STRESS, ETHNICITY, AND PROSOCIAL BEHAVIOR

Johannes Haushofer
Sara Lowes
Abednego Musau
David M. Ndetei
Nathan Nunn
Moritz Poll
Nancy Qian

Stress, Ethnicity, and Prosocial Behavior
Johannes Haushofer, Sara Lowes, Abednego Musau, David M. Ndetei, Nathan Nunn, Moritz
Poll, and Nancy Qian
NBER Working Paper No. 30363
August 2022
JEL No. O12,Z10

ABSTRACT

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

Johannes Haushofer
Department of Economics
Stockholm University
Universitetsvägen 10 A
Stockholm 106 91 Sweden
and Busara Center for Behavioral Economics,
Nairobi, Kenya
and also NBER
johannes.haushofer@ne.su.se

Sara Lowes
Department of Economics
University of California at San Diego
9500 Gilman Drive
La Jolla, CA 92093
and NBER
slowes@ucsd.edu

Abednego Musau
University Medical Centre Groningen
a.m.musau@umcg.nl

David M. Ndetei
University of Nairobi
and Africa Mental Health Research and
Training Foundation
P.O. Box 48423
Nairobi, 00100, Kenya
dmndetei@amhf.or.ke

Nathan Nunn
Vancouver School of Economics
University of British Columbia
6000 Iona Drive
Vancouver, BC V6T 2E8
Canada
and NBER
nathan.nunn@ubc.ca

Moritz Poll
Brown University
Department of Economics
64 Waterman Street
Providence, RI 02912
moritz.poll@brown.edu

Nancy Qian
Kellogg School of Management
Northwestern University
2001 Sheridan Road
Evanston, Il 60208
and NBER
nancy.qian@kellogg.northwestern.edu

1. 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, Humphreys, Posner and Weinstein, 2007, 2009, Berge, Bjorvatn, Galle, Miguel, Posner, Tungodden and Zhang, 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 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, Joëls and Holsboer, 2005, de Kloet, 2004). 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, but still be 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, Tegenthoff, Höffken and Wolf, 2010, Henckens, van Wingen, Joëls and Fernández, 2011, Kandasamy, Hardy, Page, Schaffner, Graggaber, Powlson, Fletcher, Gurnell and Coates, 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. 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 – i.e. 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, 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 pro-socially. 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 nor 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] Lastly, we find that participants report feeling closer to and more positively towards 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 – i.e., the differential effect of coethnicity under the hydrocortisone treatment – is small in magnitude, imprecisely estimated, 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, an important concern is experimenter demand effects. This is even more true in our study, which communicates the other player's ethnicity directly with information

---

[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.

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, those paired with a non-coethnic. 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 et al., 2005, de Kloet, 2004). There is substantial evidence that stress increases cortisol levels (Kirschbaum, Pirke and Hellhammer, 1993, Hines and Brown, 1936, Minkley, Schröder, Wolf and Kirchner, 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, Chemin, Jang and Abraham, 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 (Riis-Vestergaard, van Ast, Cornelisse, Joëls and Haushofer, 2018, Kandasamy et al., 2014); other paradigms for studying stress have also affected these same behaviors (Delaney, Fink and Harmon, 2014, Porcelli and Delgado, 2009).

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, Fischbacher, Kirschbaum, Fehr and Heinrichs (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, Strombach, van Wingerden, Joëls, Schwabe and Kalenscher (2015) find an increase in altruism in a dictator game for in-group but not out-group members after the TSST. In contrast, Vinkers, Zorn, Cornelisse, Koot, Houtepen, Olivier, Verster, Kahn, Boks, Kalenscher and Joëls (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, Devleeschauwer, Easterly, Kurlat and Wacziarg, 2003, Franck and Rainer, 2012, Michalopoulos, 2012, Hjort, 2014, Yanagizawa-Drott, 2014, Burgess, Jedwab, Miguel, Morjaria and Padró i Miquel, 2015, Rasul and Rogger, 2015, Depetris-Chauvin, 2016, Alesina et al., 2016, Marx et al., 2019, Alesina, Hohmann and Papaioannou, 2021) relative to those that use experimental methods (Habyarimana et al., 2007, 2009, Berge et al., 2020, Blum et al., 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-Sahara Africa. This possibility is consistent with the fact that, within the literature, coethnic preference appear to be less commonly found in behavioral games, where ethnicity is often indirectly reported, than in Implicit Association Tests (IAT) where ethnicity, by the nature of the IAT, is directly reported (e.g. Lowes, Nunn, Robinson and Weigel, 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 et al. (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, Blattman, Chytilova, Henrich, Miguel and Mitts (2016) finds that war exposure may increase prosocial behavior towards in-group members but not towards out-group members. Hjort (2014) finds evidence that Kenyan flower workers discriminate in favor of a coethnic coworker relative to a non-coethnic 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 governments efforts at post-genocide 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, their findings are likely due to other effects of conflict.

This paper proceeds as follows. Section 2 describes the experimental design. Section 3 discusses the estimating equations. Sections 4 and 5 present our results and pre-specified robustness checks. Section 6 concludes.

2. Design

The full study design and analysis were pre-specified in Haushofer, Lowes, Nunn and Qian (2019) and all pre-specified analysis is reported in the main body of the paper or in the appendix. Analysis that was not pre-specified will be noted in the paper. We also point out a number of minor design changes that occurred after the submission of the pre-analysis plan in Appendix Table A1.

2.1. 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, Kenya. 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 even ethnicity (Haushofer, Collins, De Giusti, Njoroge, Odero, Onyango, Vancel, Jang, Kuruvilla and Hughes, 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 breastfeeding. 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 Appendix 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. Beyond this, there was also a health screening. On the day of the experimental session, upon arrival at Busara, participants received a thorough health exam. Anyone who did not satisfy the health requirements did not participate in the session, although they still received the show-up and transport fee.

A summary of the different steps of the process as well as the nature of attrition is provided in Appendix 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 final sample in our paper.

We find that our final sample is slightly more likely to be male than the Kenyan and Nairobi population respectively 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 over-represents Luo, Luhya, and Kamba. The over-representation 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.

2.2. 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, Maheu, Tu, Fiocco and Schramek, 2007).

---

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

Studying stress using the oral administration of hydrocortisone to increase cortisol levels is common in the neurobiological stress literature (Henckens et al., 2011, Kandasamy et al., 2014, Schwabe et al., 2010). 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 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 et al., 2015). In contrast, hydrocortisone has similar effects on cortisol levels in our Kenyan sample as 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

---

[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.

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

2.3. 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 (Berge et al., 2020, Blum et al., 2021, Lowes et al., 2015). However, in contrast to some previous studies Berge et al. (2020), but consistent with others (Lowes et al., 2015, Blum et al., 2021), we report the ethnic identity of the other player directly.

2.4. 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 felt to this person, how likely they were to be their friend, and how much they trusted 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 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 these 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 non-coethnic 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

---

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

[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."

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; e.g. 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 mentioned above.

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 et al., 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.

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, e.g. 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 make 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.

---

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

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 three, 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 take values that range from zero to one. 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.

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:

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?" 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

---

[7] See Appendix Figure A4.

social proximity survey, where we ask participants directly about the perceived closeness of another person to them, to be most susceptible to experimenter demand. A comparison of the estimated coethnicity effect across the range of activities therefore is informative about the sensitivity of our findings to experimenter demand effects.

2.5. 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 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 the assigned cubicle.

The games and surveys are administered using touch screen computers to enable computer-illiterate individuals to participate. See Appendix 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), pre-recorded 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

14

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 Appendix 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 Appendix Figure A3. A detailed general overview of the sessions is provided in Appendix A.4.

2.6. 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 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 co-authors 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

(KEMRI), 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.

3. 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 pre-analysis plan.[8]

Our main estimating equation is as follows:

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

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 one if player i is in the hydrocortisone treatment group; I_{ij}^{Coethnic} is an indicator variable that equals one 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 $\alpha_{e(j)}$.

---

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

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 one 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:

$$\text{Hydrocortisone Effect} \quad = \quad \hat{\beta}_2 + \hat{\beta}_3 \bar{I}_{ij}^{\text{Coethnic}} + \hat{\beta}_5 \bar{I}_{ij}^{\text{Same Gender}} + \hat{\beta}_7 \bar{I}_{ij}^{\text{Same Age}}, \tag{2}$$

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:

$$\text{Coethnicity Effect} \quad = \quad \hat{\beta}_1 + \hat{\beta}_2 \bar{I}_{ij}^{\text{HC}}, \tag{3}$$

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.

3.1. 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:

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

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, 2020). Results from using a conditional logit model, which we report in Appendix 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, Ch. 3).

4. Results

4.1. Sample Balance

Balance checks for the treatment and control groups are reported in Appendix 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.

4.2. 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 Appendix 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 (in solid red) and the placebo group (in dashed blue). 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 (Riis-Vestergaard et al., 2018, Margittai, Nave, van Wingerden, Schnitzler, Schwabe and Kalenscher, 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, Cornelisse, Seinstra, Fehr, Joëls and Kalenscher, 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 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.

---

[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 one 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 sensitivity of 0.16 ng/ml (IBL, 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, Faber, Crockett and Kalenscher (2019), it led to a 5 nmol/l increase in cortisol; in Kirschbaum et al. (1993), to an increase from 10 to 16 nmol/l; in Singer, Sommer, Döhnel, Zänkert, Wüst and Kudielka (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, van Wingen, Joëls and Fernández (2012) administer 10mg of hydrocortisone, with an increase in cortisol from 10 to 45 nmol/l. Cornelisse, van Ast, Haushofer, Seinstra and Joels (2014) administer 10mg of hydrocortisone with an increase of 2–4 nmol/l. Riis-Vestergaard et al. (2018) administer 10mg and see an increase from 11 to 81.5 nmol/l. Kandasamy et al. (2014) administer 5mg 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.

Figure 1: Cortisol levels over time by treatment
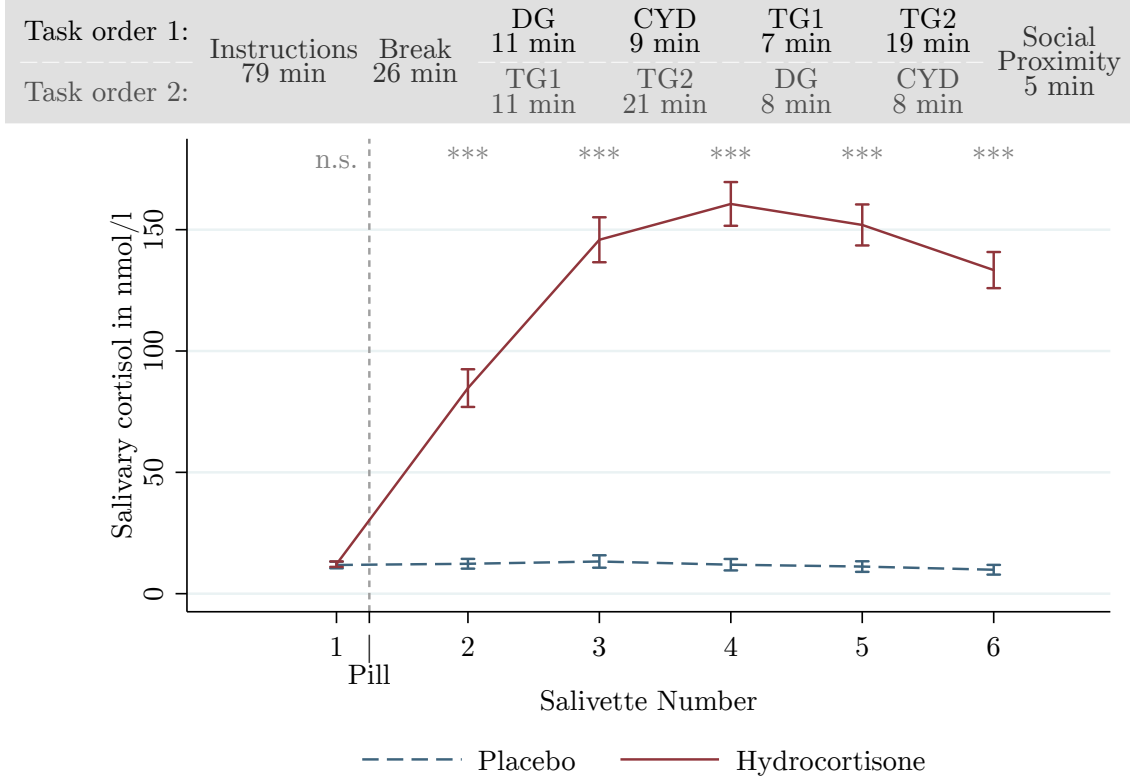
Notes: The figure depicts salivary cortisol levels during the study. The top bar describes the study activities. When applicable, we distinguish between two task orders. The study consisted of oral instructions leading up to the first (pre-pill) salivette, followed by the pill and a break, before the 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 (20mg 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, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2: Comparison of means across treatment categories
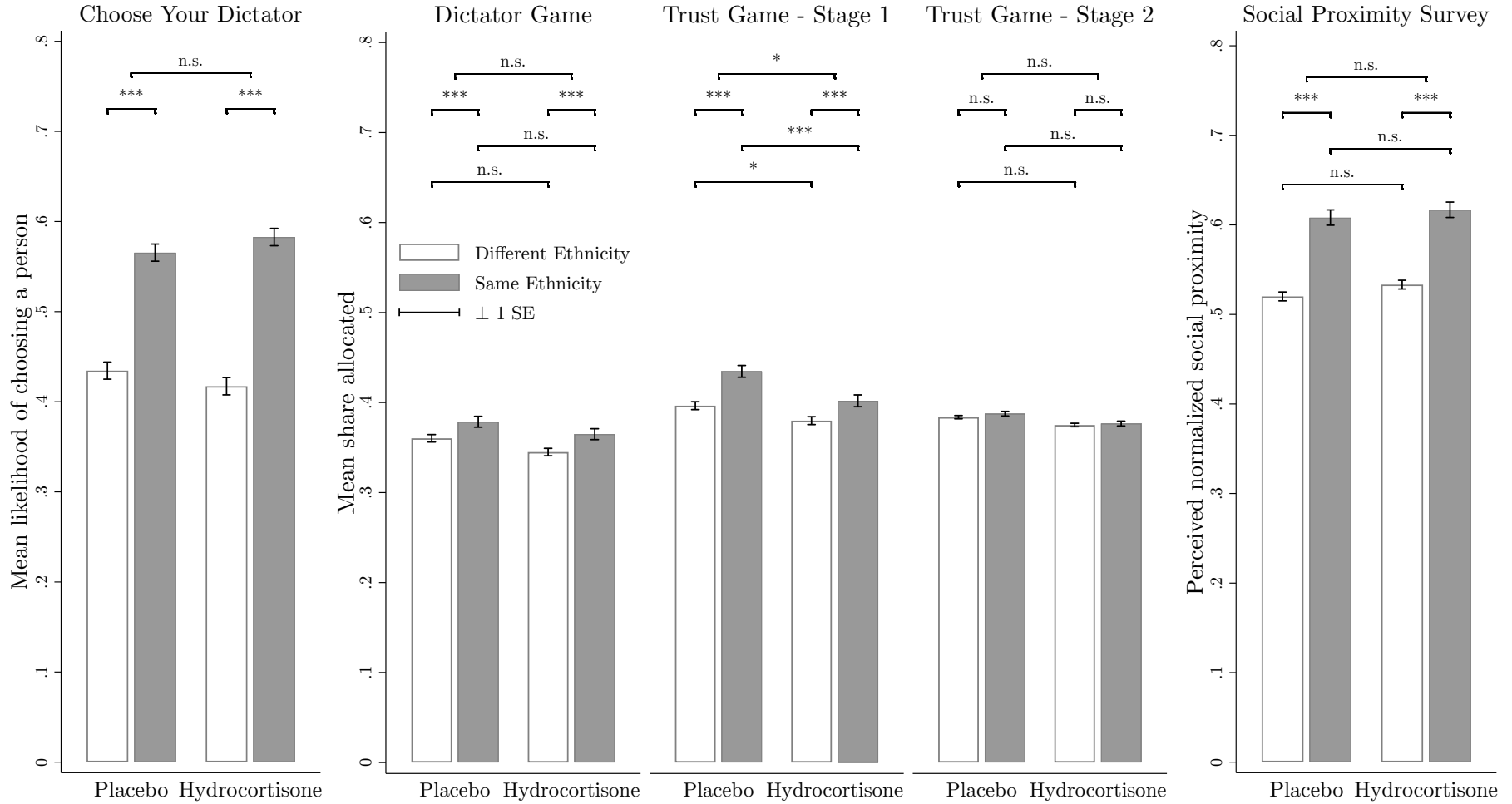
Notes: The figure compares game outcomes across the four possible treatment combinations. For each of the games, the left two bars show results for participants under placebo and the right bars under hydro. White bars symbolize non-coethnic pairings and grey bars co-ethnic pairings. The vertical brackets 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, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.3. 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 for non-coethnic pairs. The means are reported separately for the participants in the hydro 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 one 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 (3rd row) and non-coethnic pairings (4th 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

---

[11] This comparison is not possible or 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.

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.

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.

4.4. 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.

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 (p.p.) in the DG and 2.3 p.p. in TG1. The estimates are statistically significant at the 10% and 5% levels. 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 Appendix Table B3 and Appendix Figure B1, the estimates are similar if we examine each of the three questions separately.

---

[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. Appendix Table C1 reports the estimates when we re-assign 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

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Constant	0.6231***	0.3445***	0.3289***	0.3973***	0.5406***
	(0.0061)	(0.0230)	(0.0228)	(0.0197)	(0.0237)
Hydrocortisone		−0.0062	−0.0321***	−0.0139	0.0206
		(0.0114)	(0.0117)	(0.0110)	(0.0132)
Same ethnicity	0.0611***	0.0181***	0.0385***	0.0025	0.0841***
	(0.0097)	(0.0055)	(0.0060)	(0.0029)	(0.0072)
Hydrocortisone × Same ethnicity	0.0228*	0.0014	−0.0178**	−0.0019	−0.0051
	(0.0135)	(0.0074)	(0.0084)	(0.0042)	(0.0101)
Same gender	−0.0104	0.0041	−0.0135*	0.0053	0.0376***
	(0.0094)	(0.0069)	(0.0073)	(0.0058)	(0.0087)
Hydrocortisone × Same gender	−0.0327**	−0.0125	0.0154	−0.0033	−0.0201*
	(0.0132)	(0.0096)	(0.0101)	(0.0085)	(0.0122)
Same age group	0.0444**	0.0111	0.0325*	−0.0236*	−0.0106
	(0.0219)	(0.0175)	(0.0166)	(0.0142)	(0.0181)
Hydrocortisone × Same age group	−0.0400***	−0.0110	0.0236**	0.0177*	0.0096
	(0.0145)	(0.0107)	(0.0113)	(0.0092)	(0.0129)
Matched player is Luo	−0.0091	−0.0042	0.0082	−0.0008	0.0074
	(0.0081)	(0.0051)	(0.0059)	(0.0037)	(0.0055)
Matched player is Kikuyu	−0.0318***	−0.0015	0.0054	−0.0051	−0.0019
	(0.0082)	(0.0052)	(0.0057)	(0.0037)	(0.0053)
Matched player is Luhya	−0.0220***	−0.0012	0.0066	0.0025	0.0138**
	(0.0083)	(0.0052)	(0.0057)	(0.0037)	(0.0058)
Matched player is female	0.0054	0.0054	0.0119**	0.0005	0.0124**
	(0.0066)	(0.0048)	(0.0051)	(0.0042)	(0.0061)
Matched player is middle-aged	0.0318	0.0198	0.0550***	−0.0015	−0.0366**
	(0.0214)	(0.0164)	(0.0157)	(0.0141)	(0.0169)
Matched player is old	0.0364	0.0307*	0.0793***	−0.0066	−0.0440**
	(0.0225)	(0.0167)	(0.0162)	(0.0145)	(0.0178)
Decision maker is Luo		0.0116	0.0047	−0.0017	0.0059
		(0.0163)	(0.0155)	(0.0157)	(0.0164)
Decision maker is Kikuyu		0.0167	0.0273	0.0120	0.0025
		(0.0173)	(0.0168)	(0.0170)	(0.0176)
Decision maker is Luhya		−0.0199	0.0007	−0.0242*	−0.0173
		(0.0150)	(0.0145)	(0.0144)	(0.0151)
Decision maker is female		−0.0130	0.0026	0.0058	−0.0285***
		(0.0094)	(0.0093)	(0.0094)	(0.0100)
Decision maker is middle-aged		0.0501**	0.0141	0.0490**	−0.0095
		(0.0242)	(0.0195)	(0.0221)	(0.0261)
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

Notes: 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Effects of hydrocortisone, coethnity and their interaction

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5740	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.4945)	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		−0.0157*	−0.0225**	−0.0103	0.0123
		(0.0093)	(0.0093)	(0.0095)	(0.0099)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0724***	0.0188***	0.0297***	0.0015	0.0817***
	(0.0069)	(0.0038)	(0.0043)	(0.0022)	(0.0052)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0228*	0.0014	−0.0178**	−0.0019	−0.0051
	(0.0135)	(0.0074)	(0.0084)	(0.0042)	(0.0101)
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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. The estimates for the CYD game, reported in column (1), imply that participants are 7.2 p.p. 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 p.p. 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 p.p. 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 column (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 towards 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,

---

[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 Appendix Table B2.

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).

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 column (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.

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 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

---

[15] See Appendix Table B2.

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 (column (1)), while in the TG2, the interaction effect is negative and statistically significant at the 5% level (column (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.

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 towards 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 the social proximity survey. For each game, the outcomes are measures that range from zero to one. The average effects are estimated using 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 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 social proximity survey). The remaining columns include the social proximity questions, either as a single average index (columns (3) and (4)) or including each question individually (columns (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.

---

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

Table 3: Average effects across all outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Sample mean	0.3799	0.3793	0.3946	0.4215	0.4173	0.4215
Sample standard deviation	(0.2509)	(0.2552)	(0.2559)	(0.2660)	(0.2751)	(0.2806)
Panel A: Average hydrocortisone effect[a]						
Hydrocortisone effect	-0.0126	-0.0158**	-0.0104	-0.0088	-0.0070	-0.0088
	(0.0084)	(0.0079)	(0.0078)	(0.0067)	(0.0071)	(0.0067)
Panel B: Average Coethnicity effect[b]						
Coethnicity effect	0.0081***	0.0167***	0.0135***	0.0309***	0.0223***	0.0309***
	(0.0018)	(0.0022)	(0.0018)	(0.0021)	(0.0018)	(0.0021)
Panel C: Interaction of hydrocortisone and coethnicity[c]						
Interaction effect	-0.0035	-0.0057	-0.0042	-0.0071*	-0.0052	-0.0071*
	(0.0035)	(0.0042)	(0.0034)	(0.0042)	(0.0036)	(0.0042)
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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 game (columns (1) and (2)), as well as the social proximity survey index (columns (3) and (4)) or the social proximity survey questions separately (columns (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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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 baselines estimates from Table 2 with the correction and report these in Appendix Table B4. Our conclusions remain unchanged. We continue to find a non-robust and weak hydrocortisone effect, a strong and robust coethnicity effect, and no evidence for a robust and consistent interaction effect.

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 as the main specification in Panel B of Table 2. Panel B of Table 4 shows that the interaction effect is similarly inconclusive as in Panel C of Table 2.[18]

4.5. 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.

---

[17] The full estimates are reported in Appendix Table B5.

[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.

Table 4: Effects of hydrocortisone, coethnity and their interaction: with player 1 fixed effects

	(1) Dictator Game	(2) Trust Game 1	(3) Trust Game 2	(4) Social Proximity
Sample mean	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average coethnicity effect[a]				
Coethnicity effect	0.0184***	0.0301***	0.0028	0.0819***
	(0.0038)	(0.0043)	(0.0022)	(0.0052)
Panel B: Interaction of hydrocortisone and coethnicity[b]				
Interaction effect	0.0015	−0.0178**	−0.0021	−0.0052
	(0.0074)	(0.0084)	(0.0041)	(0.0101)
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

Notes: The table reports estimates of a version of equation (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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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) 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

---

[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.

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 et al. (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 et al. (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 (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 et al. (2018), testing how the allocated amount varies between the two groups allows one to gauge the size of experimenter demand effects in a study, i.e. the extent to which participants change their behavior due to a belief on 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

---

[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.

Table 5: Experimenter demand effects

| | Share Allocated in Experimenter Demand Module | | | | |
| | (1) | (2) | (3) | (4) | (5) |
	Full sample	Hydrocortisone pill	Placebo pill	Coethnic pairing	Non-coethnic pairing
Sample mean	0.4256	0.4283	0.4229	0.4213	0.4350
Sample standard deviation	(0.2754)	(0.2736)	(0.2772)	(0.2736)	(0.2789)
Demand Treatment	−0.0047	−0.0091	−0.0005	0.0014	−0.0106
	(0.0114)	(0.0159)	(0.0164)	(0.0161)	(0.0161)
Participants/Decisions	1,784	906	878	858	926

Notes: 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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 p.p.), 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, e.g. participants who receive the hydrocortisone treatment or who were paired with a coethnic partner, we re-estimate 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) 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.

5. Robustness and Sensitivity Checks

We report estimates from a series of pre-specified 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 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 pre-specified, 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 one if player i and j both belong to an ethnic group that is part of the same political coalition. The estimated

Figure 3: Robustness of the hydrocortisone effect
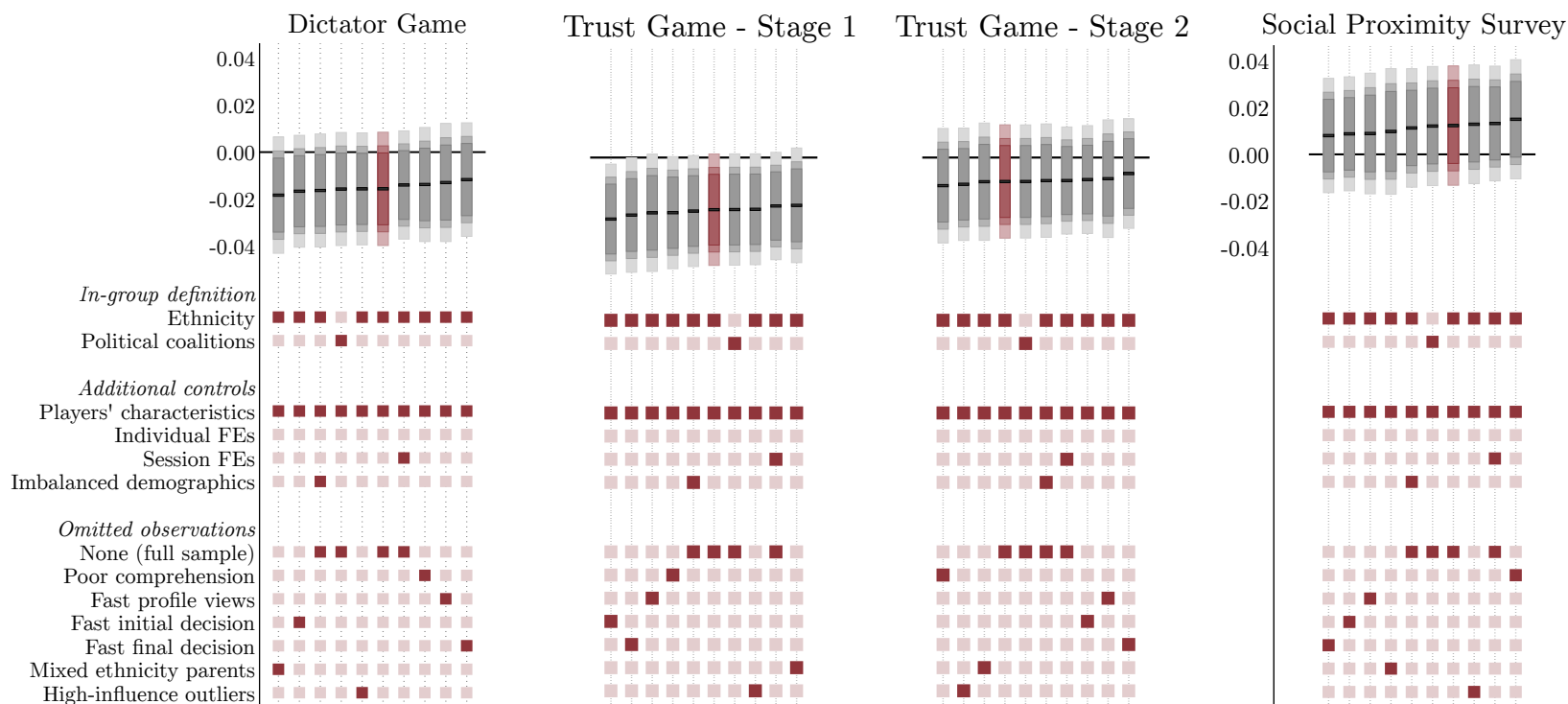
Notes: The figure shows the coefficients and standard errors for the effect of hydrocortisone on behavior in each game for alternative specifications. Coefficients are depicted by black horizontal lines. The vertical bars, from darkest to lightest, denote the 90%, 95%, and 99% confidence intervals, respectively. 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.

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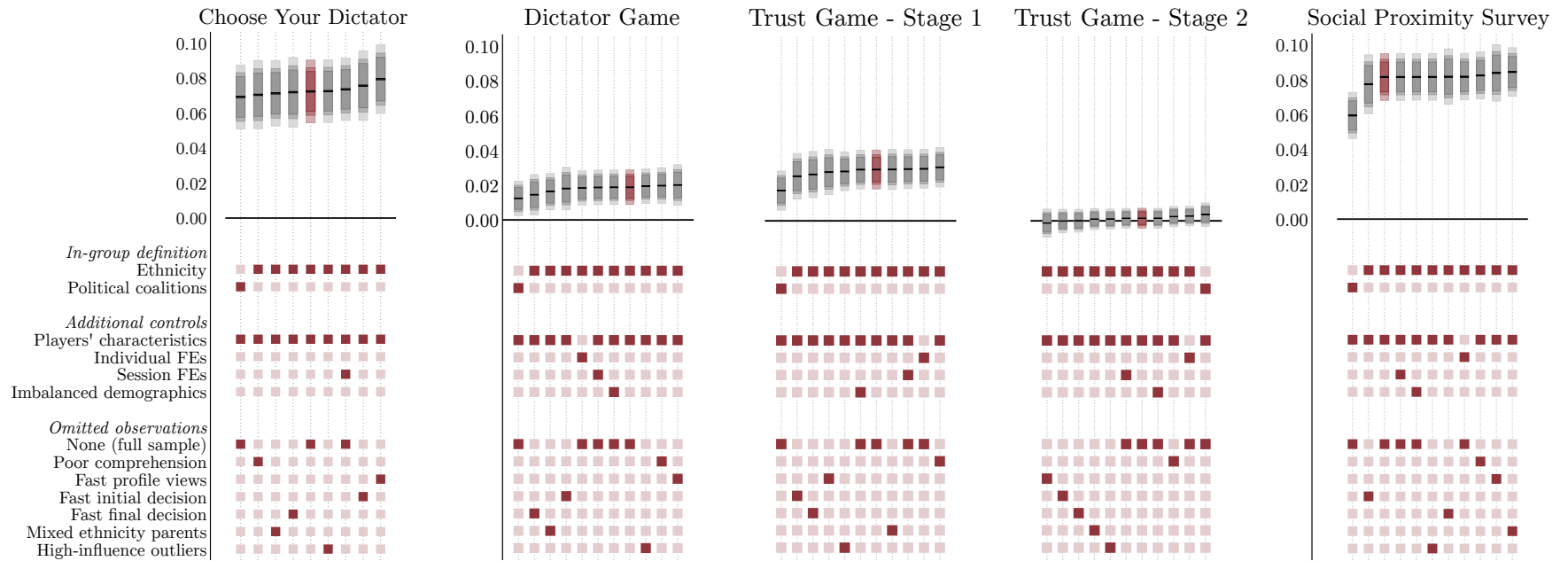
Notes: The figure shows the coefficients and standard errors for the effect of coethnicity on behavior in each game for alternative specifications. Coefficients are depicted by black horizontal lines. The vertical bars, from darkest to lightest, denote the 90%, 95%, and 99% confidence intervals, respectively.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

Figure 5: Robustness of the interaction of coethnicity and hydrocortisone

	Choose Your Dictator	Dictator Game	Trust Game - Stage 1	Trust Game - Stage 2	Social Proximity Survey

In-group definition
Ethnicity
Political coalitions

Additional controls
Players' characteristics
Individual FEs
Session FEs
Imbalanced demographics

Omitted observations
None (full sample)
Poor comprehension
Fast profile views
Fast initial decision
Fast final decision
Mixed ethnicity parents
High-influence outliers
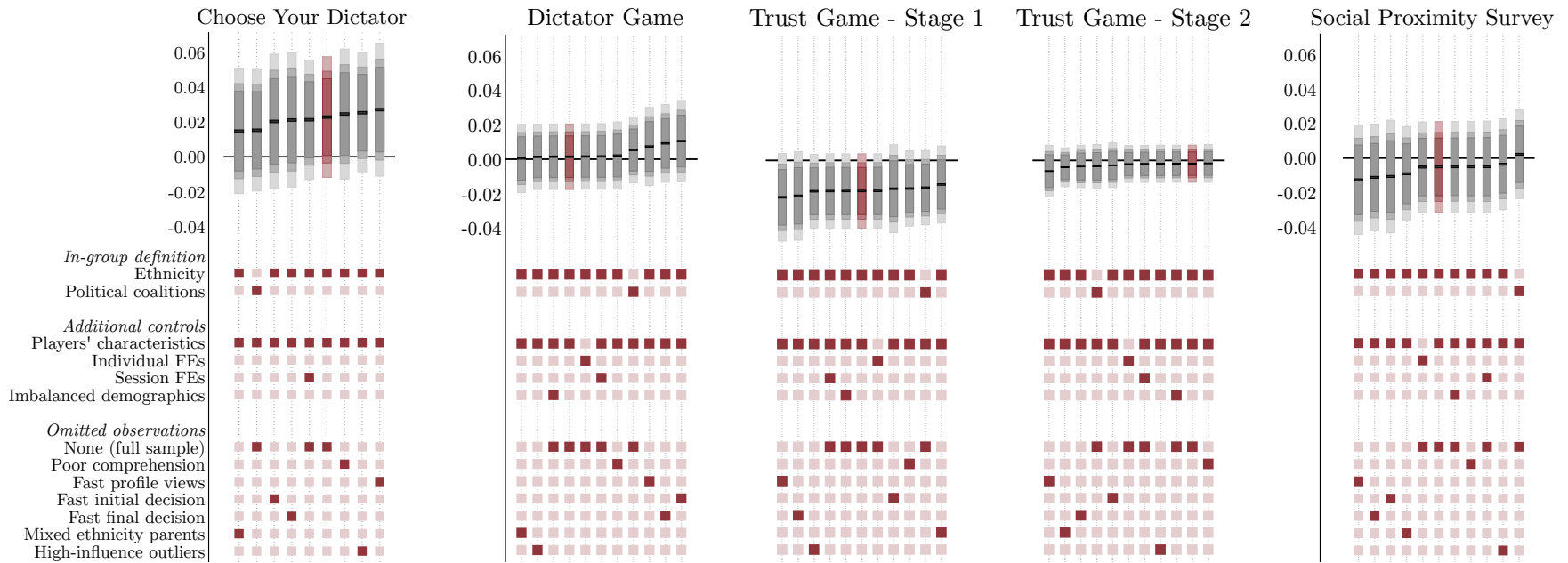
Notes: The figure shows the coefficients and standard errors for the interaction effect of coethnicity and hydrocortisone on behavior in each game for alternative specifications. Coefficients are depicted by black horizontal lines. The vertical bars, from darkest to lightest, denote the 90%, 95%, and 99% confidence intervals, respectively. The red bar indicates our main specification. The bottom panel indicates the combination of robustness checks associated with each specification.

effects are reported in Appendix 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, etc (Appendix 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 (Appendix Table C5). This check was not pre-specified.[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 (Appendix 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 for at least half of all comprehension questions across all games.[25]. The estimates are reported in Appendix 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 (Appendix Tables C8–C10). We next turn to the precision of our measure of

---

[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 Appendix 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 Appendix Figure C1.

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

38

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 (Appendix Table C11). We also examine the robustness of our findings to the omission of influential observations (Appendix 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 check 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 (Figure 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 (Figure 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. Lastly, we also find that our conclusion regarding the interaction between hydrocortisone and coethnicity is not altered by our sensitivity checks (Figure 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.

5.1. Heterogeneous effects

We now turn an examination of two pre-specified 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, as well as by providing instructions orally through in-person and pre-recorded 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 Appendix 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 Appendix Tables C15 and C16. In line with the discussion above, the estimated coethnicity effect is consistently twice as large in the DG, TG1, and the 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.

6. 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, old), gender, and mother tongue, which is a direct indicator of the ethnicity.

We found some limited but non-robust 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.

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth*, 2003, *8*, 155–194.

_ , **Reza Baquir, and William Easterly**, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 1999, *114* (4), 1243–1284.

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

_ , **Stelios Michalopoulos, and Elias Papaioannou**, "Ethnic Inequality," *Journal of Political Economy*, 2016, *124* (2), 428–488.

Angrist, Joshua D., "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors," *Journal of Business & Economic Statistics*, 2001, *19* (1), 2–28.

_ **and Jörn-Steffen Pischke**, *Mostly Harmless Econometrics*, New York: Princeton University Press, 2009.

Balliet, Daniel, Junhui Wu, and Carsten K.W. De Dreu, "Ingroup Favoritism in Cooperation: A Meta-Analysis," *Psychological Bulletin*, 2014, *140* (6), 1556–1581.

Bardsley, Nicholas, "Dictator game giving: Altruism or artefact?," *Experimental Economics*, 2008, *11* (2), 122–133.

Bauer, Michal, Christopher Blattman, Julie Chytilova, Joseph Henrich, Edward Miguel, and Tamar Mitts, "Can War Foster Cooperation?," *Journal of Economic Perspectives*, 2016, *30* (3), 249–274.

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

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel N. Posner, Bertil Tungodden, and Kelly Zhang, "Ethnically Biased? Experimental Evidence from Kenya," *Journal of the European Economic Association*, 2020, *18* (1), 134–164.

Blouin, Arthur and Sharun W. Mukand, "Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda," *Journal of Political Economy*, 2019, *127* (3), 1008–1062.

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

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

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

Chen, Yan and Sherry Xin Li, "Group Identity and Social Preferences," *American Economic Review*, 2009, *99* (1), 431–457.

Coates, J. M. and J. Herbert, "Endogenous Steroids and Financial Risk Taking on a London Trading Floor," *Proceedings of the National Academy of Sciences*, April 2008, *105* (16), 6167–6172.

Cornelisse, Sandra, Vanessa van Ast, Johannes Haushofer, Maayke Seinstra, and Marian Joels, "Time-Dependent Effect of Hydrocortisone Administration on Intertemporal Choice," *SSRN Electronic Journal*, 2014.

de Kloet, E. Ron, "Hormones and the Stressed Brain," *Annals of the New York Academy of Sciences*, 2004, *1018* (1), 1–15.

_ , Marian Joëls, and Florian Holsboer, "Stress and the Brain: From Adaptation to Disease," *Nature Reviews Neuroscience*, 2005, *6*, 463–475.

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review*, 2018, *108* (11), 3266–3302.

Delaney, Liam, Günther Fink, and Colm P. Harmon, "Effects of Stress on Economic Decision-Making: Evidence from Laboratory Experiments," 2014. IZA Discussion Paper No. 8060.

Depetris-Chauvin, Emilio, "State History and Contemporary Conflict: Evidence from Sub-Saharan Africa," 2016. ISSN Working Paper 475.

Easterly, William and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics*, 1997, *112*, 1203–1250.

Ferree, Karen E., Clark C. Gibson, and James D. Long, "Voting Behavior and Electoral Irregularities in Kenya's 2013 Election," *Journal of Eastern African Studies*, 2014, *8* (1), 153–172.

Fershtman, Chaim and Uri Gneezy, "Discrimination in a Segmented Society: An Experimental Approach," *Quarterly Journal of Economics*, 2001, *116*, 351–377.

Franck, Raphael and Ilia Rainer, "Does the Leader's Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa," *American Political Science Review*, 2012, *106* (2), 294–325.

Gomila, Robin, "Logistic or Linear? Estimating Causal Effects of Experimental Treatment on Binary Outcomes Using Regression Analysis," *Journal of Experimental Psychology*, 2020.

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein, "Why Does Ethnic Diversity Undermine Public Goods Provision?," *American Political Science Review*, 2007, *101* (4), 709–725.

_ , _ , _ , and _ , *Coethnicity: Diversity and the Dilemmas of Collective Action* 2009.

Haushofer, Johannes and Jeremy Shapiro, "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya," *Quarterly Journal of Economics*, November 2016, *131* (4), 1973–2042.

_ , Chaning Jang, and John Lynham, "Stress and Temporal Discounting: Do Domains Matter?," 2015. Working paper.

_ , Marie Collins, Giovanna De Giusti, Joseph Njoroge, Amos Odero, Cynthia Onyango, James Vancel, Chaning Jang, Maneesh Kuruvilla, and Conor Hughes, "A Methodology for Laboratory Experiments in Developing Countries: Examples from the Busara Center," 2014. Working paper, Busara Center.

_ , **Matthieu Chemin, Chaning Jang, and Justin Abraham**, "Economic and Psychological Effects of Health Insurance and Cash Transfers: Evidence from a Randomized Experiment in Kenya," *Journal of Development Economics*, 2020, *144*, 102416.

_ , **Sandra Cornelisse, Maayke Seinstra, Ernst Fehr, Marian Joëls, and Tobias Kalenscher**, "No Effects of Psychosocial Stress on Intertemporal Choice," *PLoS ONE*, 2013, *8*, e78597.

_ , **Sara Lowes, Nathan Nunn, and Nancy Qian**, "Stress and In-Group Preferences: Pre-Analysis Plan," 2019.

Henckens, Marloes J.A.G., Guido A. van Wingen, Marian Joëls, and Guillén Fernández, "Time-Dependent Corticosteroid Modulation of Prefrontal Working Memory Processing," *Proceedings of the National Academy of Sciences*, 2011, *108* (14), 5801–5806.

_ , _ , _ , **and** _ , "Time-Dependent Effects of Cortisol on Selective Attention and Emotional Interference: A Functional MRI Study," *Frontiers in Integrative Neuroscience*, 2012, *6.*

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

Hewstone, Miles, Mark Rubin, and Hazel Willis, "Intergroup Bias," *Annual Review of Psychology*, 2002, *53*, 575–604.

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

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

Jakiela, Pamela and Owen Ozier, "The Impact of Violence on Individual Risk Preferences: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 2019, *101* (3), 547–559.

Kandasamy, Narayanan, Ben Hardy, Lionel Page, Markus Schaffner, Johann Graggaber, Andrew S. Powlson, Paul C. Fletcher, Mark Gurnell, and John Coates, "Cortisol Shifts Financial Risk Preferences," *Proceedings of the National Academy of Sciences*, 2014, *111* (9), 3608–3613.

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

_ , **Karl-Martin Pirke, and Dirk H. Hellhammer**, "The 'Trier Social Stress Test' – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting," *Neuropsychobiology*, 1993, *28* (1-2), 76–81.

Levitt, Steven D. and John A. List, "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?," *Journal of Economic Perspectives*, 2007, *21* (2), 153–174.

List, John A., "On the Interpretation of Giving in Dictator Games," *Journal of Political Economy*, 2007, *115* (3), 482–493.

_ , "Non Est Disputandum de Generalizability? A Glimpse into the External Validity Trial," 2020. NBER Working Paper 27535.

_ , **Azeem M. Shaikh, and Yang Xu**, "Multiple Hypothesis Testing in Experimental Economics," *Experimental Economics*, 2019, *22* (4), 773–793.

Lowes, Sara, Nathan Nunn, James A. Robinson, and Jonathan Weigel, "Understanding Ethnic Identity in Africa: Evidence from the Implicit Association Test," *American Economic Association: Papers and Proceedings*, 2015, *105* (5), 340–345.

Lupien, S.J., F. Maheu, M. Tu, A. Fiocco, and T.E. Schramek, "The Effects of Stress and Stress Hormones on Human Cognition: Implications for the Field of Brain and Cognition," *Brain and Cognition*, 2007, *65*, 209–237.

Macchiavello, Rocco and Ameet Morjaria, "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports," *American Economic Review*, 2015, *105* (9), 2911–2945.

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

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

Marx, Benjamin, Thomas M. Stoker, and Tavneet Suri, "There Is No Free House: Ethnic Patronage in a Kenyan Slum," *American Economic Journal: Applied Economics*, October 2019, *11* (4), 36–70.

Michalopoulos, Stelios, "The Origins of Ethnolinguistic Diversity," *American Economic Review*, 2012, *102* (4), 1508–39.

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

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

Porcelli, Anthony J. and Mauricio R. Delgado, "Acute Stress Modulates Risk Taking in Financial Decision Making," *Psychological Science*, 2009, *20* (3), 278–283.

Rasul, Imran and Daniel Rogger, "The Impact of Ethnic Diversity in Bureaucracies: Evidence from the Nigerian Civil Service," *American Economic Review*, 2015, *105* (5), 457–461.

Riis-Vestergaard, Michala Iben, Vanessa A. van Ast, Sandra Cornelisse, Marian Joëls, and Johannes Haushofer, "The Effect of Hydrocortisone Administration on Intertemporal Choice," *Psychoneuroendocrinology*, 2018, *88*, 173–182.

Schwabe, Lars, Martin Tegenthoff, Oliver Höffken, and Oliver T Wolf, "Concurrent Glucocorticoid and Noradrenergic Activity Shifts Instrumental Behavior from Goal-Directed to Habitual Control," *Journal of Neuroscience*, 2010, *30* (24), 8190–8196.

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

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

Vinkers, Christiaan H., Jelle V. Zorn, Sandra Cornelisse, Susanne Koot, Lotte C. Houtepen, Berend Olivier, Joris C. Verster, René S. Kahn, Marco P.M. Boks, Tobias Kalenscher, and Marian Joëls, "Time-Dependent Changes in Altruistic Punishment Following Stress," *Psychoneuroendocrinology*, 2013, *38* (9), 1467–1475.

von Dawans, Bernadette, Urs Fischbacher, Clemens Kirschbaum, Ernst Fehr, and Markus Heinrichs, "The Social Dimension of Stress Reactivity: Acute Stress Increases Prosocial Behavior in Humans," *Psychological Science*, 2012, *23* (6), 651–660.

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

Yanagizawa-Drott, David, "Propaganda and Conflict: Evidence from the Rwandan Genocide," *Quarterly Journal of Economics*, 2014, *129* (4), 1947–1994.

Stress, Ethnicity, and Prosocial Behavior
(Online Appendix – For Online Publication)[*]

Johannes Haushofer[†]   Sara Lowes[‡]   Abednego Musau[§]   David Ndetei[¶]

Nathan Nunn[‖]   Moritz Poll[**]   Nancy Qian[††]

30 July 2022

Appendix A. Study Logistics and Sample

A.1. Pre-Analysis Plan Deviations

Table A1: Pre-Analysis Plan Deviations

Description	Timing
Added an additional line in the instructions to underline that Participant 2 in the dictator game does not take any action	Before start of data collection
Added an additional comprehension question to the dictator game that tests whether they retain the fact that Participant 2 in the dictator game does not take action.	Before start of data collection
Added an additional dictator game screen after the task where we inform the participants that they are a Participant 2 for someone else as well and will receive an additional payout accordingly.	Before start of data collection
Analyzed average results across games in Table 3.	During analysis.
Robustness check controlling not only for the political coalition present at the time of data collection (as pre-specified), but also for the prevalent coalition of the 2007 presidential election in Tables C2 and C3.	During analysis.
Robustness check controlling for imbalanced demographics in Table C5.	During analysis.

Notes: This table lists the timing of changes to the study design and analysis after the submission of the pre-analysis plan as well as their timing.
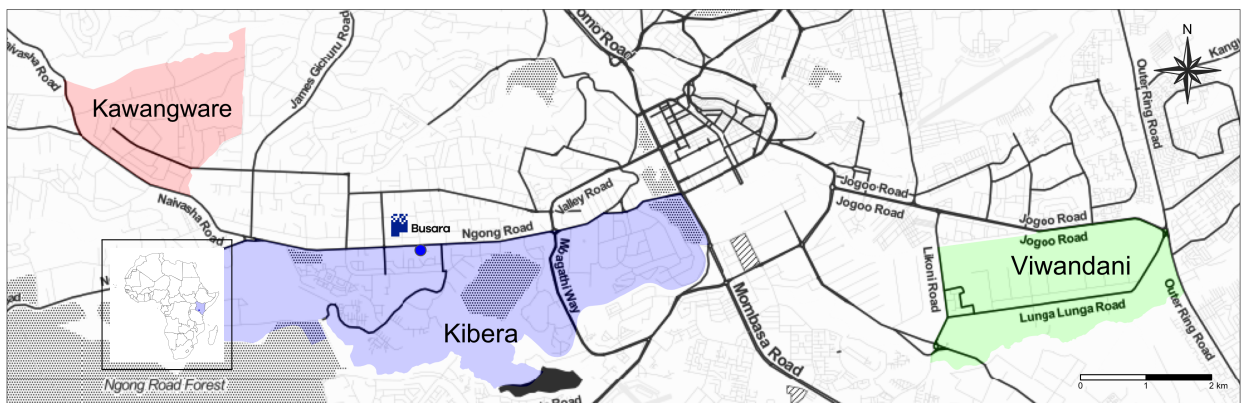
A.2. Sampling area

Figure A1: Map of Nairobi, Kenya, with shaded settlements of origin of participant pool
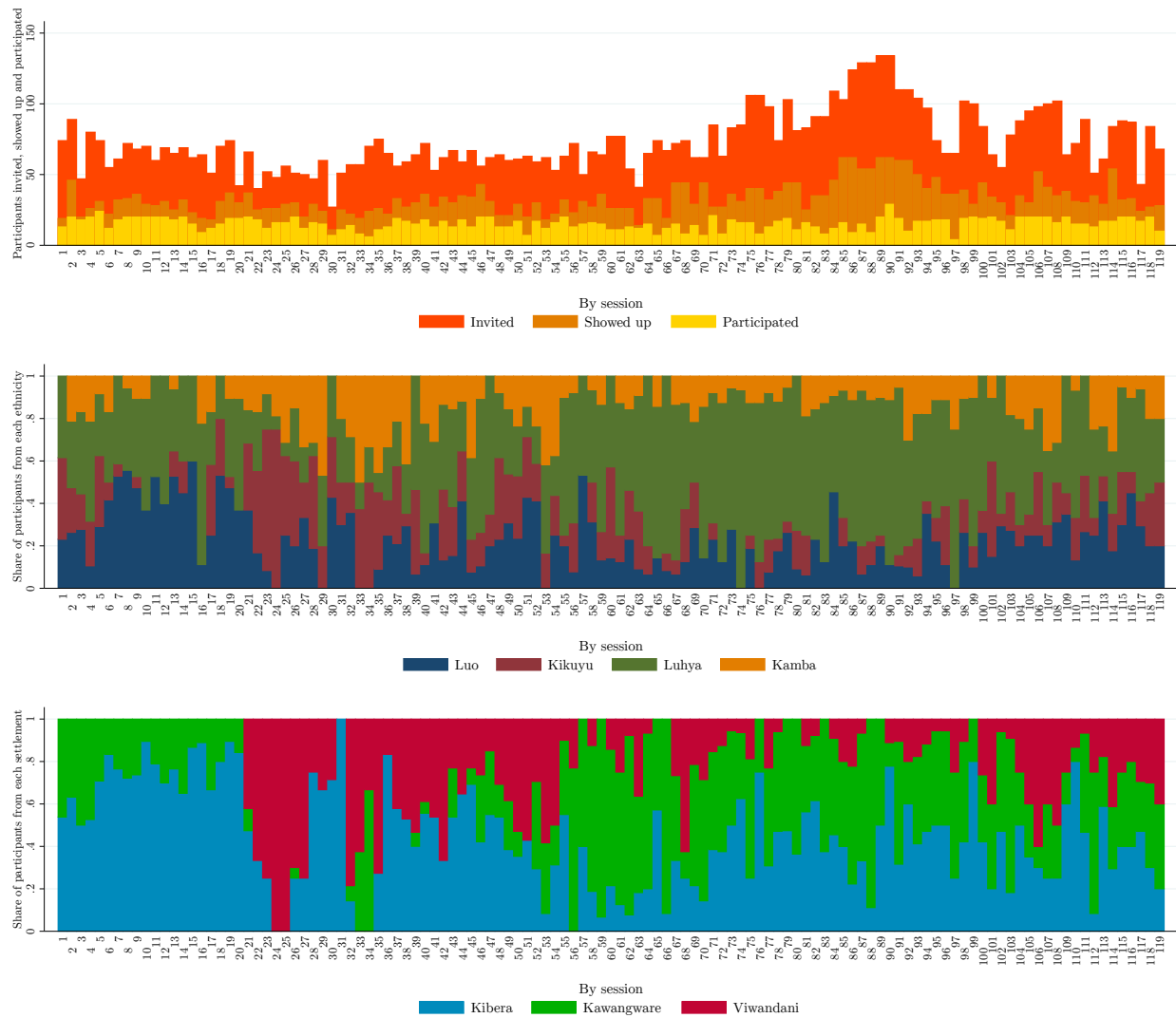
A.3. Sample selection

Figure A2: Show-up, ethnic and settlement session composition across experimental sessions (1–119)

Table A2: Sample selection

	(1) Kenya	(2) Nairobi	(3) Pool	(4) Sampled	(5) Contacted	(6) Reached	(7) Signed up	(8) Showed up	(9) Participated	(10) Final sample
Female										
Mean	0.5049	0.4885	0.4556	0.5039	0.5146	0.5157	0.4499	0.4678	0.4637	0.4641
SD	(0.5000)	(0.4999)	(0.4980)	(0.5000)	(0.4998)	(0.4998)	(0.4975)	(0.4990)	(0.4988)	(0.4989)
P((i)=(i-1))		[0.0000]***	[0.0000]***	[0.0000]***	[0.0129]**	[0.8242]	[0.0000]***	[0.0846]*	[0.7792]	[0.9801]
Individuals	47,564,296	3,138,369	76,002	30,986	24,592	14,211	7,886	3,273	1,874	1,784
Young (18–35)										
Mean	0.3057	0.4286	0.7636	0.8792	0.8742	0.8846	0.9105	0.9212	0.9584	0.9574
SD	(0.4607)	(0.4949)	(0.4249)	(0.3259)	(0.3317)	(0.3195)	(0.2854)	(0.2695)	(0.1998)	(0.2020)
P((i)=(i-1))		[0.0000]***	[0.0000]***	[0.0000]***	[0.0743]*	[0.0025]***	[0.0000]***	[0.0686]*	[0.0000]***	[0.8829]
Individuals	47,564,296	4,396,828	75,677	31,068	24,676	14,244	7,903	3,273	1,874	1,784
Middle-Aged (36–50)										
Mean	0.1330	0.1623	0.1636	0.0999	0.1016	0.1040	0.0797	0.0776	0.0416	0.0426
SD	(0.3395)	(0.3687)	(0.3699)	(0.2999)	(0.3021)	(0.3052)	(0.2709)	(0.2676)	(0.1998)	(0.2020)
P((i)=(i-1))		[0.0000]***	[0.3373]	[0.0000]***	[0.5293]	[0.4487]	[0.0000]***	[0.7066]	[0.0000]***	[0.8829]
Individuals	47,564,296	4,396,828	75,677	31,068	24,676	14,244	7,903	3,273	1,874	1,784
Luo										
Mean	0.1065		0.2062	0.2139	0.2228	0.2337	0.2187	0.2536	0.2396	0.2410
SD	(0.3085)		(0.4046)	(0.4101)	(0.4161)	(0.4232)	(0.4134)	(0.4351)	(0.4269)	(0.4278)
P((i)=(i-1))			[0.0000]***	[0.0065]***	[0.0116]**	[0.0138]**	[0.0111]**	[0.0001]***	[0.2636]	[0.9190]
Individuals	47,564,296		63,550	30,994	24,672	14,243	7,901	3,273	1,874	1,784
Kikuyu										
Mean	0.1713		0.1707	0.2028	0.1864	0.1910	0.1742	0.1662	0.1708	0.1704
SD	(0.3768)		(0.3762)	(0.4021)	(0.3895)	(0.3931)	(0.3793)	(0.3723)	(0.3764)	(0.3761)
P((i)=(i-1))			[0.6791]	[0.0000]***	[0.0000]***	[0.2639]	[0.0019]***	[0.3109]	[0.6744]	[0.9773]
Individuals	47,564,296		63,550	30,994	24,672	14,243	7,901	3,273	1,874	1,784
Luhya										
Mean	0.1435		0.3553	0.3950	0.4120	0.4024	0.4396	0.4397	0.4456	0.4467
SD	(0.3505)		(0.4786)	(0.4889)	(0.4922)	(0.4904)	(0.4964)	(0.4964)	(0.4972)	(0.4973)
P((i)=(i-1))			[0.0000]***	[0.0000]***	[0.0000]***	[0.0615]*	[0.0000]***	[0.9928]	[0.6811]	[0.9429]
Individuals	47,564,296		63,550	30,994	24,672	14,243	7,901	3,273	1,874	1,784
Kamba										
Mean	0.0981		0.1163	0.1531	0.1599	0.1580	0.1628	0.1341	0.1430	0.1418
SD	(0.2974)		(0.3206)	(0.3601)	(0.3665)	(0.3647)	(0.3692)	(0.3408)	(0.3502)	(0.3490)
P((i)=(i-1))			[0.0000]***	[0.0000]***	[0.0282]**	[0.6244]	[0.3511]	[0.0001]***	[0.3732]	[0.9178]
Individuals	47,564,296		63,550	30,994	24,672	14,243	7,901	3,273	1,874	1,784
Primary education										
Mean	0.4978	0.3325	0.2819	0.2911	0.2995	0.2940	0.2382	0.2135	0.1729	0.1749
SD	(0.5000)	(0.4711)	(0.4499)	(0.4543)	(0.4580)	(0.4556)	(0.4260)	(0.4098)	(0.3783)	(0.3800)
P((i)=(i-1))		[0.0000]***	[0.0000]***	[0.0039]***	[0.0335]**	[0.2628]	[0.0000]***	[0.0050]***	[0.0005]***	[0.8736]
Individuals	36,212,477	3,787,354	61,231	29,727	24,014	13,805	7,691	3,256	1,874	1,784
Secondary education										
Mean	0.2453	0.3393	0.5205	0.5795	0.5698	0.5656	0.6080	0.6087	0.6307	0.6289
SD	(0.4302)	(0.4735)	(0.4996)	(0.4937)	(0.4951)	(0.4957)	(0.4882)	(0.4881)	(0.4827)	(0.4832)
P((i)=(i-1))		[0.0000]***	[0.0000]***	[0.0000]***	[0.0240]**	[0.4272]	[0.0000]***	[0.9423]	[0.1184]	[0.9097]
Individuals	36,212,477	3,787,354	61,231	29,727	24,014	13,805	7,691	3,256	1,874	1,784
College education										
Mean	0.1052	0.2203	0.1765	0.1259	0.1272	0.1360	0.1504	0.1735	0.1937	0.1934
SD	(0.3068)	(0.4145)	(0.3813)	(0.3317)	(0.3332)	(0.3428)	(0.3575)	(0.3788)	(0.3953)	(0.3951)
P((i)=(i-1))		[0.0000]***	[0.0000]***	[0.0000]***	[0.6425]	[0.0142]**	[0.0037]***	[0.0024]***	[0.0707]*	[0.9806]
Individuals	36,212,477	3,787,354	61,231	29,727	24,014	13,805	7,691	3,256	1,874	1,784
Observations	47,564,300	4,397,073	76,002	31,073	24,678	14,246	7,904	3,273	1,874	1,784
Retention		9.24%	1.73%	40.88%	79.42%	57.73%	55.48%	41.41%	57.26%	95.20%

Notes: This table illustrates the sampling funnel from the population to the final sample. Every block of four rows indicates a demographic variable. Within each block the first row indicates the sample mean. The second row in parentheses indicates the sample standard deviation. The third row indicates the p-value of a simple t-test of the block's statistic with that in the block to the left, i.e. one stage earlier in the sampling funnel. Significance is denoted by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The fourth row indicates the number of observations at that stage for which the demographic is known. Across demographics these can vary if not all statistics are available for all individuals. From left to right, the columns indicate how the potential sample narrows down into the actual sample. We start by reproducing statistics from the 2019 census in columns (1) and (2) from https://www.knbs.or.ke/publications/. Mother tongue information was not published for Nairobi. The Busara Center has a Nairobi-based participant pool described in column (3) as of the time of data collection. Based on the study's inclusion criteria, participants are sampled from this pool as shown in column (4). Enumerators would then call participants from this list until they had confirmed with enough participants for the next study session to more or less fill it. Column (5) shows demographics for all individuals ever contacted in this way. Column (6) features individuals who could be reached by phone call. Column (7) indicates those individuals who agreed to come. Column (8) describes those individuals who showed up to the session. Column (9) contains those individuals who started a session after dropping participants who were screened out for medical reasons or did not consent to the study. Column (10) concludes with the final sample after eliminating participants who left part way through the session or whose data were not properly recorded.

A.4. Schedule of tasks and treatments

The sessions followed the schedule of tasks and treatments that are outlined below. We randomized the game order at the session level so that either the dictator game and choose your dictator game precede the trust game – stage 1 and trust game – stage 2 or vice versa in both instructions and play. Figure A3 illustrates the average session timing by task order from the recorded time sheets. An example for the game screens that participants saw is provided in Figure A4.

1. Participant identification
2. Welcome
3. Consent and Nurse's Checklist
4. Vital Signs (Heart Rate, Blood Pressure, Temperature)
5. Nurse's Participant meetings
6. Send eligible participants to the lab and ineligible participants home
7. Demographics & Introduction to computer interface on zTree
8. Primer Profile
9. General session instructions
10. Dictator Game **or** Trust Game – Stage 1: Instructions and comprehension
11. Choose Your Dictator Game **or** Trust Game – Stage 2: Instructions and comprehension
12. Trust Game – Stage 1 **or** Dictator Game: Instructions and comprehension
13. Trust Game – Stage 2 **or** Choose Your Dictator Game: Instructions and comprehension
14. Salivette 1: White
15. **Placebo / Drug administration**
16. Slider Instructions
17. Vital Signs (Heart Rate, Blood Pressure)
18. Break
19. Re-Primer Profile
20. Salivette 2: Orange
21. Task 1: Dictator Game **or** Trust Game – Stage 1
22. Salivette 3: Blue
23. Task 2: Choose Your Dictator Game **or** Trust Game – Stage 2
24. Salivette 4: Green
25. Task 3: Trust Game – Stage 1 **or** Dictator Game
26. Salivette 5: Purple
27. Task 4: Trust Game – Stage 2 **or** Choose Your Dictator Game
28. Salivette 6: Red
29. Social Proximity Survey
30. Pill Guessing Module
31. Experimenter Demand Effects Module
32. Additional Demographics
33. Debrief
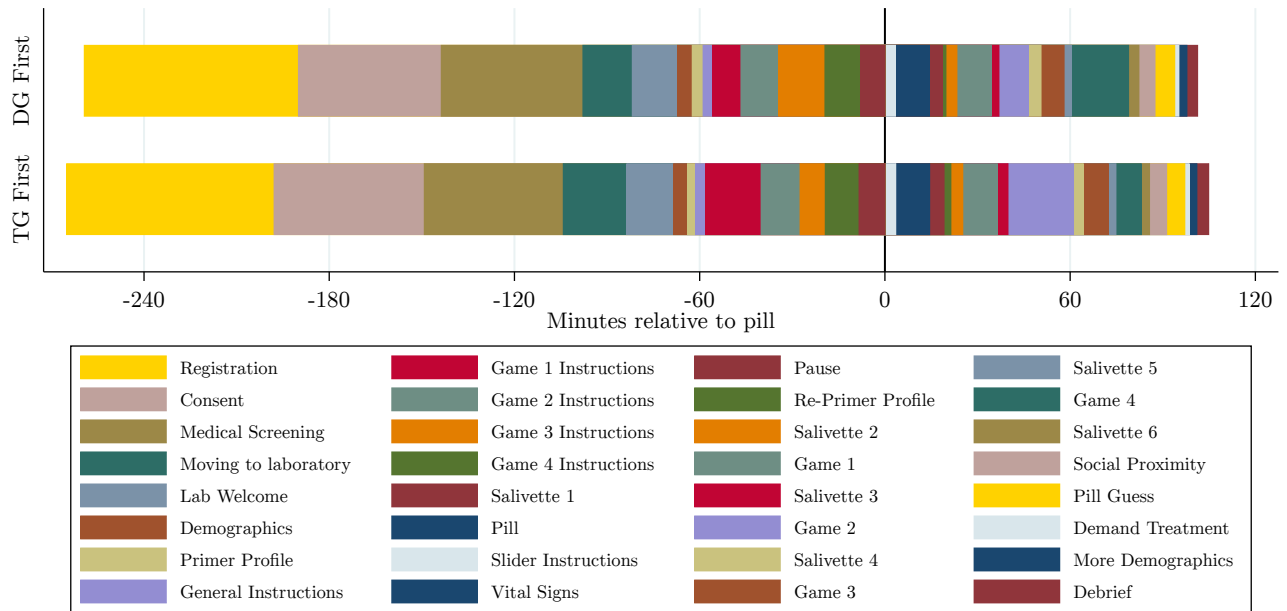
Figure A3: Session Time Visualization
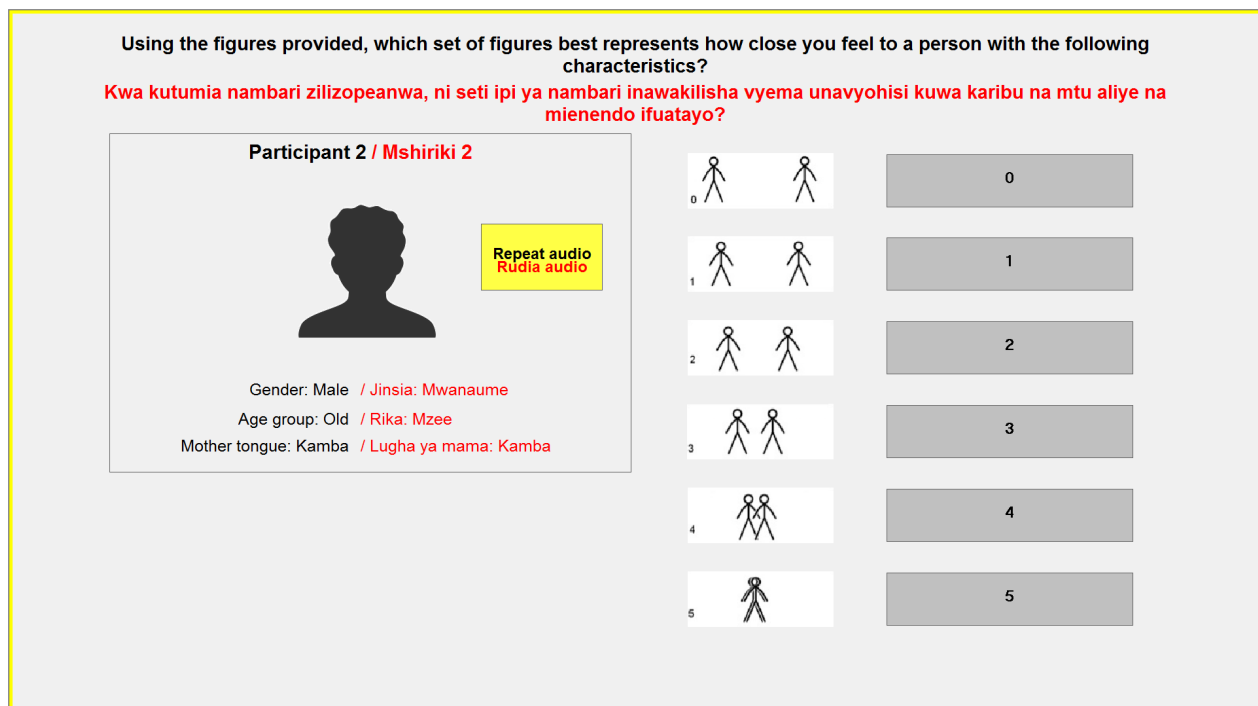
Figure A4: Example Screen

A.5. Sample balance

Table A3: Balance

Variable	(1) Placebo Mean/SE	(2) Hydrocortisone Mean/SE	(3) (1)-(2) Difference/SE
Female (%)	45.14	47.72	-2.58
	(1.65)	(1.69)	(2.36)
Age	24.92	24.57	0.35
	(0.17)	(0.17)	(0.24)
Mother tongue: Luo (%)	23.29	24.94	-1.65
	(1.41)	(1.46)	(2.03)
Mother tongue: Kikuyu (%)	15.34	18.79	-3.45*
	(1.20)	(1.32)	(1.78)
Mother tongue: Luhya (%)	46.14	43.17	2.97
	(1.66)	(1.67)	(2.35)
Mother tongue: Kamba (%)	15.23	13.10	2.13
	(1.19)	(1.14)	(1.65)
From Kibera (%)	46.91	45.33	1.58
	(1.66)	(1.68)	(2.36)
From Kawangware (%)	30.79	32.00	-1.21
	(1.53)	(1.58)	(2.20)
From Viwandani (%)	22.30	22.67	-0.37
	(1.38)	(1.41)	(1.98)
Years lived in Nairobi	12.72	12.66	0.06
	(0.28)	(0.27)	(0.39)
Average monthly income (KES)	7338.25	7585.77	-247.52
	(247.96)	(272.73)	(368.14)
Unemployed (%)	68.21	68.91	-0.69
	(1.55)	(1.56)	(2.20)
Self-employed (%)	15.12	14.58	0.54
	(1.19)	(1.19)	(1.69)
Attained some primary education (%)	18.43	16.51	1.92
	(1.29)	(1.25)	(1.80)
Attained some secondary education (%)	64.35	61.39	2.96
	(1.59)	(1.64)	(2.29)
Attained some college education (%)	17.11	21.64	-4.53**
	(1.25)	(1.39)	(1.87)
Number of children	0.90	0.88	0.03
	(0.04)	(0.04)	(0.06)
Married or cohabiting (%)	35.65	33.83	1.82
	(1.59)	(1.60)	(2.26)
Single, divorced or widowed (%)	63.91	65.83	-1.92
	(1.60)	(1.60)	(2.26)
Christian (%)	96.91	97.84	-0.93
	(0.58)	(0.49)	(0.76)
Principal component of asset list	-0.06	0.06	-0.12*
	(0.05)	(0.05)	(0.07)
N	906	878	
F-test of joint significance (p-value)			0.15
F-test, number of observations			1784

Notes: The value displayed for t-tests are the differences in the means across the groups. The value displayed for F-tests are p-values. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

A.6. Distributions

(a) Dictator Game

(b) Trust Game – Stage 1

(c) Trust Game – Stage 2

(d) Social Proximity Survey
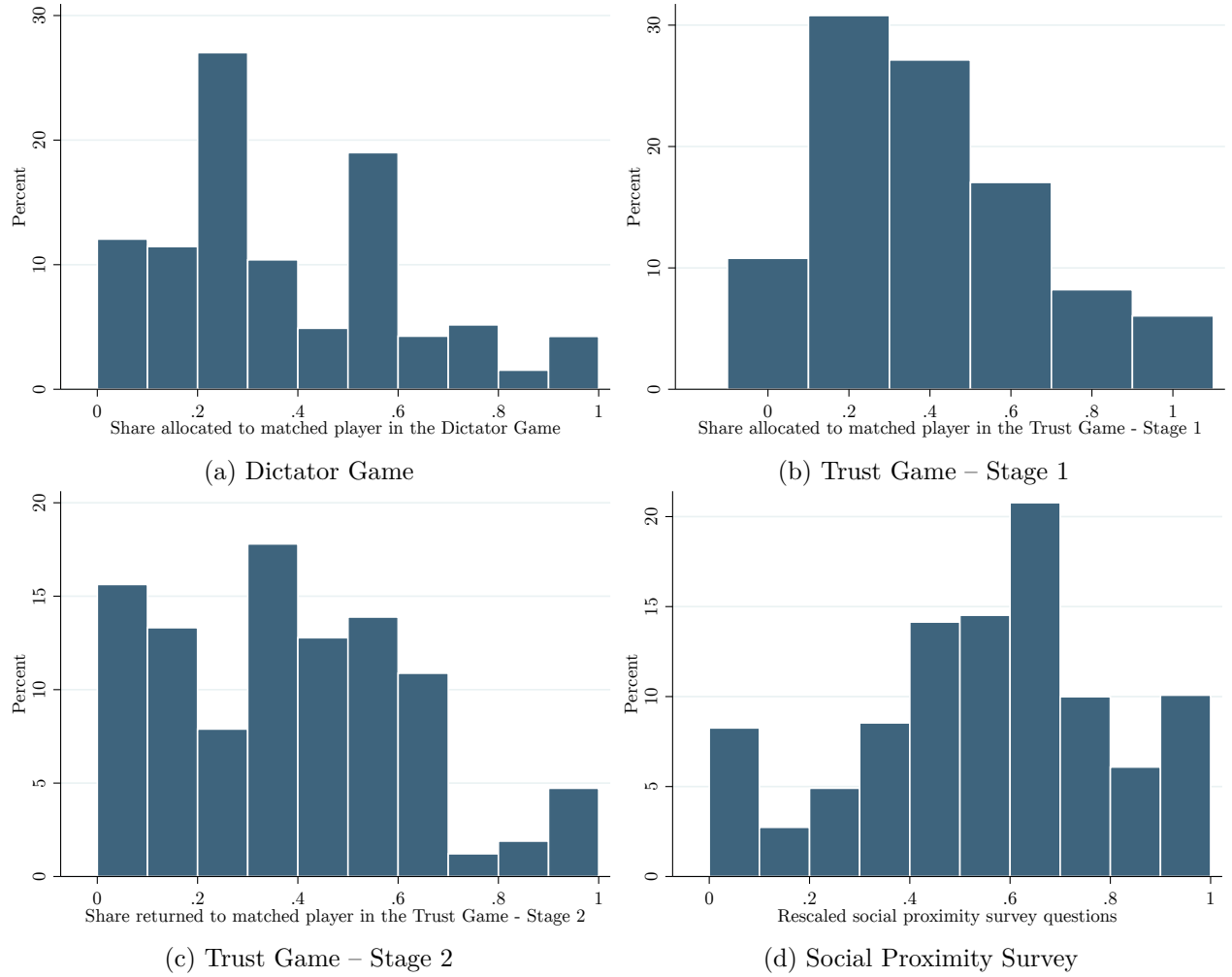
Figure A5: Distributions of allocations in the dictator and trust games and survey responses in the social proximity survey.

Appendix B. Additional results

B.1. Choose your dictator game: Linear probability versus conditional logit

Table B1 compares the results of the choose your dictator game in two different specifications, a linear probability model based on OLS and a conditional logit specification. While the interpretation of coefficient magnitude differs, significance levels are comparable.

B.2. Trust game – stage 2: By amount allocated

This section breaks up the trust game – stage 2 results by how much was allocated to the player in trust game – stage 1. This allows us to look for biases in scenarios where Player 1 was (not) particularly trusting. The first column of Table B2 presents the results of the trust game – stage 2

Table B1: Choose your dictator game: Linear probability versus conditional logit

| | Choose your dictator game LPM and logit | |
| | (1) | (2) |
	Choose Your Dictator Game Linear Probability	Choose Your Dictator Game Conditional Logit
Constant	0.6231***	
	(0.0061)	
Same ethnicity	0.0611***	0.0607***
	(0.0097)	(0.0097)
Hydrocortisone × Same ethnicity	0.0228*	0.0234*
	(0.0135)	(0.0136)
Same gender	−0.0104	−0.0104
	(0.0094)	(0.0093)
Hydrocortisone × Same gender	−0.0327**	−0.0328**
	(0.0132)	(0.0132)
Same age group	0.0444**	0.0445**
	(0.0219)	(0.0222)
Hydrocortisone × Same age group	−0.0400***	−0.0403***
	(0.0145)	(0.0145)
Matched player is Luo	−0.0091	−0.0091
	(0.0081)	(0.0081)
Matched player is Kikuyu	−0.0318***	−0.0319***
	(0.0082)	(0.0082)
Matched player is Luhya	−0.0220***	−0.0220***
	(0.0083)	(0.0083)
Matched player is female	0.0054	0.0053
	(0.0066)	(0.0066)
Matched player is middle-aged	0.0318	0.0320
	(0.0214)	(0.0217)
Matched player is old	0.0364	0.0366
	(0.0225)	(0.0229)
Participants	1,784	1,784
Decisions per participant	6	6
Decisions	10,704	10,704

Notes: The table presents the results of the choose your dictator game in both a linear probability and conditional logit specification. Standard errors clustered at the participant level are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

results using allocated amount fixed effects, i.e. an amount-specific intercept, while the remaining columns present separate estimations for each possible amount allocated in stage 1.

B.3. Social proximity: By components

We report estimates for by each of three components of the social proximity index: likelihood to be friends, trust and perceived closeness.

B.4. Multiple Hypothesis Testing

Table B4 reports the results of the main specification after column-wise correcting for multiple hypothesis testing.

Table B2: Trust game – stage 2 by amount allocated

	(1) Trust Game 2 Amount FE	(2) Trust Game 2 120 KES	(3) Trust Game 2 240 KES	(4) Trust Game 2 360 KES	(5) Trust Game 2 480 KES	(6) Trust Game 2 600 KES
Sample mean	0.3804	0.3919	0.3767	0.3783	0.3718	0.3833
Sample standard deviation	(0.2477)	(0.2557)	(0.2419)	(0.2455)	(0.2443)	(0.2504)
Panel A: Average hydrocortisone effect[a]						
Hydrocortisone effect	−0.0103	−0.0114	−0.0071	−0.0093	−0.0103	−0.0136
	(0.0095)	(0.0100)	(0.0096)	(0.0098)	(0.0099)	(0.0102)
Panel B: Average coethnicity effect[b]						
Coethnicity effect	0.0015	−0.0002	0.0010	0.0043	0.0037	−0.0012
	(0.0022)	(0.0035)	(0.0031)	(0.0031)	(0.0030)	(0.0031)
Panel C: Interaction of hydrocortisone and coethnicity[c]						
Interaction effect	−0.0019	−0.0054	−0.0102*	0.0065	−0.0012	0.0007
	(0.0042)	(0.0066)	(0.0061)	(0.0058)	(0.0058)	(0.0058)
Participants	1,784	1,784	1,784	1,784	1,784	1,784
Decisions per participant	30	6	6	6	6	6
Decisions	53,520	10,704	10,704	10,704	10,704	10,704

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Standard errors clustered at the participant level are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Social proximity by components

	(1) Social Proximity	(2) Likely to be friends	(3) Trust	(4) Closeness
Sample mean	0.5481	3.4385	2.9713	2.7088
Sample standard deviation	(0.2575)	(1.2287)	(1.2661)	(1.5245)
Panel A: Average hydrocortisone effect[a]				
Hydrocortisone effect	0.0123	0.0978**	0.0313	0.0232
	(0.0099)	(0.0450)	(0.0477)	(0.0584)
Panel B: Average coethnicity effect[b]				
Coethnicity effect	0.0817***	0.3066***	0.3607***	0.3908***
	(0.0052)	(0.0251)	(0.0268)	(0.0306)
Panel C: Interaction of hydrocortisone and coethnicity[c]				
Interaction effect	−0.0051	0.0090	−0.0385	−0.0388
	(0.0101)	(0.0496)	(0.0523)	(0.0591)
Participants	1,784	1,784	1,784	1,784
Decisions per participant	4	4	4	4
Decisions	7,136	7,136	7,136	7,136

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Likelihood to be friends and trust are set on Likert scales from 1 to 5 while closeness is set on a Likert scale from 0 to 5. Standard errors clustered at the participant level are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.5. 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. The estimates full estimates are
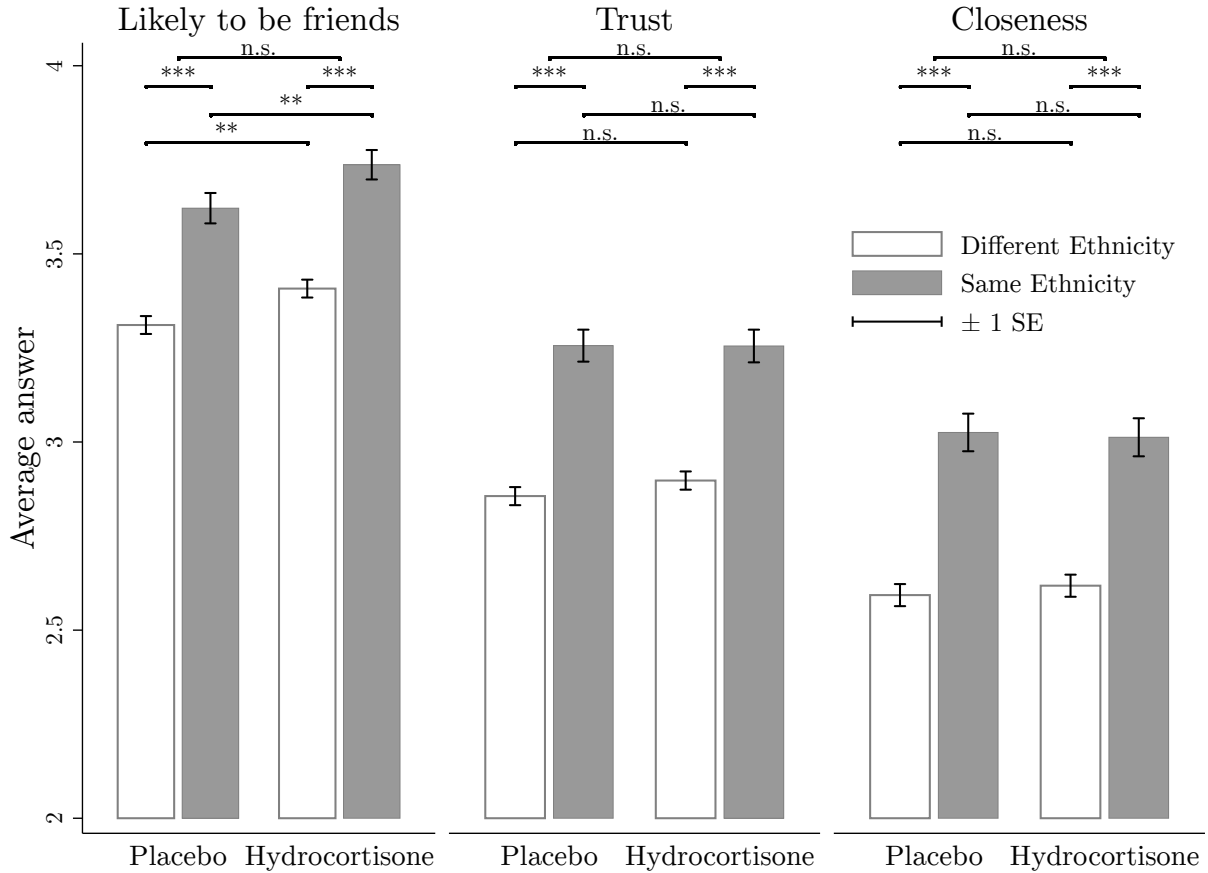
Figure B1: Mean comparison of social proximity survey components

reported in Table B5.

Appendix C. Robustness checks

C.1. Randomization inference

In this section, we reproduce our results using randomization inference. We re-assign all levels of treatment assignment (pill, ethnicity, gender, and age pairing) 10,000 times and obtain the randomization inference *p*-value as the share of pseudo treatment assignments that resulted in more extreme results than ours. The results are practically identical.

C.2. Political coalitions

We explore political coalitions as an alternative in-group definition in a robustness check. Political coalitions change over time. Table C2 reports the estimates using the 2013 and 2017 election configuration (Luo, Luhya, and Kamba in a coalition and Kikuyu in another), which is the coalition at the time the experiments took place. Previous studies, such as Hjort (2014) and Berge, Bjorvatn, Galle, Miguel, Posner, Tungodden and Zhang (2020), use the coalition structure from the 2007 election (Luo and Luhya in a coalition and Kikuyu and Kamba in another). For comparability to

Table B4: Multiple Hypothesis Testing Correction

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.2870	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.4524)	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		-0.0157	-0.0225**	-0.0103	0.0123
		[0.1873]	[0.0328]	[0.6224]	[0.3768]
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0724***	0.0188***	0.0297***	0.0015	0.0817***
	[0.0001]	[0.0001]	[0.0001]	[0.7352]	[0.0001]
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0228*	0.0014	-0.0178**	-0.0019	-0.0051
	[0.0857]	[0.8447]	[0.0336]	[0.6578]	[0.6152]
Participants	1,784	1,784	1,784	1,784	1,784
Decisions per participant	3	6	6	30	4
Decisions	5,352	10,704	10,704	53,520	7,136

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Standard errors clustered at the participant level. The p-values are column-wise corrected for multiple hypothesis testing following List, Shaikh and Xu (2019) using 10,000 bootstrap draws and are reported in brackets. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

these studies and because ethnic preference may be affected by past coalitions as well, we also report estimates using these coalitions. These are reported in Table C3.

C.3. Additional covariates

We estimate a number of specifications that include session fixed effects (Table C4) and imbalanced participant characteristics (Table C5).

C.4. Comprehension

Participants were asked 23 comprehension questions relating to the four tasks (listed in Table C6 – correct answers are bolded; questions prefixed F are asked as recapitulation right before the games start). We recorded whether participants gave the correct answer to each question on first attempt (shares by question reported in Figure C1 and proportions of participants by the share of incorrect first answers in Figure C2). Table C7 shows the at our regression results are robust when dropping participants who got half or more of the game-specific comprehension questions wrong on first attempt. Since the social proximity survey did not have comprehension questions, we use comprehension scores across all games to get a sense of alertness and commitment to the study. We also tracked usage of a 'repeat' button for the headphone-delivered audio instructions. In total, 92% of the sample use the repeat button twice or less often, consistent with good comprehension.

Table B5: Estimates from specification with player 1 fixed effects

	(1) Dictator Game	(2) Trust Game 1	(3) Trust Game 2	(4) Social Proximity
Constant	0.3331***	0.3166***	0.3861***	0.5194***
	(0.0106)	(0.0142)	(0.0069)	(0.0116)
Same ethnicity	0.0177***	0.0389***	0.0038	0.0844***
	(0.0055)	(0.0060)	(0.0028)	(0.0071)
Same ethnicity × Hydrocortisone	0.0015	−0.0178**	−0.0021	−0.0052
	(0.0074)	(0.0084)	(0.0041)	(0.0101)
Same gender	0.0040	−0.0077	0.0026	0.0404***
	(0.0052)	(0.0059)	(0.0028)	(0.0068)
Same gender × Hydrocortisone	−0.0042	0.0086	−0.0013	−0.0168*
	(0.0071)	(0.0085)	(0.0042)	(0.0095)
Same age group	0.0045	0.0469***	−0.0119*	0.0079
	(0.0106)	(0.0138)	(0.0068)	(0.0123)
Same age group × Hydrocortisone	−0.0045	0.0017	−0.0005	−0.0029
	(0.0080)	(0.0093)	(0.0046)	(0.0106)
Matched player is Luo	−0.0001	0.0050	0.0001	0.0073
	(0.0048)	(0.0056)	(0.0028)	(0.0054)
Matched player is Kikuyu	−0.0009	0.0042	−0.0032	−0.0020
	(0.0048)	(0.0053)	(0.0028)	(0.0053)
Matched player is Luhya	0.0008	0.0038	−0.0008	0.0136**
	(0.0049)	(0.0055)	(0.0029)	(0.0058)
Matched player is female	0.0035	0.0122***	0.0016	0.0180***
	(0.0036)	(0.0043)	(0.0021)	(0.0047)
Matched player is middle-aged	0.0166*	0.0630***	−0.0075	−0.0350***
	(0.0098)	(0.0133)	(0.0065)	(0.0108)
Matched player is old	0.0329***	0.0817***	−0.0026	−0.0378***
	(0.0102)	(0.0138)	(0.0067)	(0.0118)
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

Notes: 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C1: Randomization inference

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
	Randomization inference p-values				
Sample mean	0.5740	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.4945)	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		−0.0159*	−0.0208**	−0.0101	0.0123
		[0.0895]	[0.0143]	[0.2680]	[0.2090]
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0668***	0.0188***	0.0297***	0.0015	0.0817***
	[0.0000]	[0.0000]	[0.0000]	[0.4945]	[0.0000]
Panel C: Interaction of coethnicity and hydrocortisone[c]					
Interaction effect	0.0228*	0.0014	−0.0178**	−0.0019	−0.0051
	[0.0919]	[0.8446]	[0.0311]	[0.6478]	[0.6136]
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. The p-values in brackets are obtained from 10,000 randomization inference draws and significance denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Political coalitions in 2017 as the measure of in-group

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5740	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.4945)	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		−0.0157*	−0.0226**	−0.0103	0.0123
		(0.0093)	(0.0093)	(0.0095)	(0.0099)
Panel B: Average political coalition effect[b]					
Political coalition effect	0.0601***	0.0165***	0.0256***	0.0018	0.0578***
	(0.0092)	(0.0048)	(0.0056)	(0.0030)	(0.0064)
Panel C: Interaction of hydrocortisone and political coalition[c]					
Interaction effect	−0.0176	0.0000	0.0001	0.0026	−0.0133
	(0.0135)	(0.0104)	(0.0109)	(0.0092)	(0.0114)
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. As an alternative in-group definition we employ the 2017 political coalitions of Kikuyu only versus Luo, Luhya, and Kamba. Standard errors clustered at the participant level are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Political coalitions in 2007 as the measure of in-group

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5740	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.4945)	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			-0.0156^*	-0.0225^{**}	-0.0104	0.0123
			(0.0093)	(0.0093)	(0.0095)	(0.0099)
Panel B: Average political coalition effect[b]					
Political coalition effect	0.0415^{***}	0.0097^{**}	0.0199^{***}	-0.0005	0.0536^{***}
	(0.0061)	(0.0038)	(0.0044)	(0.0027)	(0.0044)
Panel C: Interaction of hydrocortisone and political coalition[c]					
Interaction effect	-0.0094	0.0026	-0.0197^{**}	0.0048	-0.0054
	(0.0116)	(0.0070)	(0.0083)	(0.0052)	(0.0084)
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. As an alternative in-group definition we employ the 2007 political coalitions of Kikuyu and Kamba versus Luo and Luhya. Standard errors clustered at the participant level are reported in parentheses. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Session fixed effects

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5740	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.4945)	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			-0.0140	-0.0209^{**}	-0.0099	0.0132
			(0.0089)	(0.0090)	(0.0090)	(0.0095)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0736^{***}	0.0187^{***}	0.0300^{***}	0.0015	0.0817^{***}
	(0.0069)	(0.0039)	(0.0044)	(0.0022)	(0.0052)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0213	0.0016	-0.0179^{**}	-0.0020	-0.0050
	(0.0133)	(0.0074)	(0.0084)	(0.0042)	(0.0102)
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: Controlling for imbalanced participant characteristics

	(1) Dictator Game	(2) Trust Game 1	(3) Trust Game 2	(4) Social Proximity
Sample mean	0.3591	0.3984	0.3804	0.5481
Sample standard deviation	(0.2490)	(0.2670)	(0.2477)	(0.2575)
Panel A: Average hydrocortisone effect[a]				
Hydrocortisone effect	−0.0163*	−0.0231**	−0.0100	0.0113
	(0.0093)	(0.0093)	(0.0095)	(0.0098)
Panel B: Average coethnicity effect[b]				
Coethnicity effect	0.0188***	0.0297***	0.0015	0.0817***
	(0.0038)	(0.0043)	(0.0022)	(0.0052)
Panel C: Interaction of hydrocortisone and coethnicity[c]				
Interaction effect	0.0014	−0.0179**	−0.0019	−0.0050
	(0.0074)	(0.0084)	(0.0042)	(0.0101)
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

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C6: Comprehension questions by game

	Dictator Game
1	How many participants take part in the task? [1, **2**, 3, 4]
2	In this task, Participant 2 does not make any decisions. Payoffs are determined fully by Participant 1's choice. [**True**, False]
3	Now, imagine you are Participant 1, and Participant 2 has the following characteristics: {Gender: [Male or Female]; Age: [Youth, Middle Aged, or Old]; Mother tongue: [Luo, Kikuyu, Luhya, Kamba]}. As Participant 1, you will now decide how much of the KSH 200 to allocate to Participant 2. If you decide to allocate KSH 50 to Participant 2. How much money do you keep? [0, 20, 50, **150**, 200]
4	How much money does Participant 2 receive? [0, 20, **50**, 150, 200]
F1	How many KSH will Participant 1 receive to allocate between himself/herself and Participant 2? [0, 40, 100, **200**, 400]
F2	There are two participants in this task: Participant 1 and Participant 2. Participant 1 receives 200 KSH. Participant 1 then decides how to allocate the 200 KSH between himself/herself and Participant 2. Will you have the role of Participant 1 or Participant 2 in this task? [**Participant 1**, Participant 2]

	Choose Your Dictator Game
1	As Participant 2, you choose one person to be Participant 1 in the task. How many people do you choose from to be Participant 1? [1, **2**, 3, 4].
2	What information will Participant 1 have about you? [**Age group**, **gender**, name, **mother tongue**, height]
3	Now, imagine you choose Person A to be Participant 1 and he/she decides to allocate KSH 15 to you. How much money do you receive? [0, **15**, 100, 185, 200]
4	How much money does Person A receive? [0, 15, 100, **185**, 200]
F1	Imagine you choose Person A to be Participant 1. Person A and Person B both receive 200 KSH and decide, how much to allocate to you. Will you receive the allocation of Person A or Person B? [**Person A**, Person B]

	Trust Game – Stage 1
1	The amount of money allocated by Participant 1 for Participant 2 is multiplied by three. [**True**, False]
2	In the task, Participant 1 chooses how much of KSH 200 to allocate to Participant 2. This amount is multiplied by three. Participant 2 decides how much of this increased amount to send to Participant 1. Whose role will you play in this task? [**Participant 1**, Participant 2]

Continued on next page

3   Now, imagine you are Participant 1, and Participant 2 has the following characteristics: {Gender: [Male or Female]; Age: [Youth, Middle Aged, or Old]; Mother tongue: [Luo, Kikuyu, Luhya, Kamba]}. You receive KSH 200. You can choose to allocate any of the following amounts to Participant 2: KSH 0, 40, 80, 120, 160 or 200. Imagine you choose to allocate KSH 40 to Participant 2. Remember, this amount is multiplied by three. How much money does Participant 2 receive? [0, 40, **120**, 200, 240, 360, 480].

4   Imagine Participant 2 chooses to return KSH 40 to you. Remember, you still have KSH 160 that you kept in the first part of the task. How much money do you receive in total? [40, 160, **200**, 240, 280].

5   How much money does Participant 2 keep? Remember, Participant 2 received KSH 120 and returned KSH 40. [40, **80**, 120, 160, 200]

F1 The amount of money allocated by Participant 1 for Participant 2 is multiplied by three. [**True**, False]

F2 In the task, Participant 1 chooses how much of 200 KSH to allocate to Participant 2. This amount is multiplied by three. Participant 2 decides how much of this increased amount to send to Participant 1. Whose role will you play in this task? [**Participant 1**, Participant 2]

Trust Game – Stage 2

---

1   In this task, how much does Participant 1 receive to allocate between himself/herself and you? [0, 80, 160, **200**, 240, 400]

2   If you receive 240 KSH from Participant 1, and return 120 KSH to Participant 1, how much do you keep? [0, 80, **120**, 200, 240]

3   Imagine you chose to return 40 KSH to Participant 1, after you received 120 KSH. How much would you keep in total? [0, 40, **80**, 100, 120]

4   In this task, Participant 1 chooses how much of 200 KSH to allocate to Participant 2. This amount is multiplied by three. Participant 2 decides how much of this increased amount to return to Participant 1. Whose role will you play in this task? [Participant 1, **Participant 2**].

F1 In the task, Participant 1 chooses how much of 200 KSH to allocate to Participant 2. This amount is multiplied by three. Participant 2 decides how much of this increased amount to send to Participant 1. Whose role will you play in this task? [Participant 1, **Participant 2**]

Notes: This table lists the comprehension questions by game. Questions without an F are asked during the instruction phase and questions with an F are final questions right before the games themselves.

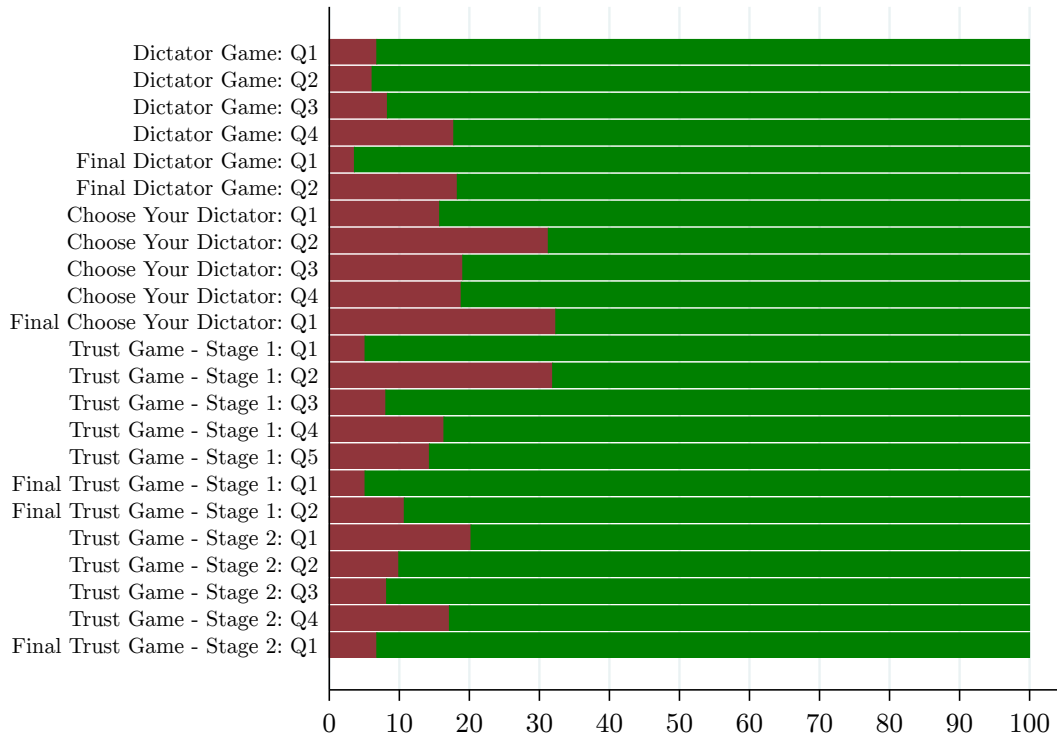Figure C1: Share of correct (green) and incorrect (red) comprehension questions on first attempt by question
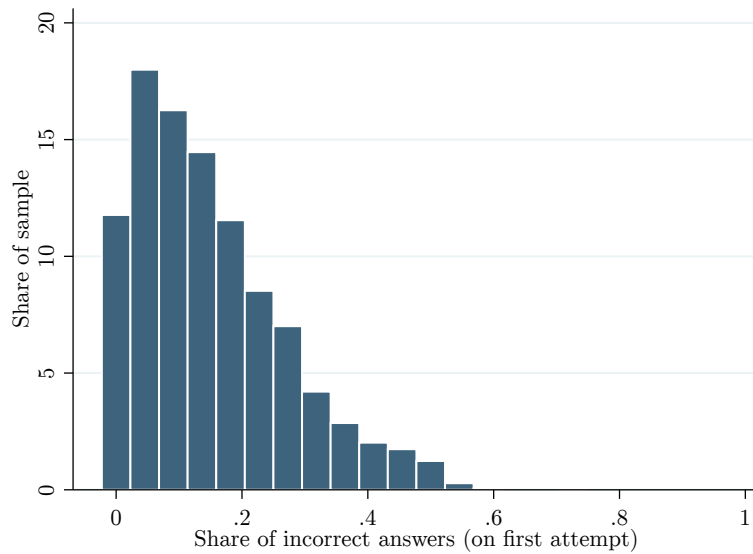

Figure C2: Incorrect comprehension questions on first attempt by participant

C.5. Response times

Tables C8, C9, and C10 display our results when dropping the 20% fastest decisions in terms of profile view time, time until initial decision, and time until leaving the decision screen, respectively.

Table C7: Estimates restricting sample to good comprehension

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5725	0.3554	0.3969	0.3771	0.5479
Sample standard deviation	(0.4948)	(0.2458)	(0.2662)	(0.2449)	(0.2572)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		−0.0138	−0.0238**	−0.0121	0.0150
		(0.0094)	(0.0094)	(0.0096)	(0.0099)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0704***	0.0198***	0.0310***	0.0025	0.0826***
	(0.0075)	(0.0039)	(0.0044)	(0.0023)	(0.0052)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0246*	0.0021	−0.0164*	−0.0018	−0.0050
	(0.0145)	(0.0076)	(0.0085)	(0.0042)	(0.0102)
Participants	1,508	1,699	1,738	1,704	1,770
Decisions per participant	6	6	6	30	4
Decisions	9,048	10,194	10,428	51,120	7,080

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C8: Dropping fastest decisions: Profile view

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5793	0.3605	0.3957	0.3802	0.5508
Sample standard deviation	(0.4937)	(0.2459)	(0.2639)	(0.2466)	(0.2525)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		−0.0128	−0.0238**	−0.0093	0.0088
		(0.0097)	(0.0098)	(0.0098)	(0.0100)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0796***	0.0200***	0.0283***	−0.0006	0.0836***
	(0.0076)	(0.0045)	(0.0051)	(0.0031)	(0.0063)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0266*	0.0075	−0.0217**	−0.0062	−0.0127
	(0.0148)	(0.0088)	(0.0099)	(0.0058)	(0.0123)
Participants	1,772	1,758	1,759	1,778	1,784
Decisions per participant	5	5	5	24	3
Decisions	8,564	8,564	8,564	42,820	5,709

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C9: Dropping fastest decisions: Initial decision

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5768	0.3653	0.4024	0.3945	0.5445
Sample standard deviation	(0.4941)	(0.2438)	(0.2612)	(0.2391)	(0.2499)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			−0.0165*	−0.0266***	−0.0095	0.0088
			(0.0092)	(0.0092)	(0.0090)	(0.0095)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0758***	0.0179***	0.0257***	−0.0001	0.0775***
	(0.0077)	(0.0047)	(0.0051)	(0.0026)	(0.0065)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0194	0.0107	−0.0165*	−0.0030	−0.0106
	(0.0150)	(0.0091)	(0.0099)	(0.0050)	(0.0126)
Participants	1,775	1,775	1,783	1,783	1,783
Decisions per participant	5	5	5	24	3
Decisions	8,564	8,564	8,564	42,820	5,777

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C10: Dropping fastest decisions: Final decision

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5720	0.3634	0.3993	0.3910	0.5468
Sample standard deviation	(0.4949)	(0.2427)	(0.2630)	(0.2409)	(0.2503)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			−0.0117	−0.0249***	−0.0069	0.0080
			(0.0094)	(0.0096)	(0.0092)	(0.0095)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0719***	0.0145***	0.0268***	0.0000	0.0818***
	(0.0077)	(0.0045)	(0.0052)	(0.0025)	(0.0061)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0213	0.0092	−0.0206**	−0.0035	−0.0110
	(0.0150)	(0.0088)	(0.0100)	(0.0048)	(0.0120)
Participants	1,778	1,771	1,777	1,784	1,783
Decisions per participant	5	5	5	24	3
Decisions	8,564	8,564	8,564	42,820	6,124

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.6. Parents of same mother tongue

We restrict the sample to participants whose mother tongue matches that of both their parents. Of the 1784 participants, 133 do not share their mother tongue with both their parents.

Table C11: Parental mother tongue

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5734	0.3561	0.3960	0.3771	0.5484
Sample standard deviation	(0.4946)	(0.2474)	(0.2656)	(0.2477)	(0.2603)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			−0.0183*	−0.0206**	−0.0104	0.0098
			(0.0096)	(0.0096)	(0.0098)	(0.0104)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0713***	0.0164***	0.0298***	0.0010	0.0845***
	(0.0072)	(0.0040)	(0.0045)	(0.0023)	(0.0054)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0147	0.0004	−0.0140	−0.0040	−0.0090
	(0.0140)	(0.0077)	(0.0087)	(0.0044)	(0.0106)
Participants	1,651	1,651	1,651	1,651	1,651
Decisions per participant	6	6	6	30	4
Decisions	9,906	9,906	9,906	49,530	6,604

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.7. Influential outliers

We identify influential observations using Welsch Distance, which is computed for an observation i as the residual r_i of an observation, scaled by its leverage score h_i and the number of observations n as follows: $W_i = r_i \frac{\sqrt{h_i(n-1)}}{1-h_i}$. The cutoff is $3\sqrt{k}$, where k denotes the number of regressors (Belsley, Kuh and Welsch, 1980). For the five games, we thereby exclude 34, 40, 21, 111, and 12 observations, respectively.

Table C12: Estimates excluding influential outliers under the Welsch distance cutoff

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5743	0.3569	0.3973	0.3792	0.5483
Sample standard deviation	(0.4945)	(0.2465)	(0.2660)	(0.2464)	(0.2570)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect		−0.0157*	−0.0223**	−0.0115	0.0129
		(0.0092)	(0.0093)	(0.0094)	(0.0099)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0726***	0.0195***	0.0286***	0.0011	0.0817***
	(0.0069)	(0.0038)	(0.0043)	(0.0022)	(0.0052)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0253*	0.0013	−0.0180**	−0.0019	−0.0035
	(0.0134)	(0.0073)	(0.0084)	(0.0042)	(0.0102)
Participants	1,784	1,783	1,784	1,784	1,784
Decisions per participant	6	6	6	30	4
Decisions	10,686	10,664	10,683	53,409	7,124

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.8. Heterogeneity by education

To account for effects of education, we split the sample by whether or not the participant reported having reached secondary education.

Table C13: Participants without any secondary education

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5722	0.3453	0.3873	0.3693	0.5422
Sample standard deviation	(0.4948)	(0.2421)	(0.2660)	(0.2482)	(0.2617)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			-0.0196^*	-0.0316^{***}	-0.0037	0.0125
			(0.0114)	(0.0118)	(0.0122)	(0.0126)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0710^{***}	0.0191^{***}	0.0279^{***}	0.0019	0.0836^{***}
	(0.0087)	(0.0047)	(0.0053)	(0.0027)	(0.0067)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	0.0372^{**}	0.0021	-0.0254^{**}	0.0001	-0.0101
	(0.0169)	(0.0091)	(0.0101)	(0.0052)	(0.0131)
Participants	1,122	1,122	1,122	1,122	1,122
Decisions per participant	6	6	6	30	4
Decisions	6,732	6,732	6,732	33,660	4,488

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C14: Participants with some secondary education or more

	(1) Choose Your Dictator Game	(2) Dictator Game	(3) Trust Game 1	(4) Trust Game 2	(5) Social Proximity
Sample mean	0.5770	0.3825	0.4173	0.3993	0.5580
Sample standard deviation	(0.4942)	(0.2586)	(0.2676)	(0.2456)	(0.2501)
Panel A: Average hydrocortisone effect[a]					
Hydrocortisone effect			-0.0116	-0.0086	-0.0236	0.0092
			(0.0160)	(0.0152)	(0.0150)	(0.0160)
Panel B: Average coethnicity effect[b]					
Coethnicity effect	0.0748^{***}	0.0185^{***}	0.0325^{***}	0.0012	0.0790^{***}
	(0.0114)	(0.0066)	(0.0076)	(0.0037)	(0.0082)
Panel C: Interaction of hydrocortisone and coethnicity[c]					
Interaction effect	-0.0031	0.0014	-0.0051	-0.0052	0.0045
	(0.0223)	(0.0127)	(0.0149)	(0.0070)	(0.0161)
Participants	662	662	662	662	662
Decisions per participant	6	6	6	30	4
Decisions	3,972	3,972	3,972	19,860	2,648

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.9. First encountered player of same or different ethnicity

To account for order effects, we restrict our sample by whether the first player encountered in each game was of the same ethnicity as the participant.

Table C15: Participants for which the first match of the game is with a non-coethnic

	(1) Dictator Game	(2) Trust Game 1	(3) Trust Game 2	(4) Social Proximity
Sample mean	0.3600	0.3953	0.3775	0.5537
Sample standard deviation	(0.2475)	(0.2644)	(0.2482)	(0.2582)
Panel A: Average hydrocortisone effect[a]				
Hydrocortisone effect	−0.0148	−0.0256**	−0.0104	0.0213*
	(0.0106)	(0.0105)	(0.0110)	(0.0113)
Panel B: Average coethnicity effect[b]				
Coethnicity effect	0.0244***	0.0350***	−0.0003	0.0952***
	(0.0045)	(0.0051)	(0.0025)	(0.0061)
Panel C: Interaction of hydrocortisone and coethnicity[c]				
Interaction effect	−0.0009	−0.0140	−0.0029	−0.0010
	(0.0086)	(0.0099)	(0.0048)	(0.0119)
Participants	1,355	1,327	1,346	1,341
Decisions per participant	6	6	30	4
Decisions	8,130	7,962	40,380	5,364

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C16: Participants for which the first match of the game is with a coethnic

	(1) Dictator Game	(2) Trust Game 1	(3) Trust Game 2	(4) Social Proximity
Sample mean	0.3564	0.4075	0.3894	0.5311
Sample standard deviation	(0.2535)	(0.2743)	(0.2458)	(0.2549)
Panel A: Average hydrocortisone effect[a]				
Hydrocortisone effect	−0.0199	−0.0150	−0.0098	−0.0145
	(0.0193)	(0.0198)	(0.0188)	(0.0204)
Panel B: Average coethnicity effect[b]				
Coethnicity effect	0.0019	0.0164**	0.0067	0.0395***
	(0.0075)	(0.0080)	(0.0045)	(0.0099)
Panel C: Interaction of hydrocortisone and coethnicity[c]				
Interaction effect	0.0093	−0.0291*	0.0010	−0.0213
	(0.0143)	(0.0153)	(0.0085)	(0.0190)
Participants	429	457	438	443
Decisions per participant	6	6	30	4
Decisions	2,574	2,742	13,140	1,772

Notes: [a] The average hydrocortisone effect is calculated as in equation (2). [b] The average coethnicity effect is calculated as in equation (3). [c] The interaction effect is given by $\hat{\beta}_3$ in equation (1). 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. Significance is denoted by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D. Summary of safety protocols

Despite a large literature on the role of ethnic identity in social interactions, there remains significant equipoise when it comes to the effect of coethnicity, in particular when paired with stress, as we detail in the introduction. We do not foresee a plausible risk that the findings of the study may be misused by any interested party. This study underwent thorough ethical review by Princeton and KEMRI IRB. The IRBs of Harvard, Northwestern, Bocconi, and the NBER ceded ongoing IRB review to Princeton after initial approval. These IRBs covered the PIs as per their affiliation at the time of data collection. The use of hydrocortisone was approved by the Kenyan Pharmacy and Poisons Board. The dosage of 20mg is very low. It is not uncommon in hospital settings where hydrocortisone is used to abate allergic reactions to administer dosages on the order of 500mg or 1000mg. In order to minimize the risk of side effects, the study sample was restricted to participants aged 18-40 who were in good health (as assessed by a team of trained nurses using a detailed health screening that can be found below and that was developed with the IRB, as well as vital signs). The nurses were all fluent and medically conversant to have patient conversations in both English and Swahili. Pregnant or breastfeeding women were excluded from the study. The determination was based on self-report and a number of probing questions in a private setting between a female nurse and the participant. Participants were informed of the potential side effects of hydrocortisone and a nurse was present at all times during the study in case of adverse events. They were given a prescription for 20mg hydrocortisone regardless of their treatment status should they need it for their records or later medical purposes. On it was indicated a phone number they could call in case they developed adverse side effects after study completion. The phone was kept by one of the nurses at any time during the study and for one month after the completion of data collection. An IRB-approved adverse event protocol was in place (see below) and the study team made arrangements with one of the local hospitals that would receive study participants and bill the costs of any tests to the study budget that were needed in order to determine whether symptoms might be linked to hydrocortisone administration. If a link to hydrocortisone could not be excluded, the study would bear the cost of treatment. In order to protect participants' privacy, no other medical diagnosis was shared by the hospital, only symptoms and the possibility of a link to study participation. Adverse events related to the study (mainly dizziness and nausea) were extremely rare, in line with expected occurrence according to the medication's leaflet, and reported to the IRBs immediately. There were no serious adverse events. The study design did not expose staff or non-participants to any discernible harm. The researchers declare no financial conflict of interest. One of the researchers, Johannes Haushofer, holds a non-remunerated advisory position at the Busara Center for Behavioral Economics and another, Moritz Poll, was employed at the Busara Center and coordinated study implementation on the ground. Neither of them, nor the Busara Center stand to gain directly from the study findings.

Consent was obtained in Swahili and participants could choose between consent forms in English, Swahili or both. There was ample room for questions and participants were informed that they could withdraw from the study at any point without having to state a reason and without any repercussions. Participants withdrawing from the study would receive their full show-up and

transportation fee, as well as any experimental payouts they had earned up to that point. In practice, all study withdrawals happened around the time of consenting. If participants did not consent to participating, withdrew shortly after, or were ineligible due to the medical screening or occasionally due to capacity constraints, they were still paid their full show-up and transportation fee. The determination of whether a participant who did not pass the medical screening would remain part of the participant pool for potential later reinvitation was at the discretion of the nurses and guided by whether the disqualifying characteristic was likely permanent (pregnancy, chronic disease, etc.) or likely to be resolved on another day (vital signs slightly out of range, use of alcohol in the past 24 hours etc.). In case of capacity constraints, participants to be sent away were selected at random, so not to induce sample selection on punctuality, and were offered to reschedule.

Appendix E. Comparison to related studies

Table E1: Comparison to other studies in this literature

	This study	Berge et al. (2020)	Blum et al. (2021)	Habyarimana et al. (2007)
Location	Kibera, Kawangware, and Viwandani (Nairobi, Kenya)	Kibera and Viwandani (Nairobi, Kenya)	Kibera and Kawangware (Nairobi, Kenya)	Kawempe (Kampala, Uganda)
Games played	CYD, DG, TG, Social Proximity	CYD, DG, PGG, IAT	CYD, DG, PGG, AMP, FAAT, WMT	DG, Puzzle Game, Network Game, Prisoners' Dilemma
Sample size	1,784	1,362	182	300
Observations per participant	60 (6 CYD, 6 DG, 36 TG, 12 Social Proximity)	10 (2 CYD, 3 DG, 3 PGG, 2 IAT)	9 game rounds (2 CYD, 3 DG, 4 PGG) + 170 attribution tasks (90 AMP, 40 FAAT, 40 WMT)	24 (12 DG, 3 Puzzle Game, 1 Network Game, 8 Prisoners' Dilemma)
Ethnic groups	Luo, Kikuyu, Luhya, Kamba	Luo (pooled with Luhya and Kisii), Kikuyu	Luo, Kikuyu	10+ different groups
Information	Profiles (age group, gender, mother tongue)	Profiles (education, age, hometown) and No information	Profiles (photo, age, education, hometown) and No information	Profiles (photo) and No information
Proxy for ethnicity	Mother tongue	Hometown	Hometown	Self-reported; perceived coethnicity from photos
Audio profiles	Audio	Audio	Audio	No[a]
Coethnicity effects found in at least one version of the game	CYD, DG, TG, Social Proximity	CYD, IAT	CYD (profiled version only), AMP, WMT	DG (profiled version only), Network Game, Prisoners' Dilemma

Notes: Information is based on the main specifications reported in each paper. DG is the dictator game. In Habyarimana et al. (2007), there are two recipients in each round. CYD is the choose your dictator game. PGG is the public goods game. TG is the trust game. IAT is the implicit association task. AMP is the Affect Misattribution Procedure. FAAT is the Face Anger Attribution Task. WMT is the Weapon Misidentification Task. Profiled version means they were given information on the other player. [a] Habyarimana et al. (2009) report additional rounds in which a small part of the sample is introduced to the player they play with by means of a video that features them speaking either the lingua franca or their tribal language.

References

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

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel N. Posner, Bertil Tungodden, and Kelly Zhang, "Ethnically Biased? Experimental Evidence from Kenya," *Journal of the European Economic Association*, 2020, *18* (1), 134–164.

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

Habyarimana, James, Macartan Humphreys, Daniel N. Posner, and Jeremy M. Weinstein, "Why Does Ethnic Diversity Undermine Public Goods Provision?," *American Political Science Review*, 2007, *101* (4), 709–725.

__ , __ , __ , **and** __ , *Coethnicity: Diversity and the Dilemmas of Collective Action* 2009.

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

List, John A., Azeem M. Shaikh, and Yang Xu, "Multiple Hypothesis Testing in Experimental Economics," *Experimental Economics*, 2019, *22* (4), 773–793.