with each of the following pairings across the six rounds: Luo-Kikuyu, Luo-Luhya, Luo-Kamba, Kikuyu-Luhya, Kikuyu-Kamba, Luhya-Kamba. The age group and gender of all pairings are fully randomized as in the other games. For the four rounds of the social proximity survey, each ethnicity appears exactly once in random order, and the other characteristics are randomized as described above.

[5] The exact wording was, "We will provide you with some information about participant 2, but you will not know exactly who participant 2 is. We will only provide you with the following information about participant 2: his/her age group, gender, and mother tongue. Likewise, participant 2 will not know exactly who you are. Participant 2 will only know your age group, gender, and mother tongue."

1. Dictator Game

In this activity, the participant is assigned to be player 1 in the dictator game. The decision involves choosing how much of a KES 200 endowment to send to player 2.[6] The participant decides how much to allocate to the other player by entering this amount into the computer. The outcome of interest is the share of the endowment [0, 1] that player 1 sends to player 2. A version of this game has been previously used in a number of studies to test for the presence of coethnic preference in similar contexts (e.g., Habyarimana et al. 2007; Berge et al. 2020; Blum, Hazlett, and Posner 2021). We interpret game play as a measure of prosocial behavior (List 2007) and address social desirability concerns (Bardsley 2008) in a dedicated experimenter demand module discussed below.

2. Choose Your Dictator Game

In this game, the participant is presented with two possible player 1s for the DG, called person A and person B. The participant chooses one of the two individuals to be their player 1. As described above, the participant is given information on each of the possible player 1s; for example, they have information on the age group, gender, and mother tongue of person A and person B. The participant also knows that person A and person B have the same information on the participant when each makes their allocation decisions. The participant indicates which person they prefer to have as their player 1. The outcome of interest is which player, person A or B, the participant chooses as their player 1.

3. Trust Game

This activity consists of two stages. In the first stage, player 1 chooses an amount from an endowment of KES 200 to send to player 2. The amount sent is multiplied by 3, and player 2 receives the tripled amount. Player 2 then decides how much of the amount received to send back to player 1. Player 2 keeps the amount they do not send back to player 1, and player 1 receives the amount they have not allocated to player 2, as well as the amount player 2 returns to them. The amounts that player 1 can allocate to player 2 are limited to KES 0, 40, 80, 120, 160, or 200, and player 2 indicates how much of the tripled amount they would return for each of these allocations. The participants complete the task both as player 1 and player 2. Thus, the outcomes of interest are (i) the share of the KES 200 endowment that player 1 sends to player 2, and (ii) the share of the tripled amount received that player 2 sends back to player 1. Both variables

[6] KES 100 was equal to approximately USD 1 at the time of the experiment.

take values that range from 0 to 1. The share player 1 sends to player 2 can be viewed as a measure of trust, and the share player 2 sends back to player 1 as a measure of trustworthiness.

4. Social Proximity Survey

To measure participants' self-reported proximity to people from different ethnic groups, we administer a "social proximity" survey in which participants are asked to report how close they feel to four individuals, one of each ethnicity. The age group and gender of these other individuals are randomly assigned. Our outcome variable is an average of the three social proximity questions rescaled to lie between 0 and 1. The questions are as follows:

1. "How likely are you to be friends with a person with the following characteristics?" Answer choices: Very unlikely, unlikely, neither likely nor unlikely, likely, very likely.
2. "How much do you trust a person with the following characteristics?" Answer choices: Not at all, a bit, somewhat, mostly, completely.
3. "Using the figures provided, which set of figures best represents how close you feel to a person with the following characteristics?" The question is accompanied by a figure showing social proximity.[7] Answer choices range from 0 (not at all close) to 5 (very close).

In addition to being standard measures of prosocial behavior, our measures also provide some variation in the extent to which they might be affected by experimenter demand. We expect the social proximity survey, where we ask participants directly about the perceived closeness of another person to them, to be most susceptible to experimenter demand. Therefore, a comparison of the estimated coethnicity effect across the range of activities is informative about the sensitivity of our findings to experimenter demand effects.

E. Experimental Setting

The study includes 1,784 participants from 119 experimental sessions. On average, each session has 15 participants. The CYD, DG, and TG1 are played six times by each participant, resulting in 10,704 total observed decisions for each of these games. The 30 TG2 and 4 social proximity survey answers yield 53,520 and 7,136 observed decisions, respectively. Each participant receives a participation and transport fee of KES 350 and an additional KES 50 if they arrive on time. Participants also receive money

[7] See fig. A4.

based on their decisions during game play. The amounts are typically between KES 500 and KES 700. Average participant earnings are KES 639 (KES 629 for placebo and KES 648 for participants who received hydrocortisone). The money is transferred to the participants via M-Pesa, generally within 24 hours after the completion of each session.

Upon arrival, we verify participants' identity with the invitation roster and refer them to a waiting area with a consent form in English, Swahili (the commonly spoken local language), or both, according to their preference. After welcoming participants to the Busara Center, we then read out the consent form to participants in Swahili, allowing for questions. A detailed, IRB-approved medical screening including vital signs is administered by full-time nurses in a private setting. In appendix D we offer a more in-depth discussion of participant safety measures. Participants who do not meet the inclusion criteria or who choose not to consent are paid in full for showing up and for their transportation costs. The remaining participants are randomly assigned seats and proceed to a computer laboratory, where each participant is seated in an assigned cubicle.

The games and surveys are administered using touch screen computers to enable computer-illiterate individuals to participate. See figure A4 for a sample of the screen. Enumerators read instructions to the participants in Swahili to maximize comprehension. Instructions also appear on the screen in both English and Swahili. For instructions that vary by participant (e.g., the information on the player they were paired with), prerecorded audio instructions are read aloud to the participants, all of whom wear headphones while playing the games. The purpose of this approach is twofold: it ensures that even illiterate participants had the full set of information available to them when making decisions, and it also increases the salience of the characteristics of the person with whom a participant is paired. For all games, every decision is preceded by audio recordings of the description of the person with whom the participant is paired for that round of that game. Participants are provided with a "repeat" button if they want to hear the information again.

To ensure that the game instructions are well understood, the lab administrators review several examples with the participants. Then, the participants complete a series of test questions that they have to answer correctly prior to continuing with the task (listed in table C6). Laboratory staff are available to answer clarifying questions in both English and Swahili. We record the number of attempts a participant needs to correctly answer each comprehension question, as well as their first answer. We also track how often participants made use of the "repeat" button for audio recordings. A summary is provided in appendix C.4.

The average duration of each session, from the time the pill was taken to the end of the experimental games and survey questions, is 103 minutes. Details on the average duration of each component of the session are provided in figure A3. A detailed general overview of the sessions is provided in appendix A.4.

F. *Additional Considerations for the Experimental Protocol*

Until recently, studies of the effects of hydrocortisone focused almost exclusively on populations from wealthy countries of European descent (e.g., Kandasamy et al. 2014; Riis-Vestergaard et al. 2018). The literature suffered from the now well-recognized WEIRD (Western, educated, industrialized, rich, and democratic) bias in research, where non-European populations are excluded from the research process both as participants and as authors (Henrich, Heine, and Norenzayan 2010). One of the contributions of this study is to broaden the diversity and inclusiveness of the hydrocortisone literature to include non-WEIRD participants.

We assemble a team of coauthors from multiple disciplines (economics, neuroscience, psychiatry, psychology, and public health) and backgrounds, including those who work actively in the local community where we conducted field work. This collaboration helps ensure that our experiment minimizes risk and is culturally sensitive. The implementing staff and the nurses conducting the screening are exclusively Kenyan. In addition to receiving IRB approval from the universities of the principal investigators, we also obtained approval from the Kenya Medical Research Institute, the Kenyan government body overseeing medical research in Kenya; and the Kenya Pharmacy and Poisons Board, which is the Kenyan government body overseeing the administration of pharmaceutical substances in the country. The hydrocortisone dosage was chosen to be on the lower end of what is common in the psychology literature. A rigorous medical screening protocol was put into place, overseen by two physicians and administered by three full-time staff nurses, to minimize the risk of adverse reactions. A detailed hospital referral protocol with a partner hospital was in place for adverse events and side effects. No severe adverse events occurred as a result of the study. See appendix D for a detailed discussion of our protocols, ethical considerations, inclusion criteria, and adverse event reporting protocol.

III. Estimating Equations

We ask three questions. First, does hydrocortisone affect prosociality? Second, does coethnicity affect prosociality? Third, what is the interaction effect of coethnicity and hydrocortisone on prosociality? We will present two main specifications, both of which are as stated in the preanalysis plan.[8]

[8] The specifications below are mathematically equivalent to those in the preanalysis plan. We choose a slightly different representation to ease interpretation and minimize notation.

Our main estimating equation is as follows:

$$\begin{aligned} y_{ij} = {} & \beta_1 I_{ij}^{\text{coethnic}} + \beta_2 I_i^{\text{HC}} + \beta_3 I_{ij}^{\text{coethnic}} \times I_i^{\text{HC}} \\ & + \beta_4 I_{ij}^{\text{same gender}} + \beta_5 I_{ij}^{\text{same gender}} \times I_i^{\text{HC}} + \beta_6 I_{ij}^{\text{same age}} + \beta_7 I_{ij}^{\text{same age}} \times I_i^{\text{HC}} \quad (1) \\ & + \alpha_{e(i)} + \alpha_{e(j)} + \alpha_{g(i)} + \alpha_{g(j)} + \alpha_{a(i)} + \alpha_{a(j)} + \varepsilon_{ij}, \end{aligned}$$

where i indexes the player making the decision and j the other player in the game. The unit of observation is a choice made by a player, and we estimate the equation separately for each stage of each game (i.e., DG, TG1, TG2, etc.). The variable y_{ij} denotes the choice made by player i when paired with player j; I_i^{HC} is an indicator variable that equals 1 if player i is in the hydrocortisone treatment group; I_{ij}^{coethnic} is an indicator variable that equals 1 if the self-reported ethnicity of player i is the same as that of player j. The equation also includes fixed effects for the ethnicity of player i, which we denote as $\alpha_{e(i)}$, and fixed effects for the ethnicity of player j, which we denote as $\alpha_{e(j)}$.

The equation also includes several additional controls: gender fixed effects for both players, $\alpha_{g(i)}$ and $\alpha_{g(j)}$; age-group fixed effects for players i and j, which we denote $\alpha_{a(i)}$ and $\alpha_{a(j)}$; an indicator variable that equals 1 if player i and j belong to the same age group, $I_{ij}^{\text{same age}}$; and an indicator variable if they are both of the same gender, $I_{ij}^{\text{same gender}}$. In addition, we include the interactions of the two latter variables with the indicator for hydrocortisone treatment, I_i^{HC}, to allow the hydrocortisone effect to vary by same gender and same age. Standard errors for all estimates are clustered at the level of player i.

From the estimates of equation (1), we can calculate the average effect of hydrocortisone on prosocial behavior. This effect, evaluated at the means of the other variables in the model, is given by

$$\text{hydrocortisone effect} = \hat{\beta}_2 + \hat{\beta}_3 \bar{I}_{ij}^{\text{coethnic}} + \hat{\beta}_5 \bar{I}_{ij}^{\text{same gender}} + \hat{\beta}_7 \bar{I}_{ij}^{\text{same age}}, \quad (2)$$

where the $\bar{I}_{ij}^{k}$ indicators denote the share of observations that fall into the category denoted by the superscript.

The effect of coethnicity on prosocial behavior, evaluated at the mean of the treatment variable, is

$$\text{coethnicity effect} = \hat{\beta}_1 + \hat{\beta}_2 \bar{I}_{ij}^{\text{HC}}, \quad (3)$$

where $\bar{I}_{ij}^{\text{HC}}$ denotes the average share of observations for which player i received the hydrocortisone treatment.

The interaction between coethnicity and hydrocortisone is simply $\hat{\beta}_3$ from equation (1).

In auxiliary regressions, we also estimate a variant of equation (1) that includes player i fixed effects rather than controls for the player's ethnicity, age group, and gender. Since coethnic preferences are the differences in each player i's decisions when playing with players from other ethnic groups, isolating within-participant variation with player fixed effects can be viewed as a robustness check to our interpretation. We do not use it as our primary specification because the fixed effects absorb the uninteracted effect of the hydrocortisone treatment and therefore we would be unable to estimate the average hydrocortisone effect.

Choose your dictator game specification.—In this game, each participant chooses a dictator who may be from the same or a different ethnic group. We estimate coethnic preference using the following specification:

$$\begin{aligned} I_{ijk}^{\text{chosen}} = {} & \beta_1 I_{ij}^{\text{coethnic}} + \beta_3 I_{ij}^{\text{coethnic}} \times I_i^{\text{HC}} \\ & + \beta_4 I_{ij}^{\text{same gender}} + \beta_5 I_{ij}^{\text{same gender}} \times I_i^{\text{HC}} + \beta_6 I_{ij}^{\text{same age}} + \beta_7 I_{ij}^{\text{same age}} \times I_i^{\text{HC}} \\ & + \alpha_{e(j)} + \alpha_{g(j)} + \alpha_{a(j)} + \alpha_k + \varepsilon_{ijk}, \end{aligned} \tag{4}$$

where i indexes the participant who is making the choice and j indexes the other player (the one who may be chosen to be dictator). Decisions are indexed by k. The unit of observation is an option, defined as a person j who could be chosen to be the dictator by participant i in decision k. The other variable definitions are the same as in equation (1). We include fixed effects for the characteristics of player j: ethnicity $\alpha_{e(j)}$, age $\alpha_{a(j)}$, and gender $\alpha_{g(j)}$.

Unlike in equation (1), we cannot include player i characteristic fixed effects. This is because player i must always choose one of two options and the effect of player i characteristics are constant across the two options in each decision made by player i.

Since our interest is in the average effects, we estimate equation (4) using a linear probability model for simplicity (Gomila 2021). Results from using a conditional logit model, which we report in table B1, are virtually identical. This is not surprising since the average effects are evaluated at the mean where the logistic function is approximately linear (Angrist 2001; Angrist and Pischke 2009, chap. 3).

IV. Results

A. *Sample Balance*

Balance checks for the treatment and control groups are reported in table A3. Among the 21 characteristics examined, 18 of the treatment − control differences are insignificant using a 10% threshold, 20 are insignificant using a 5% threshold, and all are insignificant using a 1% threshold. A joint F-test across all characteristics is also not statistically significant at conventional levels. As we discuss below, our estimates are very similar when we control for participant characteristics that are not balanced across treatment and control.

B. Effectiveness of the Hydrocortisone Treatment

We now examine whether the hydrocortisone treatment effectively increases cortisol levels. Saliva samples are collected at six predetermined points during the experiment: immediately before administration of the pill and immediately before each stage of each game and the survey (see app. A.4). These samples are later analyzed for salivary cortisol levels.[9]

In figure 1, we summarize the effects of the hydrocortisone treatment on the measured cortisol levels. The six saliva samples ("salivettes") are shown on the *x*-axis. The *y*-axis reports cortisol levels, measured in nanomoles per liter (nmol/L). For each salivette, average cortisol levels are reported for the treatment group (solid red curve) and the placebo group (dashed blue curve). Also reported are 95% confidence bands. The top of the figure reports the experimental activities and their average duration for each of the two activity orders.

We find that the hydrocortisone pills are effective at elevating participants' cortisol levels and that this increase persists throughout the experimental session. The magnitude of this increase—from about 20 nmol/L to about 160 nmol/L—is similar to other studies of hydrocortisone administration (Margittai et al. 2018; Riis-Vestergaard et al. 2018), but, as is typical for hydrocortisone administration, somewhat larger than the effects of other stress induction protocols, such as the Trier social stress test (e.g., Haushofer et al. 2013; Vinkers et al. 2013).[10] Cortisol levels in the treatment and control groups are virtually identical before the pills are taken. The cortisol levels of the participants in the hydrocortisone group increase quickly after the pills are taken, while the cortisol levels

[9] Samples are collected using a "salivette," which is a small cotton swab stored in a plastic tube. Participants chew on the cotton swab for 1 minute before placing it back into the tube. Salivettes are stored at −25°C and then shipped to Technische Universität Dresden, Germany, using a cold chain. Analysis is performed using a chemiluminescence immunoassay (CLIA) with a sensitivity of 0.16 ng/mL (Immuno-Biological Laboritories, Hamburg, Germany).

[10] For example, the TSST in Vinkers et al. (2013) increased cortisol to 18 nmol/L from a baseline of 10 nmol/L; in Schweda et al. (2019), it led to a 5 nmol/L increase in cortisol; in Kirschbaum, Pirke, and Hellhammer (1993), to an increase from 10 to 16 nmol/L; in Singer et al. (2017), to an increase from 7 to 12 nmol/L; in von Dawans et al. (2012), to an increase of about 10 to 15 nmol/L. In contrast, hydrocortisone administration leads to larger increases in cortisol. Henckens et al. (2012) administer 10 mg of hydrocortisone, with an increase in cortisol from 10 to 45 nmol/L. Cornelisse et al. (2013) administer 10 mg of hydrocortisone with an increase of 2–4 nmol/L. Riis-Vestergaard et al. (2018) administer 10 mg and see an increase from 11 to 81.5 nmol/L. Kandasamy et al. (2014) administer 5 mg of hydrocortisone and achieve an increase of about 12.4 nmol/L. In an observational study, Chemin, de Laat, and Haushofer (2013) find that farmers in Kenya exposed to weather shocks have an increase of 39.6 nmol/L.

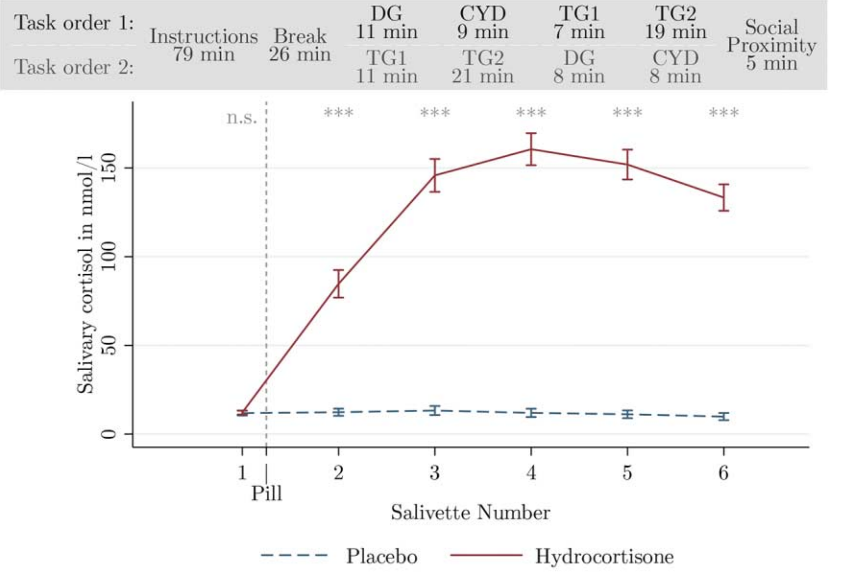

FIG. 1.—Cortisol levels over time by treatment. The figure depicts salivary cortisol levels during the study. The top bar describes the study activities. When applicable, we distinguish between two task orders. The study consisted of oral instructions leading up to the first salivette (before pill), followed by the pill and a break, before the five tasks in their respective order, with the social proximity survey last. The average duration of each activity is shown in minutes below each study activity. The games are abbreviated as dictator game (DG), choose your dictator game (CYD), trust game–stage 1 (TG1), and trust game–stage 2 (TG2). The horizontal axis marks the six saliva samples and indicates when the pill (20 mg hydrocortisone or placebo) was taken. The vertical axis shows the salivary cortisol concentration in nanomoles per liter (nmol/L). The two curves depict average cortisol levels by treatment arm over the course of the study, with 95% confidence intervals denoted by vertical lines, and the statistical significance levels of a t-test of the treatment − control difference above, shown as n.s. for not statistically significant and *** for $p < .01$.

of those in the placebo group remain practically constant. Tests of the equality of cortisol levels between the treatment and control groups reject the null hypothesis of equal levels at the 1% level for all points in time after the pill is taken.

C. *Average Differences by Hydrocortisone, Coethnicity, and Their Interaction*

We now turn to an initial examination of how prosociality is affected by the hydrocortisone treatment and by the ethnicity of the other player. A comparison of the average outcomes in each of our cross-randomized treatments is presented in figure 2. Gray bars represent the mean outcome for coethnic pairs, and white bars represent the mean outcome

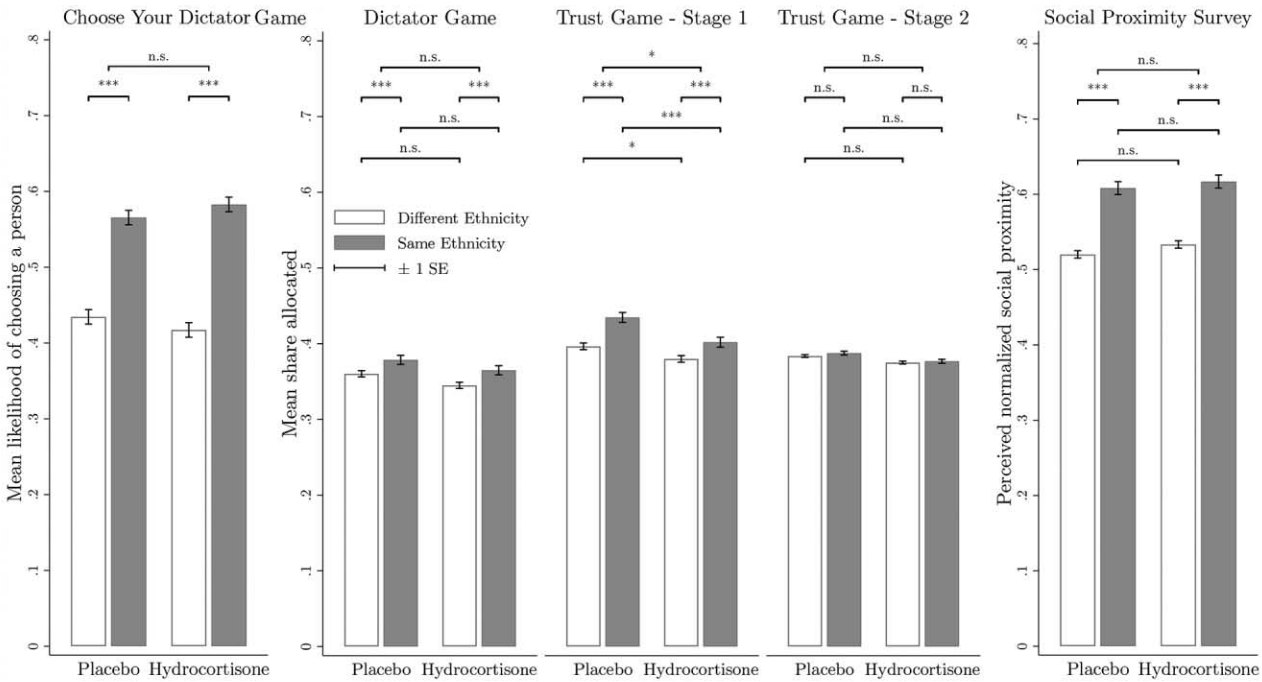

FIG. 2.—Comparison of means across treatment categories. The figure compares game outcomes across the four possible treatment combinations. For each of the games, the left two bars show results for participants under placebo and the right bars under hydrocortisone. White bars symbolize noncoethnic pairings and gray bars coethnic pairings. The vertical brackets show ±1 standard errors of the mean. At the top of the figure, we report tests across pills (bottom two rows of brackets), across pairings (second row), and of the interaction (top row). Statistical significance levels are shown as n.s. for not statistically significant, * for $p < .1$, and *** for $p < .01$.

for non-coethnic pairs. The means are reported separately for the participants in the hydrocortisone and placebo groups. The reported *y*-axis for the CYD game is the average probability of choosing a coethnic dictator (gray bar) or non-coethnic dictator (white bar). For the DG, TG1, and TG2, the *y*-axis represents the shares allocated to coethnic and non-coethnic partners. Finally, for the social proximity survey, the *y*-axis represents the perceived proximity of coethnic and non-coethnic individuals, averaged across the three measures of social proximity. The error bars correspond to 1 standard error of the mean.

Horizontal lines and asterisks above the bars report the levels of statistical significance of pairwise group comparisons. The third and fourth rows report the difference between the hydrocortisone and control treatments for coethnic pairings (third row) and non-coethnic pairings (fourth row).[11] As shown, in general we do not find differences in behavior for participants in the hydrocortisone and placebo groups. The one exception is for the first stage of the trust game, where we find participants allocate less money to the other player.

The second row reports the difference in behavior when participants are paired with a coethnic and non-coethnic player. This is reported separately (on the same row) for the hydrocortisone and placebo conditions. In contrast to the mixed hydrocortisone differences, we find clear evidence of a coethnic preference. Participants are more likely to choose coethnics than non-coethnics in the CYD game, they allocate more money to coethnics than non-coethnics in the DG and in TG1, and they report being closer to coethnics than non-coethnics in the social proximity survey. The one exception is that we do not observe a difference in behavior in TG2, the second stage of the trust game. Thus, all forms of prosociality, except trustworthiness, are greater among coethnic players.[12]

The final comparison of interest is given by the top row, which reports the levels of significance of the difference between the difference in the average for coethnic and non-coethnic pairings for participants in the hydrocortisone treatment group relative to those in the placebo group. This statistic provides insight into whether our documented coethnic preferences are affected by the hydrocortisone treatment. We find no compelling evidence of such an effect. For four of the five outcomes examined the second-difference effect is statistically insignificant at conventional levels. It is significant at the 10% level for the TG1.

[11] This comparison is not possible for the CYD game because our identification strategy for outcomes in this game relies on the comparison between the two potential dictators, and a participant's individual characteristics, including their hydrocortisone treatment status, are therefore not identified.

[12] We discuss explanations for the lack of coethnic preference in TG2 in further detail below.

Overall, our preliminary examination of the mean outcomes of the different treatment groups suggests strong evidence for coethnic preference, but a limited effect of hydrocortisone on prosocial behavior or on the strength of coethnic preference.

D. Estimated Effects

We now turn to estimates of equation (1), which we report in table 1. We present the estimated joint effects for the three hypotheses in table 2.

1. The Effect of Hydrocortisone

Panel A of table 2 reports the average effect of hydrocortisone on participants' behavior, as in equation (2).[13] Averaged across coethnic and non-coethnic interactions, we find that hydrocortisone leads to reductions in allocations to the other player of 1.6 percentage points in the DG and 2.3 percentage points in TG1. The estimates are statistically significant at the 10% and 5% levels, respectively. Relative to the sample means, which are reported at the top of the table, the effects correspond to reductions of 4.4% ($-0.0157/0.3591 = -0.0437$) and 5.6% ($-0.0225/0.3984 = -0.0564$). These results suggest hydrocortisone may reduce altruism and trust.

We find no effect of hydrocortisone on the amount given in the TG2 or on self-reported social proximity to the other player. Both estimates are small in magnitude and not statistically different from zero. Social proximity is measured as an average of the three survey questions. As we report in table B3 and figure B1, the estimates are similar if we examine each of the three questions separately.

2. The Effect of Coethnicity

We next turn to the average effect of coethnicity, which is given by equation (3). As reported in panel B of table 2, we find evidence of coethnic preference in the CYD, DG, TG1, and social proximity questions. The estimated effects are all positive and statistically significant at the 1% level. To assess the magnitude of the effects, we compare the joint coefficients presented in panel B to the sample means stated at the top of the table.

[13] The reported estimates cluster standard errors at the participant level. We find that we obtain almost identical statistical precision if we instead use randomization inference to calculate *p*-values. Table C1 reports the estimates when we reassign all levels of treatment assignment (hydrocortisone, ethnicity, gender, and age pairing) randomly 10,000 times. We calculate the *p*-value by finding the proportion of the randomization distribution that is larger than our observed test statistic.

TABLE 1
BASELINE REGRESSION ESTIMATES

	Choose Your Dictator Game (1)	Dictator Game (2)	Trust Game 1 (3)	Trust Game 2 (4)	Social Proximity (5)
Constant	.6231***	.3445***	.3289***	.3973***	.5406***
	(.0061)	(.0230)	(.0228)	(.0197)	(.0237)
Hydrocortisone		−.0062	−.0321***	−.0139	.0206
		(.0114)	(.0117)	(.0110)	(.0132)
Same ethnicity	.0611***	.0181***	.0385***	.0025	.0841***
	(.0097)	(.0055)	(.0060)	(.0029)	(.0072)
Hydrocortisone × same ethnicity	.0228*	.0014	−.0178**	−.0019	−.0051
	(.0135)	(.0074)	(.0084)	(.0042)	(.0101)
Same gender	−.0104	.0041	−.0135*	.0053	.0376***
	(.0094)	(.0069)	(.0073)	(.0058)	(.0087)
Hydrocortisone × same gender	−.0327**	−.0125	.0154	−.0033	−.0201*
	(.0132)	(.0096)	(.0101)	(.0085)	(.0122)
Same age group	.0444**	.0111	.0325*	−.0236*	−.0106
	(.0219)	(.0175)	(.0166)	(.0142)	(.0181)
Hydrocortisone × same age group	−.0400***	−.0110	.0236**	.0177*	.0096
	(.0145)	(.0107)	(.0113)	(.0092)	(.0129)
Matched player is Luo	−.0091	−.0042	.0082	−.0008	.0074
	(.0081)	(.0051)	(.0059)	(.0037)	(.0055)
Matched player is Kikuyu	−.0318***	−.0015	.0054	−.0051	−.0019
	(.0082)	(.0052)	(.0057)	(.0037)	(.0053)

Matched player is Luhya	−.0220***	−.0012	.0066	.0025	.0138**
	(.0083)	(.0052)	(.0057)	(.0037)	(.0058)
Matched player is female	.0054	.0054	.0119**	.0005	.0124**
	(.0066)	(.0048)	(.0051)	(.0042)	(.0061)
Matched player is middle-aged	.0318	.0198	.0550***	−.0015	−.0366**
	(.0214)	(.0164)	(.0157)	(.0141)	(.0169)
Matched player is old	.0364	.0307*	.0793***	−.0066	−.0440**
	(.0225)	(.0167)	(.0162)	(.0145)	(.0178)
Decision maker is Luo		.0116	.0047	−.0017	.0059
		(.0163)	(.0155)	(.0157)	(.0164)
Decision maker is Kikuyu		.0167	.0273	.0120	.0025
		(.0173)	(.0168)	(.0170)	(.0176)
Decision maker is Luhya		−.0199	.0007	−.0242*	−.0173
		(.0150)	(.0145)	(.0144)	(.0151)
Decision maker is female		−.0130	.0026	.0058	−.0285***
		(.0094)	(.0093)	(.0094)	(.0100)
Decision maker is middle-aged		.0501**	.0141	.0490**	−.0095
		(.0242)	(.0195)	(.0221)	(.0261)
Participants	1,784	1,784	1,784	1,784	1,784
Decisions per participant	6	6	6	30	4
Decisions	10,704	10,704	10,704	53,520	7,136

NOTE.—This table presents the full regression output of our main specification. All terms are described in the text. In this specification, we control for the interaction of a same-gender and same-age-group indicator with hydrocortisone as well as gender, age group, and ethnicity fixed effects of both players. Social proximity refers to the average measures of likelihood to be friends, trust, and closeness rescaled to lie between 0 and 1. Standard errors clustered at the participant level are reported in parentheses.

* $p < .1$.

** $p < .05$.

*** $p < .01$.

TABLE 2
Effects of Hydrocortisone, Coethnicity, and Their Interaction

	Choose Your Dictator Game (1)	Dictator Game (2)	Trust Game 1 (3)	Trust Game 2 (4)	Social Proximity (5)
Sample mean	.5740	.3591	.3984	.3804	.5481
Sample standard deviation	(.4945)	(.2490)	(.2670)	(.2477)	(.2575)
	A. Average Hydrocortisone Effect[a]				
Hydrocortisone effect		−.0157*	−.0225**	−.0103	.0123
		(.0093)	(.0093)	(.0095)	(.0099)
	B. Average Coethnicity Effect[b]				
Coethnicity effect	.0724***	.0188***	.0297**	.0015	.0817***
	(.0069)	(.0038)	(.0043)	(.0022)	(.0052)
	C. Interaction of Hydrocortisone and Coethnicity[c]				
Interaction effect	.0228*	.0014	−.0178**	−.0019	−.0051
	(.0135)	(.0074)	(.0084)	(.0042)	(.0101)
Participants	1,784	1,784	1,784	1,784	1,784
Decisions per participant	6	6	6	30	4
Decisions	10,704	10,704	10,704	53,520	7,136

Note.—All terms are described in the text. In this specification, we control for the interaction of a same-gender and same-age-group indicator with hydrocortisone as well as gender, age group, and ethnicity fixed effects of both players. The sample mean and standard deviation for the choose your dictator game refer to the share of decisions in which an in-group member was chosen among the decisions where one was available. Social proximity refers to the average measures of likelihood to be friends, trust, and closeness rescaled to lie between 0 and 1. Standard errors clustered at the participant level are reported in parentheses.

[a] The average hydrocortisone effect is calculated as in eq. (2).

[b] The average coethnicity effect is calculated as in eq. (3).

[c] The interaction effect is given by $\hat{\beta}_3$ in eq. (1).

* $p < .1$.

** $p < .05$.

*** $p < .01$.

The estimates for the CYD game, reported in column 1, imply that participants are 7.2 percentage points more likely to choose a player of their own ethnicity. Since the probability of any player being chosen is 50% by construction, this effect is equal to 14.5% (0.0724/0.5 = 0.145) of the sample mean. For the DG, reported in column 2, participants allocate 1.88 percentage points more to coethnic individuals, which is 5.2% (0.0188/0.3591) of the sample mean. In the TG1, reported in column 3, participants give 2.97 percentage points more of their endowment to coethnic individuals, an amount equal to 7.5% (0.0297/0.3984 = 0.0745) of the sample mean.

In contrast to the findings for the DG and TG1, for the TG2, we find no evidence of a coethnic effect (see col. 4).[14] This is surprising, especially given that we find a coethnic effect in the DG, which is similar to the TG2. In both the DG and TG2, the participant chooses the amount of an endowment to allocate to the other player. There are several potential explanations for the difference. A key difference between the TG2 and the DG is that in the TG2, there is a history of interactions that precede the participant's decision. This is not the case in the DG. In particular, it is possible that coethnicity affects altruism as measured by the DG but does not affect reciprocity conditional on an initial display of trust (i.e., through player 1's allocation).

Another potential explanation for the lack of a coethnicity effect in the TG2 is that participant fatigue may be biasing the estimates toward zero. In the TG2, we used the strategy method. Thus, participants had to make a decision for every possible allocation they could have received. Because they had to decide how much to send back to player 1 for each of the five possible player 1 allocations (KES 40, 80, 120, 160, or 200) for six iterations of the game, each participant made 30 decisions in the TG2. The concern is that making so many choices may lead to fatigue. Participants may have begun to make decisions very quickly without mentally registering the identity of the other player or the amount of the TG1 allocation.

An examination of the data reveals patterns that are consistent with player fatigue. Recall, participants are presented with each of the possible amounts they could receive from player 1—KES 40, 80, 120, 160, or 200—and choose the share to return to player 1 for each of these amounts. First, the average share returned by player 2 does not depend on the amount sent by player 1. Regardless of the initial allocation by player 1, the average share returned by player 2 is always between 37% and 39%.[15] This is consistent with participants choosing the same position on the slider each time they are asked about a different amount sent by player 1.

Second, we can examine how quickly participants respond to each of the five amounts player 1 can allocate to player 2. Participants are always presented with the five possible player 1 allocations in the same order, starting with KES 40 and ending with KES 200. We find that average response times decrease as participants respond to each subsequent amount; that is, participants answer more and more quickly. Average response times (in seconds) are 28.7 (for KES 40), 19.4, 18.7, 17.1, and 15.1 (for KES 200).

[14] The results are similar when we include fixed effects for the amount that player 1 sends to player 2, and when we estimate effects separately for each amount sent by player 1. These estimates are reported in table B2.

[15] See table B2.

Third, we also find that the variation in participants' chosen share returned decreases monotonically with each additional player 1 allocation they are asked to respond to. For all players, we calculate the standard deviation of the share returned for each of the five player 1 allocations they respond to. When we look at the average standard deviation across all players for each amount they respond to, we find that the variation declines monotonically: 0.124, 0.113, 0.109, 0.105, and 0.103. This indicates that the effect of the other player's identity on the amount allocated is declining as the participant is asked to make additional choices. Also, note that the average standard deviation of participants' choices in the TG2 is lower for all possible player 1 allocations than it is for the DG, which has an average standard deviation of 0.133.

As with CYD, DG, and TG1, for the social proximity survey, we also find a coethnic effect (see col. 5). The coefficient estimate is 0.08 and is statistically significant at the 1% level. The estimated effect is sizeable and equal to 14.9% (0.0817/0.5481 = 0.0745) of the sample mean.

3. The Interaction of Hydrocortisone and Coethnicity

The estimated interaction effect between hydrocortisone and coethnicity is reported in panel C of table 2. Since the effect is just $\hat{\beta}_3$ from equation (1), the effect in the panel restates this coefficient from table 1. In general, we do not find a consistent pattern in the estimated effects. For most outcomes—DG in column 2, TG2 in column 4, and social proximity in column 6—the interaction effect is small in magnitude and not statistically different from zero. For the CYD and TG1, we find effects that are statistically different from zero although the significance is marginal and the signs are different for the two games. In the CYD, the interaction effect is positive and statistically significant at the 10% level (col. 1), while in the TG2, the interaction effect is negative and statistically significant at the 5% level (col. 3).

Taken together, we interpret these results as mixed and inconclusive. We do not find systematic evidence that hydrocortisone increases the effect of coethnicity on prosocial behavior.

4. Average Effects across All Outcomes

Our analysis examines multiple measures, each of which is intended to capture the same underlying outcome, prosocial preferences of one player toward another. In an attempt to synthesize the effects across all outcomes of interest, we estimate average effects across decisions in the DG, TG1, TG2, and social proximity survey. For each game, the outcomes are measures that range from 0 to 1. The average effects are estimated using

TABLE 3
AVERAGE EFFECTS ACROSS ALL OUTCOMES

	(1)	(2)	(3)	(4)	(5)	(6)
Sample mean	.3799	.3793	.3946	.4215	.4173	.4215
Sample standard deviation	(.2509)	(.2552)	(.2559)	(.2660)	(.2751)	(.2806)
	A. Average Hydrocortisone Effect[a]					
Hydrocortisone effect	−.0126	−.0158**	−.0104	−.0088	−.0070	−.0088
	(.0084)	(.0079)	(.0078)	(.0067)	(.0071)	(.0067)
	B. Average Coethnicity Effect[b]					
Coethnicity effect	.0081***	.0167***	.0135***	.0309***	.0223***	.0309***
	(.0018)	(.0022)	(.0018)	(.0021)	(.0018)	(.0021)
	C. Interaction of Hydrocortisone and Coethnicity[c]					
Interaction effect	−.0035	−.0057	−.0042	−.0071*	−.0052	−.0071*
	(.0035)	(.0042)	(.0034)	(.0042)	(.0036)	(.0042)
Weights	No	Yes	No	Yes	No	Yes
Social proximity	No	No	Index	Index	Full	Full
Participants	1,784	1,784	1,784	1,784	1,784	1,784
Decisions per participant	42	42	46	46	54	54
Decisions	74,928	74,928	82,064	82,064	96,336	96,336

NOTE.—All terms are described in the text. In this specification, we control for the interaction of a same-gender and same-age-group indicator with hydrocortisone as well as gender, age group, and ethnicity fixed effects of both players. This table reports the results of stacking the data of the dictator and trust games (cols. 1 and 2), as well as the social proximity survey index (cols. 3 and 4) or the social proximity survey questions separately (cols. 5 and 6), respectively. Odd-numbered columns use the data as is while even-numbered columns weight all games to have equal influence despite trust game stage 2 having 30 observations per participant, while the dictator game and trust game stage 1 have 6 observations per participant, the social proximity index 4, and the individual questions 12. Standard errors clustered at the participant level are reported in parentheses.

[a] The average hydrocortisone effect is calculated as in eq. (2).
[b] The average coethnicity effect is calculated as in eq. (3).
[c] The interaction effect is given by $\hat{\beta}_3$ in eq. (1).
* $p < .1$.
** $p < .05$.
*** $p < .01$.

stacked data (i.e., multiple outcomes per participant), while controlling for game fixed effects and clustering standard errors at the participant level.

The estimates are reported in table 3. Each column reports estimates using a slightly different methodology for calculating the average effect.[16] The odd-numbered columns report estimates where each observation is

[16] Since the calculation of average effects was not prespecified, we present results with different strategies for calculating average effects. Our conclusions do not depend on the exact specification chosen.

given equal weight, which implicitly gives relatively more weight to games like the TG2 that has more observations per participant because of our use of the strategy method. The even-numbered columns weigh each observation so that each game is given equal weight. Columns 1 and 2 include the behavioral games only (i.e., excluding the social proximity survey). The remaining columns include the social proximity questions, either as a single average index (cols. 3 and 4) or including each question individually (cols. 5 and 6).

We obtain conclusions that are consistent across all specifications. We find robust evidence for a strong effect of coethnicity on prosocial behavior. Consistent with the game-specific estimates, the hydrocortisone effect is negative on average but generally insignificant at conventional levels. We also find the same for the interaction effect. Thus, while we find strong evidence for coethnic preferences, we find no evidence that these are greater under stress.

5. Multiple Hypothesis Testing

Our analysis examines three hypotheses related to the effects of coethnicity, hydrocortisone, and their interaction. Given that we are testing multiple hypotheses in our study, we test the sensitivity of our conclusions to correcting for multiple hypothesis testing. We follow the procedure outlined by List, Shaikh, and Xu (2019), correcting for the fact that we test three distinct hypotheses. We reproduce our baseline estimates from table 2 with the correction and report these in table B4. Our conclusions remain unchanged. We continue to find a nonrobust and weak hydrocortisone effect, a strong and robust coethnicity effect, and no evidence for a robust and consistent interaction effect.

6. Including Player 1 Fixed Effects

In addition to the main specification, equation (1), we estimate a similar specification with player 1 fixed effects in place of controls for player 1 characteristics. This checks that the results for the average coethnic effect and the interaction effect of coethnicity and hydrocortisone are not driven by omitted player 1 characteristics (i.e., features of the player that we do not observe and thus cannot control for in the main specification). The estimated effects, which are reported in table 4, show that our findings are similar with these alternative controls.[17]

Panel A of table 4 shows that the average coethnic effects are similar in sign, magnitude, and precision to the main specification in panel B of

[17] The full estimates are reported in table B5.

TABLE 4
EFFECTS OF HYDROCORTISONE, COETHNICITY, AND THEIR INTERACTION: WITH PLAYER 1 FIXED EFFECTS

	Dictator Game (1)	Trust Game 1 (2)	Trust Game 2 (3)	Social Proximity (4)
Sample mean	.3591	.3984	.3804	.5481
Sample standard deviation	(.2490)	(.2670)	(.2477)	(.2575)
	A. Average Coethnicity Effect[a]			
Coethnicity effect	.0184***	.0301***	.0028	.0819***
	(.0038)	(.0043)	(.0022)	(.0052)
	B. Interaction of Hydrocortisone and Coethnicity[b]			
Interaction effect	.0015	−.0178**	−.0021	−.0052
	(.0074)	(.0084)	(.0041)	(.0101)
Participants	1,784	1,784	1,784	1,784
Decisions per participant	6	6	30	4
Decisions	10,704	10,704	53,520	7,136

NOTE.—The table reports estimates of a version of eq. (1) with player 1 fixed effects. Variables are as described in the text. Social proximity refers to the average measures of likelihood to be friends, trust, and closeness rescaled to lie between 0 and 1. Standard errors clustered at the participant level are reported in parentheses.
[a] The average coethnicity effect is calculated as in eq. (3).
[b] The interaction effect is given by $\hat{\beta}_3$ in eq. (1).
** $p < .05$.
*** $p < .01$.

table 2. Panel B of table 4 shows that the interaction effect is similarly inconclusive to that in panel C of table 2.[18]

E. *Experimenter Demand Effects*

One possible concern is that our results may be influenced by experimenter demand effects. Because participants know they are being studied, they may change their behavior to conform with what they perceive to be the experimenters' expectation. This is particularly important because we communicated the ethnicity of the partner to participants more directly than previous studies. In addition, participants also know that the study is about stress hormones, and therefore beliefs related to the pharmacological treatment may also affect behavior.

There are several reasons why we believe that it is unlikely that demand effects are a significant threat to our results. Due to the experimental design—specifically our use of hydrocortisone—the stress effect cannot be confounded by experimenter demand since there is no way for participants to know which of the two identical pills (hydrocortisone or placebo)

[18] Individual fixed effects are not identified in the choose your dictator game for reasons discussed earlier. Similarly, we cannot estimate the direct (uninteracted) hydrocortisone effect because it is absorbed by player 1 fixed effects.

they were given. To confirm this, at the end of the laboratory session, we ask every participant to guess which of the two pills they received. The guess is incentivized with KES 50 (about USD 0.50). If there is a way for participants to guess their treatment status based on physiological reactions or other observations, we expect these guesses to be correct for significantly more than 50% of the participants. Instead, participants guess correctly in only 48.9% of cases, which is not significantly different from a random guess at conventional levels.[19] Thus, because participants are unaware of their hydrocortisone treatment status, the hydrocortisone treatment effect cannot be affected by experimenter demand.

There is also the possibility of demand effects that bias the estimated coethnicity effects. Our protocols are designed to minimize such effects. The ethnicity of the other player is only one of several characteristics reported to the participant. This feature of the design is intended to obfuscate the study's interest in ethnicity, while precisely transmitting information about the ethnicity of the other player. The focus on ethnicity is also obscured by the fact that participants' interactions with coethnic and non-coethnic partners were randomly ordered.[20]

If demand effects are present, we expect that they likely work to attenuate coethnicity effects. Kenya has robust social norms against discrimination based on ethnicity, and the perception is that the population in general, including Westerners and Christians, looks down on behavior that favors coethnics. Thus, demand effects, if present, would likely lead to a compensatory increase in prosocial behavior when playing with a non-coethnic partner, which would bias our observed ethnicity effects downward. This is consistent with evidence from Blum, Hazlett, and Posner (2021), which indicates that demand effects and social desirability induce a downward bias in the estimated effect of coethnicity.

As a formal assessment of the importance of demand effects in our setting, we undertake a direct test using the method proposed by De Quidt, Haushofer, and Roth (2018). The intuition of this method is to explicitly tell participants what the experimenter expects and measure the behavioral response to this information. If this behavioral response is small, demand effects are less of a concern. In our version of the method, participants play an additional single round of the DG at the very end of the experiment. Participants are randomly placed into one of two groups

[19] We also ask respondents to report why they guessed they were in the hydrocortisone or placebo treatment. Individuals in treatment are no more likely to report physiological or psychological symptoms as the reason for their guess.

[20] In a pilot session with Busara staff who are trained in the design of experimental games, we probed whether they noticed any patterns in the matching. They did not observe any patterns and, importantly, they did not recognize that pairings were stratified by coethnic status.

(with equal probability) for this final activity. In one group, we provide basic instructions, simply explaining the additional round of play. In the second group, we also add the following statement (in English and Swahili): "You are participating in this study in groups. We expect that people in your group will give more than they otherwise would." The statement is self-referential and allows participants to interpret it with reference to any group that they might have in mind. As argued by De Quidt, Haushofer, and Roth (2018), testing how the allocated amount varies between the two groups allows one to gauge the size of experimenter demand effects in a study, that is, the extent to which participants change their behavior because of a belief about what the experimenter expects from them.

We report the results of this exercise in table 5. Each column reports estimates from a specification where the dependent variable is the share allocated to the other player in the dictator game. The independent variable is an indicator variable for whether the participant received the "demand treatment." The coefficient for this variable, therefore, measures how much behavior changes when experimenter expectations are made explicit. The first column reports estimates for the full sample of 1,784 participants. We find no evidence of demand effects in our setting. The estimated coefficient, which suggests that the treatment decreases the share given by 0.0047 (0.47 percentage points), is very small in magnitude and statistically indistinguishable from zero.

To check for the possibility that demand effects might only be present for subsets of participants or interactions, for example, participants who receive the hydrocortisone treatment or who were paired with a coethnic partner, we reestimate the specification from column 1 for the following subsamples: individuals who receive the hydrocortisone treatment, column 2; individuals in the control group, column 3; individuals (randomly)

TABLE 5
Experimenter Demand Effects

	Share Allocated in Experimenter Demand Module				
	Full Sample (1)	Hydrocortisone Pill (2)	Placebo Pill (3)	Coethnic Pairing (4)	Non-Coethnic Pairing (5)
Sample mean	.4256	.4283	.4229	.4213	.4350
Sample standard deviation	(.2754)	(.2736)	(.2772)	(.2736)	(.2789)
Demand treatment	−.0047	−.0091	−.0005	.0014	−.0106
	(.0114)	(.0159)	(.0164)	(.0161)	(.0161)
Participants/ decisions	1,784	906	878	858	926

Note.—The table presents the results of the experimenter demand module. Participants played an additional round of the dictator game and were randomly allocated to either an explicit experimenter demand or not. Standard errors are reported in parentheses.

paired with a coethnic partner in the demand effects dictator game, column 4; and individuals paired with a non-coethnic partner, column 5. The estimated effects for each of the subsamples are small in magnitude and insignificant at conventional levels. In addition, three of the four demand effects, as well as the effect for the full sample, have negative signs, suggesting that if anything, increased salience of demand leads to lower levels of prosocial behavior.

The lack of evidence for demand effects is consistent with the fact that we identify very similar coethnicity effects across our range of outcomes. In the experimental design, we intentionally included a range of activities that we expect to be more or less susceptible to experimenter demand. In addition to behavioral games, we also administered our social proximity survey after all games were completed. Because the surveys ask directly about perceptions of the other players (who are from different ethnic groups), we except this measure to be the most strongly biased if demand effects are present. However, consistent with an absence of demand effects, we find very similar coethnicity effects when this outcome is examined.

Overall, our findings suggest that it is very unlikely that demand effects are a factor in explaining our results.

V. Robustness and Sensitivity Checks

We report estimates from a series of prespecified sensitivity tests to assess the robustness and stability of the findings. A summary of how estimates of the three effects of interest (hydrocortisone, coethnicity, and their interaction) varies across the different robustness checks is provided in figures 3–5. Each figure reports estimates and confidence intervals for an effect of interest from different specifications. The estimates are shown in ascending order based on their magnitude. The bottom of each figure reports information on the specification and how it differs from the baseline specification. For comparison, the figures also report the baseline estimates. The estimated effects from equation (1) are colored maroon. The estimates with player 1 fixed effects (reported in table 2) are also included in the figures.

The first sensitivity check that we perform is that we vary our definition of coethnicity. For our main estimates, we define our coethnicity indicator variable I_{ij}^{coethnic} as players i and j having the same mother tongue. However, previous papers have used an alternative definition of coethnicity based on political coalitions. We thus present an alternative definition of coethnicity, defining player i and j as members of the same group if they belong to the same political coalition.

Political coalitions shift over time. During the 2007 election, the Kamba ethnic group supported the Kikuyu candidate against a coalition of Luo

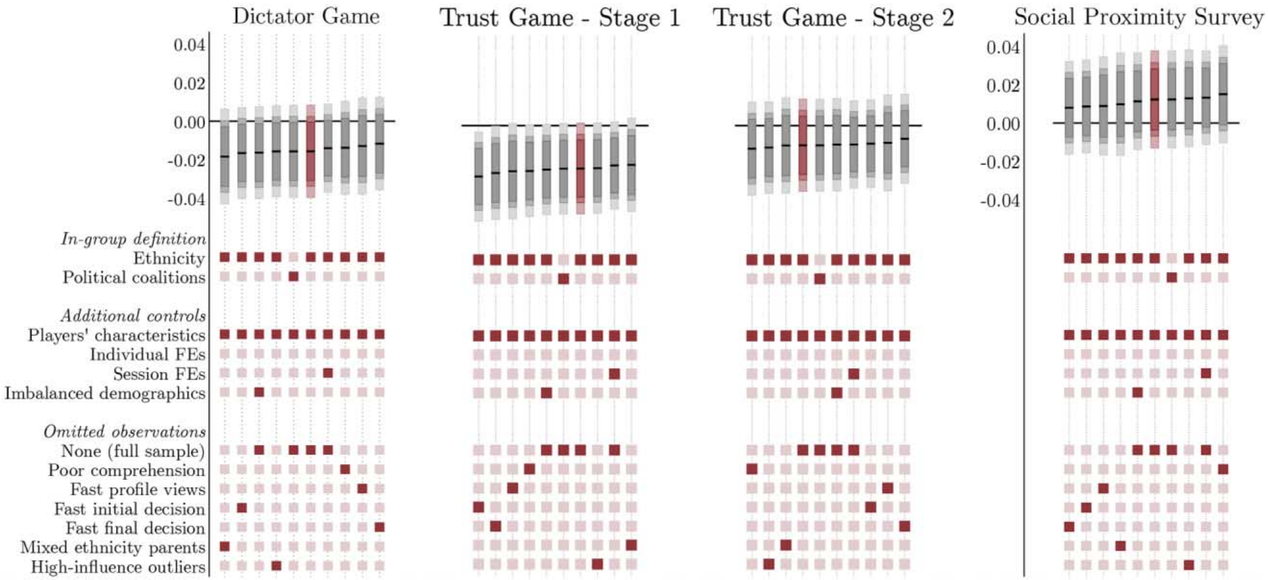

FIG. 3.—Robustness of the hydrocortisone effect. The figure shows the coefficients and standard errors for the effect of hydrocortisone on behavior in each game for alternative specifications. Coefficients are depicted by black horizontal lines. The vertical bars, from darkest to lightest, denote the 90%, 95%, and 99% confidence intervals, respectively. The red bar indicates our main specification. The bottom panel indicates the combination of robustness checks associated with each specification. The effect of hydrocortisone is not identified in the choose your dictator game, nor in individual fixed-effects specification, and these are therefore not displayed here. FEs = fixed effects.

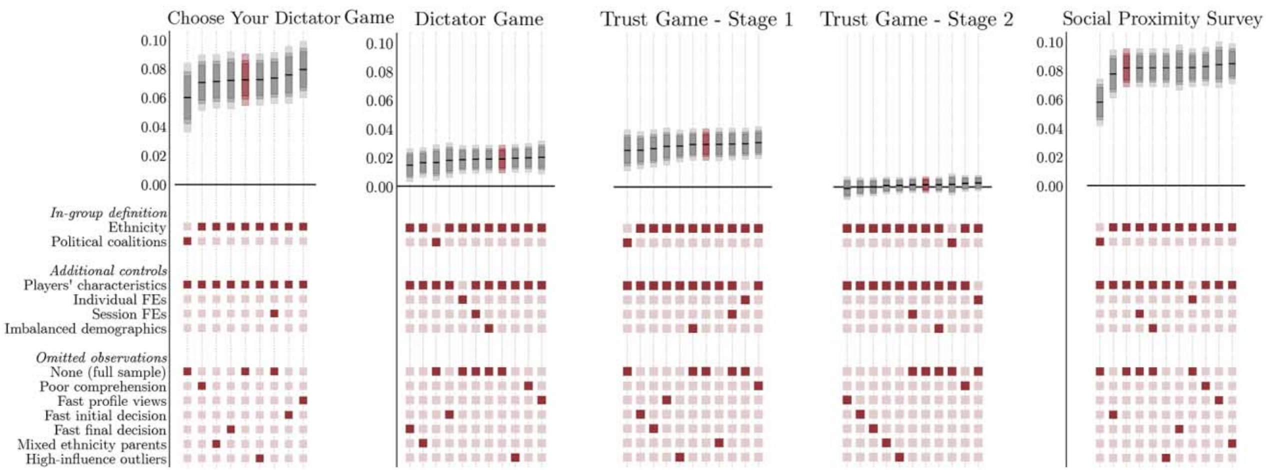

FIG. 4.—Robustness of the coethnicity effect. The figure shows the coefficients and standard errors for the effect of coethnicity on behavior in each game for alternative specifications. Coefficients are depicted by black horizontal lines. The vertical bars, from darkest to lightest, denote the 90%, 95%, and 99% confidence intervals, respectively. The red bar indicates our main specification. The bottom panel indicates the combination of robustness checks associated with each specification. FEs = fixed effects.

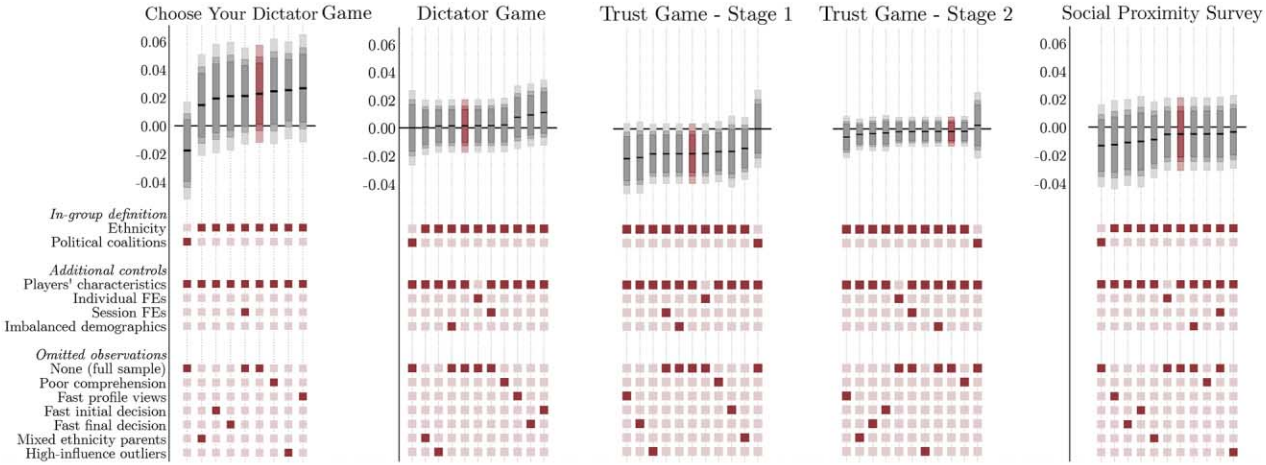

FIG. 5.—Robustness of the interaction of coethnicity and hydrocortisone. The figure shows the coefficients and standard errors for the interaction effect of coethnicity and hydrocortisone on behavior in each game for alternative specifications. Coefficients are depicted by black horizontal lines. The vertical bars, from darkest to lightest, denote the 90%, 95%, and 99% confidence intervals, respectively. The red bar indicates our main specification. The bottom panel indicates the combination of robustness checks associated with each specification. FEs = fixed effects.

and Luhya. This is the coalition that has been used to define groups in previous studies (e.g., Hjort 2014; Berge et al. 2020). After 2007 and during the 2013 and 2017 elections, the Kamba joined the Luo-Luhya coalition (Ferree, Gibson, and Long 2014). This is the coalition structure that was present during our experiment.

As prespecified, we check the sensitivity of our findings to defining coethnicity using the coalition structure that emerged after the 2013 elections. We use an indicator variable that equals 1 if player i and j both belong to an ethnic group that is part of the same political coalition. The estimated effects are reported in table C2.[21]

We check the robustness of the estimates to the inclusion of additional covariates. We control for session fixed effects, which capture factors that are potentially important for prosocial behavior, such as the day of the week, the time of day, the temperature at the time, recent political events, laboratory staff behavior, and so forth (table C4). We also check the sensitivity of the estimates to the inclusion of participant characteristics that are not balanced across the hydrocortisone and placebo treatments (table C5). This check was not prespecified.[22]

We also examine the sensitivity of our findings to omitting observations. In any experimental setting, there is concern about poor respondent comprehension, lack of focus by respondents, and respondent fatigue. Motivated by this, we check the sensitivity of our findings to the omission of observations with lower comprehension, as measured by a correct initial responses to comprehension questions (table C6). On average, participants initially answer incorrectly in 14% of the comprehension questions they are asked.[23] In the sensitivity test, we remove all participants who provided incorrect answers on their first attempt for more than half of the comprehension questions asked about that game.[24] For the social proximity survey, since there are no comprehension questions, we exclude participants who answered incorrectly on their first attempt

[21] As noted, earlier studies use the coalition structure from the 2007 election (Luo and Luhya in a coalition and Kikuyu and Kamba in another). For comparability to these studies and because ethnic preference may be affected by previous coalitions, we also report estimates using the 2007 coalitions. These are reported in table C3. We obtain similar estimates. In both cases, the findings appear to be driven by an own-ethnicity effect. If we include both an own-ethnicity indicator and an own-coalition but different-ethnicity indicator in our regressions, the former effect is strong and the latter effect tends to be small and statistically insignificant at conventional levels.

[22] Since a within-player choice is made in every round of the choose your dictator game, similar to when we added player 1 fixed effects, player characteristics are absorbed and do not affect the estimates. We therefore do not estimate effects for the choose your dictator game.

[23] In these cases, participants are then able to attempt to answer again, after any further instruction or clarification, if needed. Participants could not continue until they answered the comprehension questions correctly.

[24] The number of observations that we exclude differ across games and is reported in fig. C1.

for at least half of all comprehension questions across all games.[25] The estimates are reported in table C7.

To address cases where participants may have had limited focus or attention, we also check the sensitivity of our findings to omitting decisions with the fastest response times, which may be indicative of inattention. We drop the fastest 20% of decisions according to three metrics: the average time spent viewing the profile of the other participant, the average time to make a first selection about an allocation or choice, and the average time to confirm a choice and moving on to the next decision (tables C8–C10). We next turn to the precision of our measure of coethnicity. In some cases, a participant may have parents who belong to two different ethnicities. Motivated by this, we check the robustness of our findings to restricting the sample to participants who share the same mother tongue with both their mother and their father (table C11). We also examine the robustness of our findings to the omission of influential observations (table C12). We omit observations that are found to have the greatest influence on the regression estimate and deemed to be outliers based on their calculated Welsch distance (Belsley, Kuh, and Welsch 1980).

As shown in figures 3–5, the estimated effects are remarkably stable and, in general, they do not alter our conclusions. In addition, we find that the robustness checks do not systematically move our point estimates one direction or another. The baseline estimates tend to fall in the middle of the ordered set of estimates. The estimated hydrocortisone effect continues to appear marginally significant at best (fig. 3). In particular, we find no evidence that omitting observations that might yield noisier estimates, such as those with poor comprehension or fast response times, yields estimates that are more precise or larger in magnitude. In contrast, the robustness checks appear to confirm the strength of the estimated coethnicity effect (fig. 4). In every specification, the estimated coethnicity effect is positive for the choose your dictator game, the dictator game, the first decision of the trust game, and for the social proximity survey. The estimated effect for the second decision in the trust game is always a very precisely estimated zero effect. Thus, the effect of coethnicity appears very robust. Finally, we also find that our conclusion regarding the interaction between hydrocortisone and coethnicity is not altered by our sensitivity checks (fig. 5). The estimated interaction effect is never statistically different from zero.

Overall, our sensitivity checks reinforce the baseline findings. We find very limited evidence of an effect of hydrocortisone, strong evidence for a coethnic preference, and no evidence that coethnic preference is stronger under hydrocortisone.

[25] See fig. C2 for the distribution of this measure.

Heterogeneous effects.—We now turn to an examination of two prespecified heterogeneous effects. We begin by considering the educational attainment of participants. Our study intentionally includes less literate populations. This was accomplished both through extensive recruiting and by providing instructions orally through in-person and prerecorded instructions, rather than written text only. To examine the heterogeneity of our results by education, we run our analysis separately for participants without any secondary education and with at least some secondary education. The results can be found in tables C13 and C14. A coethnicity effect that is similar in magnitude is found in both subsamples. Finally, the hydrocortisone effect and its interaction remain unimportant in both subsamples.

The second form of heterogeneity that we examine is motivated by the possibility that behavior across the multiple rounds of each game may be influenced by whether the first match is with a coethnic or with a non-coethnic. Due to how we randomized, approximately 25% of the sample is paired with a coethnic in the first round and a non-coethnic in the second round, while 75% are paired with a non-coethnic in the first round and then a coethnic in the following round.

Whether a player is first paired with a coethnic may matter for the following reason. When the first decision made in a game is an allocation to a coethnic, then in the second round, making an allocation that favors the coethnic requires allocating *less* to the non-coethnic in the second round. By contrast, when the first decision made is an allocation to a non-coethnic, then in the second round, making an allocation that favors the coethnic player means allocating *more* to the coethnic in the second round. That is, in the first case, implementing a second round allocation that favors the coethnic player means choosing an allocation that is less prosocial relative to the first round. In the second case, implementing a second round allocation that favors the coethnic means choosing an allocation that is more prosocial relative to the first round. It is possible that coethnic preferences emerge when favoring a coethnic is perceived as being more prosocial to coethnics rather than as being less prosocial to non-coethnics.

To check for this, for each game, we split the sample into two groups: those who were matched with a coethnic first and those matched with a non-coethnic first. The estimates for the two samples are reported in tables C15 and C16. In line with the discussion above, the estimated coethnicity effect is consistently twice as large in the DG, TG1, and social proximity survey when the first match is with a non-coethnic and, therefore, favoring the coethnic player means behaving more prosocially and giving more in the second round than in the first.

The sensitivity of the estimates to the order in which players are matched provides further evidence of the importance of the details of the protocols for experiments aimed at measuring coethnic preference. The result is consistent with the fact that how ethnicity is reported is important for detecting coethnic preference.

VI. Conclusion

We studied the relationships between stress, coethnicity, and prosocial behavior in a laboratory setting in Nairobi, Kenya. The starting point of our study was the fact that while observational studies provide evidence that ethnicity is an important determinant of economic, social, and political outcomes in sub-Saharan Africa, many laboratory studies fail to find an effect of coethnicity on behavior. We examined the possibility that this difference is explained by the fact that coethnic preference may only emerge during times of stress, which is when important real-world decisions tend to be made.

Our study tested this explanation in a laboratory setting with 1,784 participants by randomly increasing the level of the stress hormone cortisol using hydrocortisone pills. We randomly manipulated whether the other player belongs to the same or a different ethnic group. Participants were randomly paired with other participants for the different rounds of each behavioral game. While the identity of the other player was unknown, participants were provided with the other player's age group (young, middle-aged, old), gender, and mother tongue, which is a direct indicator of the ethnicity.

We found some limited but nonrobust evidence that hydrocortisone decreases prosocial behavior. We find a sizeable and robust coethnicity effect. Contrary to our hypothesis, we find no evidence that hydrocortisone increases coethnic preference. In fact, the average effect that we estimate across all outcomes suggests that, if anything, hydrocortisone may reduce coethnic preference, although the estimates are imprecise and not robust.

The fact that we find limited evidence of the importance of stress for coethnic preference should not be viewed as the final answer on how stress affects coethnic preference. Since our study used hydrocortisone, it targeted only one of two main stress systems in humans, namely, the hypothalamic-pituitary-adrenal (HPA) axis, which releases cortisol. Therefore, our findings do not speak to potential effects of manipulating the other main stress system, the sympathetic-adrenal-medullary (SAM) axis, which releases adrenaline and noradrenaline. These two systems may interact, and real-world events may trigger both systems simultaneously. In fact, previous work suggests that the two stress systems may interact to generate behavioral effects (Schwabe et al. 2010). We view this as an important avenue for future research.

Data Availability

Code replicating the tables and figures in this article can be found in Haushofer et al. (2022) in the Harvard Dataverse, https://doi.org/10.7910/DVN/TWKAZ4.

References

Alesina, Alberto, Reza Baquir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *Q.J.E.* 114 (4): 1243–84.

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *J. Econ. Growth* 8:155–94.

Alesina, Alberto, Sebastian Hohmann, Stelios Michalopoulos, and Elias Papaioannou. 2021. "Intergenerational Mobility in Africa." *Econometrica* 89 (1): 1–35.

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou. 2016. "Ethnic Inequality." *J.P.E.* 124 (2): 428–88.

Angrist, Joshua D. 2001. "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors." *J. Bus. and Econ. Statis.* 19 (1): 2–28.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics.* Princeton, NJ: Princeton Univ. Press.

Balliet, Daniel, Junhui Wu, and Carsten K. W. De Dreu. 2014. "Ingroup Favoritism in Cooperation: A Meta-Analysis." *Psychological Bull.* 140 (6): 1556–81.

Bardsley, Nicholas. 2008. "Dictator Game Giving: Altruism or Artefact?" *Experimental Econ.* 11 (2): 122–33.

Bauer, Michal, Christopher Blattman, Julie Chytilova, Joseph Henrich, Edward Miguel, and Tamar Mitts. 2016. "Can War Foster Cooperation?" *J. Econ. Perspectives* 30 (3): 249–74.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, et al. 2020. "Ethnically Biased? Experimental Evidence from Kenya." *J. European Econ. Assoc.* 18 (1): 134–64.

Blouin, Arthur, and Sharun W. Mukand. 2019. "Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda." *J.P.E.* 127 (3): 1008–62.

Blum, Ashley, Chad Hazlett, and Daniel N. Posner. 2021. "Measuring Ethnic Bias: Can Misattribution-Based Tools from Social Psychology Reveal Group Biases That Economics Games Cannot?" *Polit. Analysis* 29 (3): 385–404.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2015. "The Value of Democracy: Evidence from Road Building in Kenya." *A.E.R.* 105 (6): 1817–51.

Chemin, Matthieu, Joost de Laat, and Johannes Haushofer. 2013. "Negative Rainfall Shocks Increase Levels of the Stress Hormone Cortisol among Poor Farmers in Kenya." Working paper, Dept. Econ., McGill Univ.

Chen, Yan, and Sherry Xin Li. 2009. "Group Identity and Social Preferences." *A.E.R.* 99 (1): 431–57.

Coates, J. M., and J. Herbert. 2008. "Endogenous Steroids and Financial Risk Taking on a London Trading Floor." *Proc. Nat. Acad. Sci. USA* 105 (16): 6167–72.

Cornelisse, Sandra, Vanessa van Ast, Johannes Haushofer, Maayke Seinstra, and Marian Joels. 2013. "Time-Dependent Effect of Hydrocortisone Administration on Intertemporal Choice." https://ssrn.com/abstract=2294189.

de Kloet, E. Ron. 2004. "Hormones and the Stressed Brain." *Ann. New York Acad. Sci.* 1018 (1): 1–15.

de Kloet, E. Ron, Marian Joels, and Florian Holsboer. 2005. “Stress and the Brain: From Adaptation to Disease.” *Nature Rev. Neuroscience* 6:463–75.

Delaney, Liam, Günther Fink, and Colm P. Harmon. 2014. “Effects of Stress on Economic Decision-Making: Evidence from Laboratory Experiments.” IZA Discussion Paper no. 8060, Inst. Study Labor, Bonn.

Depetris-Chauvin, Emilio. 2015. “State History and Contemporary Conflict: Evidence from Sub-Saharan Africa.” https://ssrn.com/abstract=2679594.

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. “Measuring and Bounding Experimenter Demand.” *A.E.R.* 108 (11): 3266–302.

Easterly, William, and Ross Levine. 1997. “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *Q.J.E.* 112:1203–50.

Ferree, Karen E., Clark C. Gibson, and James D. Long. 2014. “Voting Behavior and Electoral Irregularities in Kenya’s 2013 Election.” *J. Eastern African Studies* 8 (1): 153–72.

Fershtman, Chaim, and Uri Gneezy. 2001. “Discrimination in a Segmented Society: An Experimental Approach.” *Q.J.E.* 116:351–77.

Franck, Raphael, and Ilia Rainer. 2012. “Does the Leader’s Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa.” *American Polit. Sci. Rev.* 106 (2): 294–325.

Gomila, Robin. 2021. “Logistic or Linear? Estimating Causal Effects of Experimental Treatment on Binary Outcomes Using Regression Analysis.” *J. Experimental Psychology: General* 150 (4): 700–709.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein. 2007. “Why Does Ethnic Diversity Undermine Public Goods Provision?” *American Polit. Sci. Rev.* 101 (4): 709–25.

———. 2009. *Coethnicity: Diversity and the Dilemmas of Collective Action.* New York: Russell Sage Found.

Haushofer, Johannes, Matthieu Chemin, Chaning Jang, and Justin Abraham. 2020. “Economic and Psychological Effects of Health Insurance and Cash Transfers: Evidence from a Randomized Experiment in Kenya.” *J. Development Econ.* 144:102416.

Haushofer, Johannes, Marie Collins, Giovanna De Giusti, et al. 2014. “A Methodology for Laboratory Experiments in Developing Countries: Examples from the Busara Center.” Working paper, Busara Center, Nairobi, Kenya.

Haushofer, Johannes, Sandra Cornelisse, Maayke Seinstra, Ernst Fehr, Marian Joëls, and Tobias Kalenscher. 2013. “No Effects of Psychosocial Stress on Intertemporal Choice.” *PLoS ONE* 8:e78597.

Haushofer, Johannes, Chaning Jang, and John Lynham. 2015. “Stress and Temporal Discounting: Do Domains Matter?” Working paper, Dept. Econ., Princeton Univ.

Haushofer, Johannes, Sara Lowes, Abednego Musau, et al. 2022. “Replication Data for ‘Stress, Ethnicity, and Prosocial Behavior.’” Harvard Dataverse. https://doi.org/10.7910/DVN/TWKAZ4.

Haushofer, Johannes, Sara Lowes, Nathan Nunn, and Nancy Qian. 2019. “Stress and In-Group Preferences: Pre-Analysis Plan.” https://www.socialscienceregistry.org/trials/3937.

Haushofer, Johannes, and Jeremy Shapiro. 2016. “The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya.” *Q.J.E.* 131 (4): 1973–2042.

Henckens, Marloes J. A. G., Guido A. van Wingen, Marian Joëls, and Guillén Fernández. 2011. “Time-Dependent Corticosteroid Modulation of Prefrontal Working Memory Processing.” *Proc. Nat. Acad. Sci. USA* 108 (14): 5801–06.

———. 2012. “Time-Dependent Effects of Cortisol on Selective Attention and Emotional Interference: A Functional MRI Study.” *Frontiers Integrative Neuroscience* 6:66.

Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. “The Weirdest People in the World.” *Behavioral and Brain Sci.* 33 (2/3): 1–75.

Hewstone, Miles, Mark Rubin, and Hazel Willis. 2002. “Intergroup Bias.” *Ann. Rev. Psychology* 53:575–604.

Hines, Edgar A., and George E. Brown. 1936. “The Cold Pressor Test for Measuring the Reactibility of the Blood Pressure: Data Concerning 571 Normal and Hypertensive Subjects.” *American Heart J.* 11 (1): 1–9.

Hjort, Jonas. 2014. “Ethnic Divisions and Production in Firms.” *Q.J.E.* 129 (4): 1899–946.

Jakiela, Pamela, and Owen Ozier. 2019. “The Impact of Violence on Individual Risk Preferences: Evidence from a Natural Experiment.” *Rev. Econ. and Statis.* 101 (3): 547–59.

Kandasamy, Narayanan, Ben Hardy, Lionel Page, et al. 2014. “Cortisol Shifts Financial Risk Preferences.” *Proc. Nat. Acad. Sci. USA* 111 (9): 3608–13.

Kirschbaum, Clemens, and Dirk Hellhammer. 1989. “Salivary Cortisol in Psychobiological Research: An Overview.” *Neuropsychobiology* 22 (3): 150–69.

Kirschbaum, Clemens, Karl-Martin Pirke, and Dirk H. Hellhammer. 1993. “The ‘Trier Social Stress Test’—A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting.” *Neuropsychobiology* 28 (1–2): 76–81.

Levitt, Steven D., and John A. List. 2007. “What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?” *J. Econ. Perspectives* 21 (2): 153–74.

List, John A. 2007. “On the Interpretation of Giving in Dictator Games.” *J.P.E.* 115 (3): 482–93.

———. 2020. “Non Est Disputandum de Generalizability? A Glimpse into the External Validity Trial.” Working Paper no. 27535 (July), NBER, Cambridge, MA.

List, John A., Azeem M. Shaikh, and Yang Xu. 2019. “Multiple Hypothesis Testing in Experimental Economics.” *Experimental Econ.* 22 (4): 773–93.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan Weigel. 2015. “Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test.” *A.E.R.* 105 (5): 340–45.

Lupien, S. J., F. Maheu, M. Tu, A. Fiocco, and T. E. Schramek. 2007. “The Effects of Stress and Stress Hormones on Human Cognition: Implications for the Field of Brain and Cognition.” *Brain and Cognition* 65:209–37.

Macchiavello, Rocco, and Ameet Morjaria. 2015. “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports.” *A.E.R.* 105 (9): 2911–45.

Margittai, Zsofia, Gideon Nave, Marijn van Wingerden, Alfons Schnitzler, Lars Schwabe, and Tobias Kalenscher. 2018. “Combined Effects of Glucocorticoid and Noradrenergic Activity on Loss Aversion.” *Neuropsychopharmacology* 43 (2): 334–41.

Margittai, Zsofia, Tina Strombach, Marijn van Wingerden, Marian Joels, L. Schwabe, and Tobias Kalenscher. 2015. “A Friend in Need: Time-Dependent Effects of Stress on Social Discounting in Men.” *Hormones and Behavior* 73:75–82.

Marx, Benjamin, Thomas M. Stoker, and Tavneet Suri. 2019. “There Is No Free House: Ethnic Patronage in a Kenyan Slum.” *American Econ. J. Appl. Econ.* 11 (4): 36–70.

Michalopoulos, Stelios. 2012. “The Origins of Ethnolinguistic Diversity.” *A.E.R.* 102 (4): 1508–39.

Minkley, Nina, Thomas P. Schroder, Oliver T. Wolf, and Wolfgang H. Kirchner. 2014. "The Socially Evaluated Cold-Pressor Test (SECPT) for Groups: Effects of Repeated Administration of a Combined Physiological and Psychological Stressor." *Psychoneuroendocrinology* 45:119–27.

Ndegwa, Stephen N. 1997. "Citizenship and Ethnicity: An Examination of Two Transition Moments in Kenyan Politics." *American Polit. Sci. Rev.* 91 (3): 599–616.

Porcelli, Anthony J., and Mauricio R. Delgado. 2009. "Acute Stress Modulates Risk Taking in Financial Decision Making." *Psychological Sci.* 20 (3): 278–83.

Rasul, Imran, and Daniel Rogger. 2015. "The Impact of Ethnic Diversity in Bureaucracies: Evidence from the Nigerian Civil Service." *A.E.R.* 105 (5): 457–61.

Riis-Vestergaard, Michala Iben, Vanessa A. van Ast, Sandra Cornelisse, Marian Joëls, and Johannes Haushofer. 2018. "The Effect of Hydrocortisone Administration on Intertemporal Choice." *Psychoneuroendocrinology* 88:173–82.

Schwabe, Lars, Martin Tegenthoff, Oliver Höffken, and Oliver T. Wolf. 2010. "Concurrent Glucocorticoid and Noradrenergic Activity Shifts Instrumental Behavior from Goal-Directed to Habitual Control." *J. Neuroscience* 30 (24): 8190–96.

Schweda, Adam, Nadira Sophie Faber, Molly J. Crockett, and Tobias Kalenscher. 2019. "The Effects of Psychosocial Stress on Intergroup Resource Allocation." *Sci. Reports* 9:18620.

Singer, Nina, Monika Sommer, Katrin Döhnel, Sandra Zänkert, Stefan Wüst, and Brigitte M. Kudielka. 2017. "Acute Psychosocial Stress and Everyday Moral Decision-Making in Young Healthy Men: The Impact of Cortisol." *Hormones and Behavior* 93:72–81.

Vinkers, Christiaan H., Jelle V. Zorn, Sandra Cornelisse, et al. 2013. "Time-Dependent Changes in Altruistic Punishment Following Stress." *Psychoneuroendocrinology* 38 (9): 1467–75.

von Dawans, Bernadette, Urs Fischbacher, Clemens Kirschbaum, Ernst Fehr, and Markus Heinrichs. 2012. "The Social Dimension of Stress Reactivity: Acute Stress Increases Prosocial Behavior in Humans." *Psychological Sci.* 23 (6): 651–60.

Waki, Philip N. 2008. *Report of the Commission of Inquiry into Post Election Violence (CIPEV).* Nairobi, Kenya: Government Printer.

Yanagizawa-Drott, David. 2014. "Propaganda and Conflict: Evidence from the Rwandan Genocide." *Q.J.E.* 129 (4): 1947–94.