“White Man’s Burden”? Experimental Evidence on Generosity and Foreigner Presence*

Jacobus Cilliers† Oeindrila Dube‡ Bilal Siddiqi§

October 29, 2014

Abstract

We experimentally vary white foreigner presence in dictator games across 60 villages in Sierra Leone, and find that the simple presence of a silent white foreigner increases player contributions by 19 percent. To separate the impact of the white foreigner’s race and nationality from other characteristics, we test additional predictions. First, the white foreigner’s presence may heighten demand effects, prompting players to impress the white foreigner by being more generous. This should make behavior in the game less indicative of true generosity. Consistent with this, we find that game contributions are no longer predicted by real-world public good contributions when the white foreigner is present. Second, the white foreigner’s presence may make those more familiar with aid perceive the games as a form of means testing, and therefore give less to signal that they are poor. Consistent with this, in the presence of the white foreigner, players in more aid-exposed villages give less and are more likely to believe that the games are testing them for aid suitability. Together, these results suggest that players’ giving decisions respond to the white foreigner’s race and nationality. Our findings hold direct implications for how to measure altruism, target aid, and evaluate aid effectiveness in the developing world.

Keywords: Aid, race, altruism, behavior, dictator games, lab-in-the-field experiment, Africa

JEL Classification Numbers: C91, C93, O19

*We are grateful to Sendhil Mullainathan for numerous discussions and suggestions. We also thank Abigail Barr, Erwin Bulte, Jessica Gottlieb, Rachid Laajaj, Ted Miguel, Anand Rajaram, and Debraj Ray, as well as seminar participants at Stanford, Oxford, NYU, Wageningen, the Princeton Conference on Causal Inference and the Study of Conflict and State Building, the Northeast Universities Development Consortium (NEUDC), the Berkeley Symposium on Economic Experiments in Developing Countries (SEEDEC), and the Centre for the Study of African Economies’ Annual Conference for comments. Ali Ahmed, Jessica Creighton, Anthony Mansaray, Joshua McCann and Jeffrey Steinberg provided excellent research assistance. The International Initiative for Impact Evaluation (3ie) provided funding for the research. Siddiqi acknowledges support from the AFOSR under Award No. FA9550-09-1-0314. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of any supporting institution. All errors and omissions are our own.

†University of Oxford. jacobus.cilliers@economics.ox.ac.uk

‡New York University. odube@nyu.edu

§World Bank. bsiddiqi@worldbank.org

A vast number of interactions in the developing world involve foreigners disbursing resources to locals, who differ starkly—and often visibly—in terms of their ethnic identity. The extent of such interaction has also risen dramatically in recent years, with the emergence of two trends. First, aid is increasingly disbursed through participatory approaches, which entail greater interaction between donors and recipients in the field.[1] Second, there has been a rapid increase in the use of field experiments and behavioral games in developing country contexts (Figure 1). These two trends have overlapped through greater program evaluation of development assistance projects, which often employ behavioral experiments (Avdeenko and Gilligan, 2012; Casey et al., 2012; Fearon et al., 2009; Humphreys et al., 2012).

The results from such experiments are interpreted as unbiased metrics of human behavior. Yet, in both the experimental and aid contexts, differences in race and nationality strongly signal differences in wealth, power, and authority. Understanding the extent to which these differences bias individual behavior is important, given the growing use of behavioral techniques, and the implications of such bias for aid targeting and evaluation.

Our paper directly examines this issue. We study the effect of researcher race and nationality on individual behavior in a developing country. Specifically, we conduct a lab-in-the-field experiment that randomizes the presence of a white foreigner across behavioral games in 60 villages across Sierra Leone. We assess how the presence of this "white man"[2] impacts generosity as measured by player contributions in dictator games.[3] We find that the mere presence of the white foreigner on the research team increases the amount given by a substantial 19 percent.[4] A challenge in interpreting this effect as arising from the presence of a *white* person lies in the fact that the experiment varied the presence of two individuals—one white and one black—across the games. Thus, players may have responded to some other characteristic of the white person, besides his identity as a white foreigner. For example, if he was perceived to be especially friendly, this personality trait may have increased giving. Alternatively, if he seemed more educated, this may have commanded greater respect and influenced player allocations. In short, many channels are possible.

We deal with this challenge in two ways. The first is by design. The behavioral games were implemented by teams of five individuals, four of whom were black Sierra Leoneans, including the team leader. The random variation is in the identity of the fifth member: in control areas,

[1] For example, over the last decade, the World Bank has allocated \$85 billion in participatory development assistance (Mansuri and Rao, 2013).

[2] Non-Africans in Sierra Leone are typically referred to in English and Krio as "white man," and in other dominant local languages as its literal equivalent: "opoto" in Temne, "poomui" in Mende, etc. Even though Lebanese, Chinese, or Indians may sometimes be identified as such when they are compared to one another, they are generally referred to as "white" when discussed in comparison to Sierra Leoneans or Africans. Thus we refer to our treatment as the "white man" effect. Since our study also explores the role of aid expectations and the disbursement of aid by white foreigners, our title "White Man's Burden" is an homage to the 2006 book of the same name by William Easterly.

[3] Player contribution in a dictator game is the common measure of altruism in behavioral experiments (Cardenas and Carpenter, 2008).

[4] Our paper relates to studies in the social psychology literature that have examined the impact of surveyor and experimenter characteristics, including race, on relevant outcomes—see Rosenthal (1963) for a summary. Others have also examined how researcher characteristics including race, education, income, gender, and religion influence subject response in survey data (Anderson et al., 1988; Bailar et al., 1977; Blaydes and Gillum, 2012; Cotter et al., 1982; Finkel et al., 1991; Hyman, 1954; Mensch and Kandel, 1988; Miyazaki and Taylor, 2007; Reese et al., 1986; Webster, 1996).

he, too, was a black Sierra Leonean, but in treatment areas, the fifth member was a white foreigner. The team leader—who was the same person across all games—played an active role and instructed the players. In contrast, the protocol specified a silent, limited role for the fifth member. He was not allowed to talk to the players in either treatment or control areas. His specific task was to hand out money at the start of the game, and he was not to interact otherwise with game participants. In this regard, our experiment examines how the mere *presence* of a white foreigner affects measured generosity in the games.

This design only partly mitigates concerns that other characteristics could drive the observed effects. To further address this issue, we lay out explicit hypotheses about why the white foreigner's presence could plausibly influence player contributions, and generate additional, testable predictions implied by the hypotheses.[5] We test these predictions by drawing on rich household and village-level survey data collected from game participants, which were collected as part of a baseline for a separate study on postconflict reconciliation.

Our key hypothesis is that the result can be explained by "experimenter demand effects," which arise when research subjects change behavior to conform to the perceived desire of the researcher (Levitt et al., 2011; Masling, 1966; Nichols and Maner, 2008).[6] We posit two types of such demand effects. Each implies opposite effects on giving. First, players may give *more* if they perceive that the researcher wants them to display more generosity and they have a stronger desire to impress a white foreign researcher. Demand effects have been shown to arise from perceived power differentials (Schultz, 1969), and in our setting, there are stark disparities in wealth and authority between black locals and white foreigners.[7] Second, players may perceive that the white foreigner is there to assess whether they need aid. As a result, they may act to seem aid-worthy. If subjects believe aid is targeted to the poor, they will have an incentive to act poor by giving *less* in the games where he is present. This second phenomenon is more likely to arise in communities that have more experience with aid, where people have had a chance to learn about means testing and aid-targeting methods. Thus, these two forces—the desire to impress the white foreigner by displaying generosity, and the desire to act aid-worthy—imply countervailing effects on giving.

We generate two key empirical predictions implied by these two types of demand effects. First, if people assume that the "white man" wants them to give more, his presence should erode the relationship between giving in the games and giving in the real world. Past work suggests that the desire to look good in the lab setting can induce people to act more altruistically than they would outside this setting (Benz and Meier, 2008; Levitt and List, 2007; List, 2005), and more so when their actions are known to the researcher (Andreoni and Bernheim, 2009; Hoffman et al., 1996; List et al., 2004). Taken together, these studies suggest that larger experimenter demand effects in presence of the "white man" will reduce the correlation between player's game

[5] By focusing on the experimenter's ethnicity, our paper differs from previous studies that have examined the import of players' racial, ethnic and national identities in determining behavioral game allocations (e.g. Adida et al., 2012; Burns, 2006, 2012; Cappelen et al., 2011; Habyarimana et al., 2007; Whitt and Wilson, 2007).

[6] Experimenter demand effects are related to the idea of "Hawthorne effects"—the finding that participants alter their behavior in experimental settings because they are being observed (Adair et al., 1989; Diaper, 1990; Haley and Fessler, 2005; Hoffman et al., 1994; Levitt et al., 2011; McCarney et al., 2007; Zwane et al., 2011).

[7] In the wake of Sierra Leone's long civil war, foreigners affiliated with the United Nations provided peacekeeping and sponsored the special courts that prosecuted war crimes. Foreign aid workers continue to dispense resources and restructure governance, while foreign businessmen trade in diamonds.

allocations and their real-world behavior.

We use our baseline surveys to measure how much players contributed to public goods in their villages prior to the games. We find that these contributions strongly predict contributions in the games where the "white man" is absent, but do not where he is present. This pattern is consistent with stronger demand effects eroding the relationship between behavior and measured generosity in the "white man" group.

Second, if players give based on perceptions that the "white man" is testing them for aid, they should give less in his presence, particularly in villages that have more experience with development aid. We present two pieces of evidence that support this mechanism. We find that the "white man" treatment effect falls in villages that have had longer exposure to non-governmental organizations (NGOs) dispensing aid. In fact, players in the top 20 percent most aid-exposed villages give *less* in the presence of the white foreigner. We also draw on a postgame question that asked players their perception regarding the purpose of the game. We find that players in these highly aid-exposed villages were also more likely to believe that the games conducted in the presence of the "white man" were to test them for aid, rather than to give out money or conduct research.

These results suggest that players act based on their perceptions of the experimenter as a white foreigner. Moreover, they help rule out other potential channels. For example, if other characteristics such as his demeanor drove the results, we should not observe differential impacts in either giving or aid testing beliefs based on factors such as aid exposure.

We also conduct a third empirical test. Traditional village leaders are typically the intermediaries between aid givers and the community; thus if players give based on perceptions of aid testing, then leaders should also give less, as they are more likely to be aware of how aid is given out. Consistent with this account, we find that players from households of traditional leaders also give less in the presence of the "white man," and are more inclined to believe that the purpose of the games where he is present is to test them for aid. This suggests a desire to signal the need for aid on their part. However, since these are also higher-status individuals, they may also give relatively less in treated areas because they do not perceive a larger power difference between themselves and the white foreigner. Thus both forces—the desire to signal need and smaller demand effects among powerful people—may contribute to the net lower giving by leadership households in response to the "white man."

While these effects above provide evidence that individuals make their decisions in response to the white foreigner's race and nationality, we also formulate a fourth empirical prediction to help rule out the primacy of other characteristics. If giving responded to another personality trait correlated with the white foreigner, we would plausibly expect the treatment effect to vary based on this trait among players. However, we find no differential impacts based on characteristics such as player friendliness proxied by their friendships, or psychometric measures of depression and PTSD; nor do we find differential effects based on gender and education. This casts doubt on the importance of characteristics such as demeanor in inducing differential giving in the presence of the white foreigner.

In the sections below, we provide institutional context, describe the empirical strategy and data, present our results on differential giving and its underlying mechanisms, and conclude by discussing the implications of our findings for research and policy design.

1 Institutional Context

As our study explores the role of differences in power and aid exposure in moderating subject responses to foreigner presence, this section provides a brief overview of Sierra Leone's extensive exposure to foreign aid, and its traditional authority system, including the strong local power base of chiefs, religious leaders, and secret societies.

Sierra Leone experienced a dramatic influx of foreign aid following its long and brutal civil war, which lasted from 1991 to 2001. The war contributed to the escalation of poverty, leaving the nation among the ranks of the world's poorest and least developed—in 2011, the country ranked 180th out of 187 countries on the United Nation's Human Development Index, up from 187th out of 187 in 2003. With the end of the conflict, the country experienced a surge in foreign aid (Fanthorpe, 2003). Official development assistance as a proportion of gross national income peaked at 43 percent in 2001 and remained over 24 percent in 2010, making it the 11th most aid-dependent nation in the world (World Bank, 2012). However, Sierra Leone's exposure to foreign aid dates back far longer due to religions missions, the earliest of which were established in the beginning of the 17th century (Fyfe, 1968). Given this extended dependence, exposure to development aid is highly likely to shape locals' perception of white foreigners.

Sierra Leone has a strong chieftaincy system, a key institutional legacy of British indirect rule. With the creation of the protectorate in 1898, the Paramount Chiefs—the highest-level traditional chiefs—were given substantial powers over their subjects (Acemoglu et al., 2012; Fanthorpe, 2001; Jackson, 2006). Despite postwar attempts to create more accountable, formal sources of administrative authority, de facto power still remains with the chiefs: "chiefs are regarded as legitimate traditional rulers, with broader political power bases than the council chairperson" (Jackson, 2006). They are still called upon to settle disputes (Sawyer, 2008), and they retain the power to tax, request labor, administer justice, and allocate land.

Informal power in Sierra Leone also resides in secret societies, cultural institutions or "cult associations" (Richards, 2005) that regulate social behavior through secret rituals and practices (Fanthorpe, 2007). Most Sierra Leoneans are members of secret societies, whose influence permeates all aspects of social and political life in the country (Murphy, 1980). Traditional and missionary religions overlap with these institutions to some degree. Most of the population identifies as either Muslim or Christian, though the practice of these religions is intertwined with traditional religious beliefs (Conteh, 2009). Religious organizations such as Bible and Koranic study groups play an important role in community life. Local religions imbue higher powers with chief-like authority—the word for "god" in Limba means, literally, "The Great Chief." The chief's powers are also intricately linked to the workings of the secret societies (Jackson, 2006). Given its extensive exposure to foreign aid, Sierra Leone provides an ideal setting to explore the questions that motivate this study—how foreigner presence effects measured generosity, and how this effect is moderated by aid expectations.

2 Experimental Design

To examine the impact of white foreigner presence on measured generosity, we randomized the presence of a white "supervisor" in the administration of behavioral games. The experiment was carried out in 60 villages across five districts in Sierra Leone: Bombali, Kailahun, Koinadugu, Kono, and Moyamba. Randomization was stratified by district: 12 villages were selected in each district and randomized into treatment and control. Thus there are six treatment and six control villages in each district. Figure 2 shows these treatment and control locations.

2.1 Baseline Survey

The subjects in our study were recruited in the following manner. We conducted baseline surveys of households and villages, on average, 1.3 months prior to when the games were conducted. These surveys were part of a broader study that examines postconflict reconciliation in Sierra Leone, though the survey was modified for the purposes of the "white man" experiment. For the surveys, we randomly sampled 12 respondents within each village. First, households were sampled using in-field randomization within the village population. Within each sampled household, the respondent was also randomly selected from the roster of household members. These 12 individuals typically served as the players in the games. However, if the original survey respondent could not be found, he or she was replaced by someone else from the same household, and of the same gender as the original respondent. Replacement occurred for 7.5 percent of the sample, and we do not have baseline covariates on these individuals.

2.2 Game Sessions

All games were administered by a core team of black Sierra Leoneans comprising one trained facilitator, two assistants, and a local translator recruited from the area where the games were played. Aside from the translator, the three other core members remained constant across all games. Thus all the games were run by the same facilitator, who read out instructions to the player, and interacted extensively with them. The assistants played relatively minor roles, such as preparing the envelopes in which players placed their allocations and remaining in the room to ensure that players did not speak to one another.

In addition, the team included a "supervisor" whose identity varied systematically across control and treatment villages. In the control villages, the supervisor was a Sierra Leonean from Freetown, the national capital. In the treatment villages, the supervisor was a white American. A clear concern is that some other unobserved characteristic of the two individuals may have driven participants' responses. Our design took several precautions against this. First, neither "supervisor" was made aware of the experiment being conducted in order to minimize the chance of expectancy effects (Rosenthal and Rubin, 1978). At the same time, to prevent other potential differences, such as proactivity, from confounding the "white man" treatment, all games were conducted according to a strict protocol and set script that specified a silent, background role for the supervisors. Their sole tasks were to take notes and distribute money to players in the beginning of each game. They did not speak during the administration of the game or otherwise interact with game participants in any way. They were also not allowed to be in the room when players made their allocation decisions. In other words, though they were called supervisors, their actual role was extremely limited. Thus, our experiment examines how the mere presence of a white foreigner affects giving in these games.

Second, both supervisors were chosen for similarity on key characteristics: they were both college-educated males in their 20s. Importantly, Sierra Leone's rural population is homogeneously black African and it is therefore reasonable that the starkly visible difference in the race of the white supervisor is the most salient difference between them. Though it is not possible for our design to disaggregate the impact of race from that of nationality, in research and aid settings in developing countries this tends to be a common bundle—in our setting, it would be rare, for example, to find a "white local."

However, the design of the experiment cannot fully rule out the possibility that other characteristics besides race and nationality could give rise to observed differences in giving across the two games. Therefore, we also formulate several additional empirical predictions that should hold if players act based on their perceptions of the researcher as a white foreigner. We draw on these additional tests to provide evidence of the "white man" mechanism.

Our measure of generosity aggregates giving in three versions of the standard dictator game. The advantage of the dictator game is that it provides a clean measure of other-regarding preferences, since payoffs to play are independent of beliefs over other players' expected behavior.

In each of the specific games, the player was given 4,000 leones (approximately $1), which is an amount slightly higher than the average daily income in Sierra Leone. The player could decide how much of this money to keep and how much to give to recipients.

The three specific types of the dictator game were played in the following order:

1. Game 1: Anonymous Own-Village. The recipients were 12 additional individuals randomly sampled from the players' village. The players were told who these 12 recipients were, but not which recipient they would be matched to. Players were also informed that they would be re-matched to a different recipient in each game. Player contributions were strictly anonymous, that is, no player was informed how much any other player had chosen to contribute.

2. Game 2: Non-Anonymous Own-Village. Here, we relaxed the between-subject anonymity condition. Instead, players were told that their giving decision would be announced at the end of the session, after all games had been played. All other game elements of Game 1 and 2 were identical. Recipients were the same individuals from Game 1, and as in that game, players were not told who they were matched to. To minimize feedback across games, the announcement was made in front of the players and recipients at the end of the session, after all games had been played. Recipients were not present during the game, but instead arrived after all games were played to receive their allocations.

3. Game 3: Anonymous Other-Village. Players were told that we had identified 12 recipients in "another village" who would receive the amount they chose to give. All other game elements were the same as in Game 1—no player was informed of any other player's contribution, who they were matched to, and so on.

We sum the proceeds from each of these specific games and examine the impact on total giving.[8]

Games were conducted in two rooms: (1) a public room where the group received instructions, and (2) a private room where they individually made their allocation decision. No talking was allowed during the explanation of the games, to prevent strategic interaction and framing. To ensure that players understood the game, the facilitator repeated the explanation of the game in the private room, allowing players to ask questions. After the second explanation, the facilitator would exit the room, allowing the player to make her or his allocation decision privately. The players were instructed to place their allocations in an envelope, and place the envelope into the slit of a locked box.

Finally, directly after the games were conducted, we asked participants to respond to a question about what they believed to be the purpose of the research team's visit. These questions were administered in private to each respondent.

To maximize players' comprehension, the games were also translated into seven local languages in three of the five districts.[9] To promote quality and consistency across translations, the games scripts were translated into the local languages and then back-translated again into English. However, the translation quality may have varied across districts to some degree. In three of districts, there is just one major local language, and thus one translator had to be hired. In our samples in Bombali and Koinadugu districts, there are two and three local languages, respectively, requiring the equivalent number of translators. Since higher translator turnover may have lowered the quality of communication, we verify the robustness of our results to dropping these two districts (table A.5). To account for the possibility that the teams, as a whole, may have improved their communication skills over time, we varied the ordering of treatment and control villages, so that games were conducted first in treatment villages in half the districts, and first in control villages in the other half.

2.3 Empirical Strategy

Based on this experimental design, we use the following specification to estimate the impact of the white foreigner effect:

$$y_{ivld} = \alpha_d + \delta_l + \beta(\textit{white-man}_{vld}) + \mathbf{Z}_{vld}\rho + \mathbf{X}_{ivld}\phi + \omega_{ivld} \tag{1}$$

where i is individual, v is village, l is majority language group, and d is district. y_{ivld} is the amount given in the dictator games, $\textit{white-man}_{vld}$ is an indicator of treatment to assignment, α_d are district fixed effects, and δ_l are majority-language fixed effects. We include district fixed effects to account for substantial regional heterogeneity in economic development, urbanization, war exposure, and other factors that may influence generosity. We include majority-language fixed effects to control for variation in underlying generosity levels across ethnic groups, since language groups correspond closely to ethnic/tribal affiliation in Sierra

[8]Since the order of the games was not randomized, this raises potential concerns that ordering affects how much was contributed in each game. However, we do not aim to identify different effects across games, and instead examine aggregate giving.

[9]The facilitator spoke the local language in the other two districts.

Leone. Majority-language fixed effects also reduce noise created by the varying quality of communication across games. Specifically, they absorb potential variation in the quality of the translator, since there was one translator per language group. $\mathbf{Z}_{vld}$ and $\mathbf{X}_{ivld}$ are additional individual and village-level covariates included in some specifications. β captures the "white man" treatment effect. As the randomization took place at the village level, we cluster the standard errors in all specifications at this level.

3 Data

The behavioral games data used in the analysis is supplemented by household and village-level survey data collected prior to implementation of the games.[10] We have village-level data for all 60 villages, and household-level data for 715 out of 720 participants, though the sample size varies for particular measures.

To account for ethnicity, we construct an indicator variable for each language spoken by the majority of the respondents in each village. Ethnicity is geographically concentrated and identified by language. The language spoken by the majority of villagers therefore provides a measure of the majority ethnic group in the village. There were seven ethnic groups among our sampled villages (Fullah, Kono, Koranko, Loko, Mende, Temne and Yalunka). However, there was only one Yalunka majority village, and thus, for robustness, we constructed two additional measures of ethnicity using broader linguistic parent groups from the third and fourth levels in the Atlantic language hierarchy, details of which are provided in the appendix.

We control for violence-related effects by using a measure of the number of buildings burned in each village during the war. We also measure village size based on the number of households residing in the village, and the level of market integration, with an indicator for whether there is a market in the village. We also use two proxies for village cooperation: an indicator of whether the village has communal farms and an indicator of whether it has a labor gang, which refers to a group of people who work on each others' farms.

To measure each village's exposure to foreign aid, we construct various measures of NGO involvement in the provision of local public goods.[11] The broadest indicator is based on whether an NGO owns a school or clinic, currently provides resources to the local school or clinic, or has contributed to the construction of public facilities (including schools, clinics, wells, latrines, or the palava hut[12]). This variable is called "NGO aid," and 54 villages of our 60 villages have some form of aid engagement under this metric.

Since the NGO aid indicator has little variation, we also construct two continuous measures of the extent of aid exposure. The first variable, "Years of aid with NGO-owned facility," refers to the number of years that an NGO has owned either a school or clinic in the village. It is worth noting that NGOs own these public facilities in 31 (or just over half) the sample villages. We also construct "Years of aid with NGO activity," which is the number of years an NGO has either owned a school or a clinic, or the number of years for which an NGO has contributed to

[10] Data collection was part of a larger research project in Sierra Leone, although we adapted the questionnaire to address the specific research questions of this study.

[11] In the Sierra Leonean context, nearly all large NGOs involved in providing public goods in rural villages are foreign or have foreign funders.

[12] The palava hut is often the only public space in the village for community members to congregate.

the construction of the school or clinic. This variable has the advantage of being comprehensive, capturing NGO activity in five additional villages where they do not own facilities. Therefore, we use it as our primary measure of aid exposure. However, since we cannot observe if the NGO engaged each year between the time it built the facility and its current support, we present robustness checks with the "NGO-owned facility" and "NGO aid" variables in the appendix.

To measure household socioeconomic status, we develop an asset index. As is standard in the literature (Vyas and Kumaranayake, 2006), we use principal component analysis to construct an aggregate measure based on ownership of a broad range of appliances and equipment,[13] the material of the roof and floor of the house in which the respondent lives, as well as ownership of land.[14] Data on the respondent's educational attainment is used to generate a continuous measure of years of education completed.

We are also able to examine the player's past exposure to white foreigners, because the survey tracked how often the respondent has met a white person before. In addition, we asked the respondent if someone in their household is a paramount, town, or village chief/headman; a religious leader, including reverend, imam, or pastor; or the leader of a secret society. We use this to generate a measure of whether the participant is from a household of a traditional authority leader.

To test for the correlation between giving in the games and real-world behavior, we construct two indexes of individuals' contributions to community groups and public goods. The "Monetary contributions index" aggregates the size of monetary contributions to different groups in the village over the past 3 months (religious, women, parent-teachers association, village development committee, and youth group), as well as an indicator whether a respondent has contributed towards building or maintaining public facilities. The broader "All contributions index" additionally includes labor contributions to community groups and road brushing. We use the methodology outlined by Kling et al. (2007) to aggregate these different types of measures.[15]

To gauge differential impacts based on player characteristics, we used household survey measures of psychological health to create a depression index based on five questions drawn from the Zung (1971) index and a measure of post-traumatic stress disorder (PTSD) based on a list of 11 questions from the American Psychiatric Association's 4^{th} Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), as developed by Foa et al. (1997). We also asked each respondent to list which of the other respondents they consider a very good friend, within their village. We use the number of times each player was named a good friend by someone else as a proxy of their friendliness.

In the postgame questions, participants were asked, "Why do you think the researchers were playing these games? Choose the statement you agree with most." The following three statements were then read out to them: (1) "To give money to the community in a fun and educational way," (2) "To test the community, to see which community is more deserving of aid," and (3) "To find out more about how members in this community think about each other,

[13]These include: iron, refrigerator, television, radio, mobile phone, generator, fan, bed, clock, sewing machine, modern stove, bicycle, motorcycle, car/truck, and pushcart.

[14]We cannot use measures such as income or wages to control for wealth since the majority of respondents are subsistence farmers.

[15]The Kling et al. (2007) method standardizes each of the indicators and sums across them after imputing missing values.

interact with each other, and treat each other." The last option corresponds most closely to the stated purpose of the research team's visit prior to the start of the games, which was to find out about the community. In the text and tables, we refer to the first statement as "Give money," the second as "Aid test," and the third as "Research."

3.1 Descriptive Statistics

Table A.1 presents descriptive statistics of key variables. Our main dependent variable is total giving, which sums the proceeds from the three dictator games. The average total giving was 3,216 leones out of 12,000 given to players in all three games, or an average of 1,079 leones (=$0.25) per game, which is about a quarter of the daily wage in Sierra Leone.[16] Table 1 examines key individual and village level characteristics in the control and treatment groups. There are no statistically significant differences in these characteristics across the two groups with the exception of respondent age and whether the village has a communal farm. Though these differences are only significant at the 10 percent level, in table A.4 we show that the results are robust to controlling for these characteristics.

4 Result: "White Man" Presence and Giving

We begin a with a simple graphical exploration of whether the "white man" treatment induced differential giving. Figure 3 shows the cumulative distribution function (CDF) of total giving in treatment and control regions. The treatment distribution first-order stochastically dominates the control distribution: a higher proportion of the treatment group gave more, evaluated at each level of giving, indicating that the "white man" treatment induced an upward shift in the distribution of giving in the dictator games. The Kolmogorov-Smirnov equality of distribution test rejects the null hypothesis that these two distributions are equal.[17] While the CDF presents visualization of the raw data, table 2 presents regression estimates of equation (1). Column (1) presents estimates on total giving without ethnicity fixed effects. The coefficient indicates that players gave 537 leones more in the games where the "white man" was present. Columns (2) adds in ethnicity fixed effects, which increase the precision of the estimates. Here, the coefficient of 564 indicates a 19 percent increase in total giving, relative to the control mean (2,944 leones).

In column (3), we show that this main result is not affected by controlling for replacements, who played the games but were not surveyed at baseline. Column (4) shows that the treatment effect is not significantly different for this group while column (5) demonstrates the robustness of the effect to dropping these 55 individuals. This is an important check because we are not able to include replacements in several of the remaining empirical tests, which require baseline covariates. Thus column (5) represents our preferred specification.[18]

[16] The Anonymous Other-Village game is only available for 708 out of 720 observations owing to enumeration error (the values for this game were not recorded in one of the villages).

[17] The test statistics reject the null that giving in the control group is smaller than giving in the treatment group with a p-value of 0.043, and rejects the null that the samples are drawn from the same distribution with a p-value of 0.079.

[18] Table A.3 examines effects for each specific variant of the dictator game. The effect appears to be most precisely estimated for the Anonymous Own-Village dictator game. However, table A.6 also shows that when we pool

We conduct a number of additional robustness checks. Table A.4 shows that the results are robust to the inclusion of respondent age and communal farm indicator, as well as alternative ethnicity fixed effects, generated on the basis of broader linguistic categories. Table A.5 also demonstrates that the effects are robust to dropping the two districts (Bomabali and Koinadugu) where translation quality may have been lower because larger linguistic diversity required different interpreters to join the teams intermittently.[19]

In the subsections below, we additionally draw on our household and village survey data to explore mechanisms through which white foreigner presence affects measured generosity.

5 Mechanisms

We hypothesize that giving increases in treatment areas because players act according to their perceptions of the white foreigner—they give based on what they perceive he wants them to do, and based on their beliefs about what he is there to do. There are two potential demand effects that imply opposite effects on giving. First, if players perceive that the white foreigner wants them to give more, they may increase their game allocation to please him. Second, if they perceive he is testing them to see if they need aid, they may give less in order to signal need. This latter "need-signaling" effect is more likely in villages that have received aid in the past, as this is where people have had a chance to learn about means testing. In the subsections below, we present two empirical predictions designed to test this hypothesis.

5.1 Giving to Impress the "White Man"

In this subsection, we explore the first potential demand effect. We develop the following implied empirical prediction: if people give more to impress the white foreigner, this should reduce the correlation between real-world contributions and giving in the games.

Why would we expect a weakening of this correlation? In the absence of the "white man," players give more in the games if they have more pro-social preferences. In his presence, however, people may also give out of a desire to impress him. Behavior in the experiment thus becomes a noisier measure of true generosity, as selfish people may now also act generously. This weakens the correlation between their past contributions and giving in the "white man" group.[20]

To test this hypothesis, we devise a raw prediction test. Our aim is to see if real-world contributions predict giving in the games where the "white man" is absent, but not where he is present. To implement this, we draw on several different measures of how people contribute to community groups and public goods in their communities. As shown in table A.1, the most common forms of contributing are giving resources to the construction of public facilities; contributing labor toward road-brushing; and making monetary contributions to religious

the games together, we are not able to discern statistically distinguishable treatment effects across the three specific game types.

[19] Our results are also robust to iteratively dropping each individual village with 12 players, which suggests that outliers do not drive the result. When we reestimate the treatment effect dropping each village, the p-value exceeds 0.10 in only three out of 60 cases, and marginally so, rising to 0.102, 0.106, and 0.11. These three villages are also in different districts, suggesting that there is no systematic bias at the district level.

[20] Even if preferences were additive, so the desire to please the white foreigner were added to their own pro-social preferences, we would still expect a weakening of the correlation based on ceiling effects: there is little room to increase giving in response to the white foreigner among those who already give a lot.

organizations, with 47 percent, 35 percent, and 31 percent contributing in these ways, respectively.[21] Individually, each of these variables is a noisy measure of people's willingness to give. We therefore aggregate them together into the "Monetary contributions index," and "All contributions index" described earlier. In creating these indexes, we drop one outlying observation, who reported giving 600,000 leones to religious groups.[22] In table A.5, we verify that dropping this outlier does not affect our main effect on total giving.

Table 3 presents these results. The coefficients in the first two columns indicate that both indices significantly predict giving in absence of the "white man," but this significant positive correlation does not hold in the treatment villages. For example, the coefficient on the monetary contributions index in column (1) tell us that a one standard deviation increase in this variable is associated with 453 higher giving in the games where he is absent. The significant coefficient on the interaction term with the "white man" treatment indicates that the predicted correlation is smaller in the games where he is present. Finally, an F-test of the sum of these two coefficients ($453 - 429 = 24$) is insignificant, indicating contributions have no predictive power in the "white man" group. The same account holds for the more aggregate measure in column (2). Here the sum of the two coefficients on the contributions index and its interaction is -51 and statistically insignificant—again, indicating no predictive power of behavior on giving in the "white man" group.

We place the most weight on these aggregate measures because the index helps reduce noise underlying any single measure. This can be seen when we present the same specification for the three most important contributions comprising the index in table 3. In each case, the indicator predicts giving in the games where the "white man" is absent. And, in each case, the sum of the coefficients on the indicator and its interaction with the "White man" treatment indicator suggest no predictive power in his presence. However, among the individual indicators, the interactions are insignificant, though they are all negative and substantial in magnitude.[23]

Overall, this prediction exercise demonstrates that individuals deviate from giving based on their real-world behavior in the presence of the white foreigner.

5.2 Giving Based on Perceptions of Aid-Testing

We posit a second demand effect arises from the perception that the white foreigner is there to assess their need for aid. In this case, players may act out of a desire to seem aid-worthy. If they know aid is means tested, they would act poor to signal that they cannot give money away easily—they would give less in the games where the "white man" is present. This type of need-signaling is more likely to arise when people have had more experience with aid, and had a chance to learn about means testing. This serves as our second empirical prediction. If people act based on perceptions that the white foreigner is testing them for aid, we should observe *less* giving in his presence among more aid-exposed villages. We should also observe that they are

[21] After that there is a steep drop-off, with 20 percent contributing labor to religious groups and 10 percent or less contributing to each of the other groups.

[22] This is 22 standard deviations above the mean and tenfold larger than the value contributed at the 99th percentile. It is also twice the maximum of any other giving variable (300,000 for parent-teacher associations).

[23] The other indicators, which are relatively unimportant sources of contributions, are not significantly correlated with giving, even in control areas. (This is not shown, but is available from the authors upon request.) Thus, individually, they are uninformative as tests of how the correlation falls in treated areas.

more likely to believe the games where he is present aim to test them for aid. To examine this prediction, we draw on both giving in the games as well as questions asking respondents what they believed to be the purpose of the games.

We begin with a simple but powerful test of this prediction in table 4. We ask: Is there any observed relationship between aid and giving in the control group where the "white man" is absent, and in the treatment group where he is present? We use our most comprehensive measure of aid engagement, the years of NGO activity.

We find a clear pattern in columns (1)-(2). Aid exposure has no significant relationship on giving in control areas. But it has a significant negative relationship with giving in treatment areas. Those residing in more aid-exposed villages give less, but only when the white foreigner is present.

Villages receiving aid for longer periods have other characteristics that affect measured generosity. For example, these communities might be larger, less integrated, more cooperative, richer, or have more exposure to white foreigners. These features could serve as confounds to the observed pattern if they also lead people to give less in the presence of the white foreigner. For example, people in more cooperative societies may give less because they respond negatively to the presence of an outsider. To account for these potential effects, in columns (3)-(4), we control for a number of village-level characteristics, including: village size and the presence of a market, which are important indicators of economic development and the degree of market integration; whether the village has communal farms or a labor gang, which proxies for social cooperation; and the number of buildings burned during the civil war, as postconflict aid may have been targeted to areas with higher levels of historical violence.

In columns (5)-(6), we also introduce four key individual-level covariates. This includes two proxies for economic wellbeing—a household asset index and the years of formal education attained. It also includes two indicators measuring past exposure to foreigners—never met a white person before, and met a white person between one and 10 times (both defined relative to the omitted category of met a white person more than 10 times). It is interesting to note that these two variables exert no significant impacts on our generosity measures.[24] Moreover, the inclusion of the controls do not affect the pattern that giving decreases with aid exposure, but only when the man is present.

This pattern is also difficult to square with alternative accounts that posit giving may have increased owing to some other characteristic of the "white man." For example, consider the possibility that allocations increased in treatment areas because the white foreigner had a more pleasant demeanor, and not because of his race or nationality. Then, it is hard to understand why aid exposure would have a negative impact on giving in the presence of a more pleasant person.

In Table 5, we continue building on the results in Table 4. We introduce interaction terms between aid exposure and the "white man" treatment to demonstrate that the effects observed

[24] These questions about meeting white persons before were administered in the baseline survey in treatment and control areas. The survey, on average, preceded the games by 1.3 months, though there is variation from under a week to nearly three months. If answering these questions increased the salience of issues related to white foreigners, we would expect to see a significant interaction between the treatment effect and time between survey and games. Table A.4 shows that there is no significant effect of this interaction.

in the previous table are significantly different from one another. Column 1 presents the results without controls. Column 2 includes the village-level controls, and column 3 includes both individual and village-level controls. The control variables are also included in interaction with the "White man" treatment indicator.[25] The significant coefficient on the "Years NGO activity x white man" term tells us that across these specifications, the increase in giving is relatively smaller among places that have received aid for a longer period of time.

These effects are consistent with the idea that the two types of demand effects exert opposite effects on giving. For example, consider the specification in column (1). The coefficient on the "White man" variable tells us the treatment effect among villages that had zero years of NGO aid, which represents about 40 percent of our sample. In these places, giving increased by 795 leones, or 32 percent (relative to the overall control group mean of 2,944 leones). We can think of this effect as arising from the first type of demand effect—people give more because this is what they believe the white foreigner wants them to do. But there is a second, offsetting force captured by the negative interaction term. The coefficient on this term indicates that each year of aid engagement lowers the treatment effect by 22 leones. We can think of this effect as arising from the second type of demand effect—people give less to signal need because they believe the white foreigner is testing them for aid. Given these two opposite forces, the second effect comes to dominate in the most highly aid-exposed villages. Those residing in villages that have received aid for 36 years or longer (above the 86th percentile of the aid distribution), give *less* in the presence of the white foreigner.

Columns (4)-(6) repeat these three specifications, but drop the top five percent (that is, three) most aid-exposed villages. The results are robust to their exclusion. The coefficients in column (4) indicate that the treatment effect becomes negative for those people residing in villages that have received 16 or more years of aid.[26] This corresponds to the 80th percentile of the aid distribution, excluding outlier villages. In other words, in this specification, the top 20 percent most aid-exposed villages give *less* in the presence of the "white man." This is again consistent with the idea those who have the most experience with development aid face an incentive to act poor, and signal that they have relatively little to give away by lowering their game allocations when the "white man" is in the room. The first two columns of table A.7 also show that the pattern of results remains the same using two alternate measures of aid exposure—the "NGO aid" indicator and the "Years of NGO-owned facilities" variable.

If this behavior reflects players' beliefs that the "white man" is there to assess their need for aid, then individuals from highly aid-exposed areas should also be more inclined to believe that the purpose of the games with the "white man" is to test them for aid suitability. To examine this prediction, we next turn to data on perceived beliefs regarding the purpose of the games.

First, we examine overall respondent choice across treatment and control communities (not disaggregated by aid exposure). Column (1) of table 6 uses a simple logit model where the dependent variable equals one if the respondent chooses "Aid test" and zero otherwise. We find a significant treatment effect of 0.439, equivalent to 56 percent lower odds of choosing this

[25] Since these controls are only available for a subsample of the observations, column (4) of table A.5 verifies that our main treatment effect holds in this subsample.

[26] The treatment effect in non aid-exposed villages is 1,103 leones, and the coefficient on the interaction terms tells us each year of aid reduces it by 70 leones.

option in the presence of the white researcher. However, since this logit specification groups together the money and research options, we next employ a multinomial logit model, using a categorical dependent variable of participant belief that includes all three choices. Columns (2)-(4) show the relative coefficients for each option pair: "Aid test" over "Research," "Aid test" over "Give money," and "Give money" over "Research." The results are consistent with those in the first column—in the presence of the white foreigner, the relative risk ratio of choosing "Aid test" is 54 percent lower than choosing "Give money," and 61 percent lower than choosing "Research."[27] However, the treatment has no significant effect in the choice between "Give money" and "Research." Thus, in the presence of the white foreigner, players on average were more inclined to believe that the games aimed to find out about the community, which was the stated intent of the team's visit, or to believe that the games were designed to hand out money, which is what the team actually did. This is in contrast to perceiving an underlying agenda that the team was there to test them for aid-worthiness.

However, the pattern of beliefs regarding aid testing vary substantially depending on the village's past exposure to development aid. Table 7 examines this heterogeneity by introducing interaction terms between the "white man" treatment and years of aid with NGO activity. We present these results for the full sample, but table A.8 shows that the results are robust to dropping the three most aid-exposed villages. Columns (1)-(4) present the results without any controls, while columns (5)-(8) show the results with the village-level and main individual-level controls for wealth and past exposure to foreigners.

The results show a clear pattern: while communities with no exposure are less likely to choose the "Aid test" option, this effect is countered with additional years of aid experience.

Column (1) presents the logit on "Aid test." The coefficient on the "white man" term implies that people living in communities with no aid exposure have a 76 percent lower odds of believing that the purpose of the game is to test people for aid, in the presence of the white foreigner. However, each additional year of aid erodes this effect by 4.8 percent. Thus, communities that have received more than 16 years of aid (representing the top 23 percent of the aid distribution), associate the white foreigner positively with the "Aid test" option. In other words, individuals from highly aid-exposed villages are more inclined to believe that the purpose of the games is to test them for aid, when the "white man" is present.

Columns (2)-(4) show that the results are similar with the equivalent multinomial logit specification: each year of aid exposure increases the relative risk of choosing "Aid test" over "Give money" by 4.7 percent, and over "Research" by 5.1 percent in "white man" treatment villages, relative to control. Again, the associated relative risk of choosing the research versus money options are not significantly different. Columns (5)-(8) also indicate that the results are robust to the controls.[28]

[27]The relative risk ratio in the multinomial logit specification is analogous to the odds ratio in the logit specification.

[28]The evidence on beliefs also helps us rule out an alternative interpretation of the results, that respondents give more in response to the "white man" treatment to signal they are altruistic and thus more deserving of aid. Under this "altruism-signaling" account, the less aid-exposed participants display differential levels of generosity in the hopes of securing future aid from the white foreigner. However, if this account held, then the less aid-exposed participants should also be differentially inclined to believe that the purpose of the games is to test them for aid. Instead, we find precisely the opposite result.

The coefficients in table A.8 show that the implied effects are similar when we drop the outlier villages. For the logit specification, the coefficients suggest that the top 20 percent of the aid exposure distribution excluding outliers are more inclined to believe that the purpose of the games is to test them for aid when the "white man" is present. Table A.7 also verifies that the same pattern of results holds with two alternative measures of aid exposure (the "NGO aid" indicator and "Years of aid with NGO activity"). Overall, these results provide robust evidence that those in the most aid-exposed communities gave less in the presence of the white foreigner and were more inclined to believe that the games were designed to assess them for aid when he was present. This is consistent with the idea that those who have experience with aid disbursement have an incentive to act poor by giving less in the games, as a signal that they are in need of further assistance.

These differential impacts are again very difficult to square with the idea that some other characteristic induces the "white man" effect, beyond his identity as white foreigner. If another characteristics were salient, it would be surprising to observe that heterogeneous impacts on both aid test beliefs and measured generosity based on aid exposure in the past.

Next, we conduct a third, related empirical test. If players give based on perceptions of aid testing, then village leaders should also give less, as they are more likely to be aware of how aid is given out. If, indeed, this is related to their beliefs about being tested for aid by the "white man," we should again observe that they are more inclined to believe that the purpose of the games is to test them for aid, when he is present.

To examine this account, we use our household survey to generate an indicator of whether the respondent comes from the household of a customary authority leader, which includes members belonging to the three groups that traditionally symbolize power and authority in Sierra Leone: chiefs, religious leaders, and leaders of secret societies. In table 8, we interact the treatment with this "customary authority" variable. In column (1), the coefficient on the "White man" variable indicates that those who are not from leadership households give 919 leones more in treatment areas, which represents a 31 percent increase over the control mean. The sum of this coefficient and the coefficient on the interaction term tell us that leadership households, in contrast, give 492 leones *less* in the presence of the white foreigner.

It is possible that this subgroup may be richer, and may have been exposed more to white persons in the past. It is also likely that they are from the dominant ethnic group in the village, and that older males are over-represented among this group. To account for these potential confounds, we incorporate our main individual controls, and also include three additional individual-level variables: gender, age and whether the respondent is from the village's ethnic majority, all interacted with the "white man" treatment. In addition, we retain our village-level controls. Column (2) shows that the interaction effect remains large and significant with the addition of these other covariates.

Here, both types of demand effects may be at play. First, leadership households have a relatively high status and may not perceive much of a power difference between themselves and the white foreigner. If the demand effect to give more in the presence of the "white man" comes from perceived power differentials, we expect the response to be relatively smaller among higher status players. However, this by itself would imply a smaller positive response, rather than a net negative response. In addition, leadership households may also give less in response to the white foreigner because they perceive he is there to test them for aid, and they are more familiar with how aid disbursement works. In that case, they may give less to signal that their community is in need of assistance. Columns (3)-(6) of table 8 provide some evidence favoring this account. Here, we examine how the customary authority interactions affect beliefs regarding the purpose of the games. The evidence here is weaker, in that we see no significant effect on the "Aid test" option relative to the "Give money" option. However, column (5) indicates that individuals from leadership households disproportionately believed that the games conducted in the presence of the white foreigner were designed to test them for aid relative to the "Research" option.

5.3 Ruling Out Alternative Accounts

In this section, we devise another empirical test to address the account that other characteristics of the white foreigner, besides his race and nationality, lead to the estimated effect.

If giving responded to another personality trait in the white foreigner, we would plausibly expect the treatment effect to vary based on this personality trait among players. For example, a more friendly player may respond to a more friendly researcher, if friendliness were the key factor. Or, happier players may respond more to a happier researcher, with a more smiling countenance, if happiness and demeanor were at play. To test these effects, we construct psychometric measures of depression and PTSD, and a measure of the number of good friendships each player has within the community, as a proxy for friendliness. Importantly, all of these measures are collected from players in the baseline survey prior to the games.

In addition, we look at two demographic traits—gender and education. If the "white man" was perceived to be more physically attractive, we may expect differential responses among female players. Or, if he conveyed he was more educated, even in non-speaking role, we might expect differential impacts among players who are educated. However, table 9 shows that there are no significant effects based on any of these traits. This range of null interactions casts doubt on the importance of these other characteristics in accounting for increased giving in the presence of the white foreigner.

6 Conclusion

This paper has examined whether researcher identity affects measured generosity. We use a lab-in-the-field experiment which varied the presence of a white foreigner across behavioral games in 60 Sierra Leonean villages. This white foreigner played a silent role, without directly interacting with the players. Yet, his mere presence boosted total giving by 19 percent in the dictator games.

Our findings suggest that players act based on their perception of this individual as a white foreigner. We uncover evidence of two such demand effects. First, players appear to give more because they perceive this is what he wants them to do. Consistent with this, we find that real-world contributions strongly predict giving in the games where the "white man" is absent, but not in the games where he is present. In short, giving is less indicative of true generosity under white presence.

Second, we find that players from more aid-exposed villages and households of village leaders give less in the presence of the white foreigner. When asked about the purpose of the games, they are also more likely to report that the games where he is present test them for aid suitability. This is consistent with the idea that those knowledgeable about aid perceive the games as a form of means testing and give less to signal they are poor and in need of further assistance.

These results support the "white man" mechanism: if measured generosity increased due to other researcher characteristics besides nationality and race, we would not expect the treatment effect to vary with exposure to development aid. We also present evidence against alternative mechanisms by examining if the effects vary based on other player traits. We find no differential impacts based on characteristics such as player friendliness proxied by good friendships, or psychometric measures of depression and PTSD, and nor do we find differential effects based on gender and education. Taken together, these results suggest that the researcher's identity as a white foreigner, rather than other traits such as demeanor, explain the main result.

Our study holds direct and important implications for how we interpret current measures of generosity, target aid, and evaluate aid effectiveness in the developing world. First, it shows that measures of generosity may be upward-biased relative to underlying other-regarding preferences, depending on who is carrying out the measurement. In addition, if there is heterogeneity in how players respond to foreign researchers, this makes it challenging to interpret variation in cross-national measures of generosity. For example, Henrich et al. (2006) find large differences in measured generosity across developing countries and attribute this to differences in cultural practices. However, if communities vary in their response to foreigners based on factors such as previous exposure to aid, then the results may be driven in part by different responses to foreigners rather than cultural differences per se.

Our results also hold implications for the conduct of aid-targeting exercises. Communities in aid-dependent countries such as Sierra Leone are subject to frequent "needs assessments" and other policy-specific data collection exercises that are meant to improve aid-targeting and respond to community needs. Two popular approaches toward aid-targeting include proxy means tests, which identify the poor based on outsiders' measurement of the asset and income position of households, and community-based targeting, where needs assessments are left to the community (Alatas et al., 2013). If communities that have received aid in the past act strategically to signal need and secure more aid in the future, then targeting based on proxy means tests may not work as well as intended and could favor those who have already received aid previously. As such, community-based targeting may be a more effective way of identifying the poor.

In addition, there are a number of implications for studies seeking to evaluate aid interventions. If communities receiving aid treatments respond strategically to signal need, then we may underestimate the impact of aid on economic outcomes. Aid evaluations increasingly use behavioral games and experiments to circumvent challenges associated with subjective responses and generate more objective measures of economic and social capital outcomes. However, these approaches may also underestimate the true impact of aid if players are strategically signaling need during games conducted in intervention communities. This is important in light of the fact that a number of recent studies have found insignificant effects of community-driven development (CDD) programs on social capital outcomes (Avdeenko and Gilligan, 2012; Casey et al., 2012; Humphreys et al., 2012).

Finally, if behavioral games are used for evaluation purposes, then imbalance in the racial and national composition of evaluators across treatment and control areas could generate a different type of bias. For example, if the foreign NGO implementing the program takes a greater interest in observing behavioral games in program areas, then their mere presence could lead to the spurious conclusion that the aid itself has engineered meaningful increases in generosity or other-regarding preferences. Overall, by demonstrating how the presence of a foreign researcher can affect measured altruism, our analysis points to the import of considering researcher identity broadly in measuring and interpreting behavioral outcomes.

Bibliography

Acemoglu, D., Reed, T., and Robinson, J. A. (2012). Chiefs.

Adair, J. G., Sharpe, D., and Huynh, C.-L. (1989). Hawthorne Control Procedures in Educational Experiments: A Reconsideration of Their Use and Effectiveness. *Review of Educational Research*, 59(2):215–228.

Adida, C. L., Laitin, D. D., and Valfort, M.-A. (2012). "One Muslim is Enough!" Evidence from a Field Experiment in France.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., Purnamasari, R., and Wai-Poi, M. (2013). Ordeal Mechanisms in Targeting: Theory and Evidence from a Field Experiment in Indonesia.

Anderson, B. A., Silver, B. D., and Abramson, P. R. (1988). The Effects of the Race of the Interviewer on Race-Related Attitudes. *Public Opinion Quarterly*, 52:289–324.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Avdeenko, A. and Gilligan, M. J. (2012). Community-Driven Development and Social Capital: Lab-in-the-Field Evidence from Sudan.

Bailar, B., Bailey, L., and Stevens, J. (1977). Measures of Interviewer Bias and Variance. *Journal of Marketing Research*, XIV(August):337–344.

Benz, M. and Meier, S. (2008). Do people behave in experiments as in the field? evidence from donations. *Experimental Economics*, 11(3):268–281.

Blaydes, L. and Gillum, R. (2012). Religiosity-of-Interviewer Effects: Assessing the Impact of Veiled Enumerators on Survey Response in Egypt.

Burns, J. (2006). Racial Stereotypes, Stigma and Trust in Post-Apartheid South Africa. *Economic Modelling*, 23(5):805–821.

Burns, J. (2012). Race, Diversity and Pro-Social Behavior in a Segmented Society. *Journal of Economic Behavior & Organization*, 81(2):366–378.

Cappelen, A. W., Moene, K. O., Sø rensen, E. O., and Tungodden, B. (2011). Needs Versus Entitlements—An International Fairness Experiment.

Cardenas, J. C. and Carpenter, J. (2008). Behavioural Development Economics: Lessons from Field Labs in the Developing World. *Journal of Development Studies*, 44(3):337–364.

Casey, K., Glennerster, R., and Miguel, E. (2012). Reshaping Institutions: Evidence on Aid Impacts Using a Pre-Analysis Plan. *Quarterly Journal of Economics*, 127(4):1755–1812.

Conteh, P. S. (2009). *Traditionalists, Muslims and Christians in Africa: Interreligious Encounters and Dialogue*. Cambria Press, Amherst.

Cotter, P. R., Cohen, J., and Coulter, P. B. (1982). Race-of-Interviewer Effects in Telephone Interviews. *Public Opinion Quarterly*, 46:278–284.

Diaper, G. (1990). The Hawthorne Effect: A Fresh Examination. *Educational Studies*, 16(3):261–267.

Fanthorpe, R. (2001). Neither Citizen Nor Subject? 'Lumpen' Agency and the Legacy of Native Administration in Sierra Leone. *African Affairs*, 100:363–386.

Fanthorpe, R. (2003). Humanitarian Aid in Post-War Sierra Leone: The Politics of Moral Economy.

Fanthorpe, R. (2007). Sierra Leone: The Influence of the Secret Societies, with Special Reference to Female Genital Mutilation. Technical Report August, Writenet.

Fearon, J. D., Humphreys, M., and Weinstein, J. M. (2009). Can Development Aid Contribute to Social Cohesion after Civil War? Evidence from a Field Experiment in Post-Conflict Liberia. *American Economic Review: Papers and Proceedings*, 99(2):287–291.

Finkel, S. E., Guterbock, T. M., and Borg, M. J. (1991). Race-of-Interviewer Effects in a Pre-Election Poll. *Public Opinion Quarterly*, 55:313–330.

Foa, E. B., Cashman, L., Jaycox, L., and Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The posttraumatic diagnostic scale. *Psychological assessment*, 9(4):445.

Fyfe, C. (1968). *A history of Sierra Leone*. CUP Archive.

Habyarimana, J., Humphreys, M., Posner, D. N., and Weinstein, J. M. (2007). Why Does Ethnic Diversity Undermine Public Goods Provision? *American Political Science Review*, 101(4):709–725.

Haley, K. J. and Fessler, D. M. T. (2005). Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game. *Evolution and Human Behavior*, 26(3):245–256.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., and Ziker, J. (2006). Costly Punishment across Human Societies. *Science*, 312(5781):1767–1770.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. L. (1994). Preferences, Property Rights, and Anonymity in Bargaining Games. *Games and Economic Behavior*, 7:346–380.

Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, pages 653–660.

Humphreys, M., Windt, P. V. D., and de la Sierra, R. S. (2012). Social and Economic Impacts of Tuungane. Technical Report June, Columbia University, New York.

Hyman, H. H. (1954). *Interviewing in Social Research*. University of Chicago Press, Chicago.

Jackson, P. (2006). Reshuffling an Old Deck of Cards? The Politics of Local Government Reform in Sierra Leone. *African Affairs*, 106(422):95–111.

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

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives*, pages 153–174.

Levitt, S. D., List, J. A., Blalock, H. M., and Sears, D. O. (2011). Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics*, 3(January):224–238.

List, J. A. (2005). The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions. Technical report, National Bureau of Economic Research.

List, J. A., Berrens, R. P., Bohara, A. K., and Kerkvliet, J. (2004). Examining the role of social isolation on stated preferences. *American Economic Review*, pages 741–752.

Mansuri, G. and Rao, V. (2013). *Localizing Development: Does Participation Work?* The World Bank, Washington, DC.

Masling, J. M. (1966). Role-related Behavior of the Subject and Psychologist and its Effect Upon Psychological Data. In Levine, D., editor, *Nebraska Symposium on Motivation*, pages 67–103. University of Nebraska Press, Lincoln, NB.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., and Fisher, P. (2007). The Hawthorne Effect: A Randomised, Controlled Trial. *BMC Medical Research Methodology*, 7(30).

Mensch, B. S. and Kandel, D. B. (1988). Underreporting of Substance Use in a National Longitudinal Youth Cohort: Individual and Interviewer Effects. *Public Opinion Quarterly*, 52:100.

Miyazaki, A. D. and Taylor, K. A. (2007). Researcher Interaction Biases and Business Ethics Research: Respondent Reactions to Researcher Characteristics. *Journal of Business Ethics*, 81(4):779–795.

Murphy, W. P. (1980). Secret Knowledge as Property and Power in Kpelle Society: Elders Versus Youth. *Africa: Journal of the International African Institute*, 50(2):193–207.

Nichols, A. L. and Maner, J. K. (2008). The Good-Subject Effect: Investigating Participant Demand Characteristics. *Journal of General Psychology*, 135(2):151–65.

Reese, S. D., Danielson, W. A., Shoemaker, P. J., Chang, T.-K., and Hsu, H.-L. (1986). Ethnicity-of-Interviewer Effects Among Mexican-Americans and Anglos. *Public Opinion Quarterly*, 50:563–572.

Richards, P. (2005). War as Smoke and Mirrors: Sierra Leone 1991-2, 1994-5, 1995-6. *Anthropological Quarterly*, 78(2):377–402.

Rosenthal, R. (1963). Experimenter Attributes as Determinants of Subjects' Responses. *Journal of Projective Techniques and Personality Assessment*, 27(3):324–331.

Rosenthal, R. and Rubin, D. B. (1978). Interpersonal Expectancy Effects: the First 345 Studies. *Behavioral and Brain Sciences*, 3:377–415.

Sawyer, E. (2008). Remove or Reform? A Case for (Restructuring) Chiefdom Governance in Post-Conflict Sierra Leone. *African Affairs*, 107(428):387–403.

Schultz, D. P. (1969). The Human Subject in Psychological Research. *Psychological Bulletin*, 72(3):214–228.

Vyas, S. and Kumaranayake, L. (2006). Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis. *Health Policy and Planning*, 21(6):459–468.

Webster, C. (1996). Hispanic and Anglo Interviewer and Respondent Ethnicity and Gender: The Impact on Survey Response Quality. *Journal of Marketing Research*, 33(February):62–72.

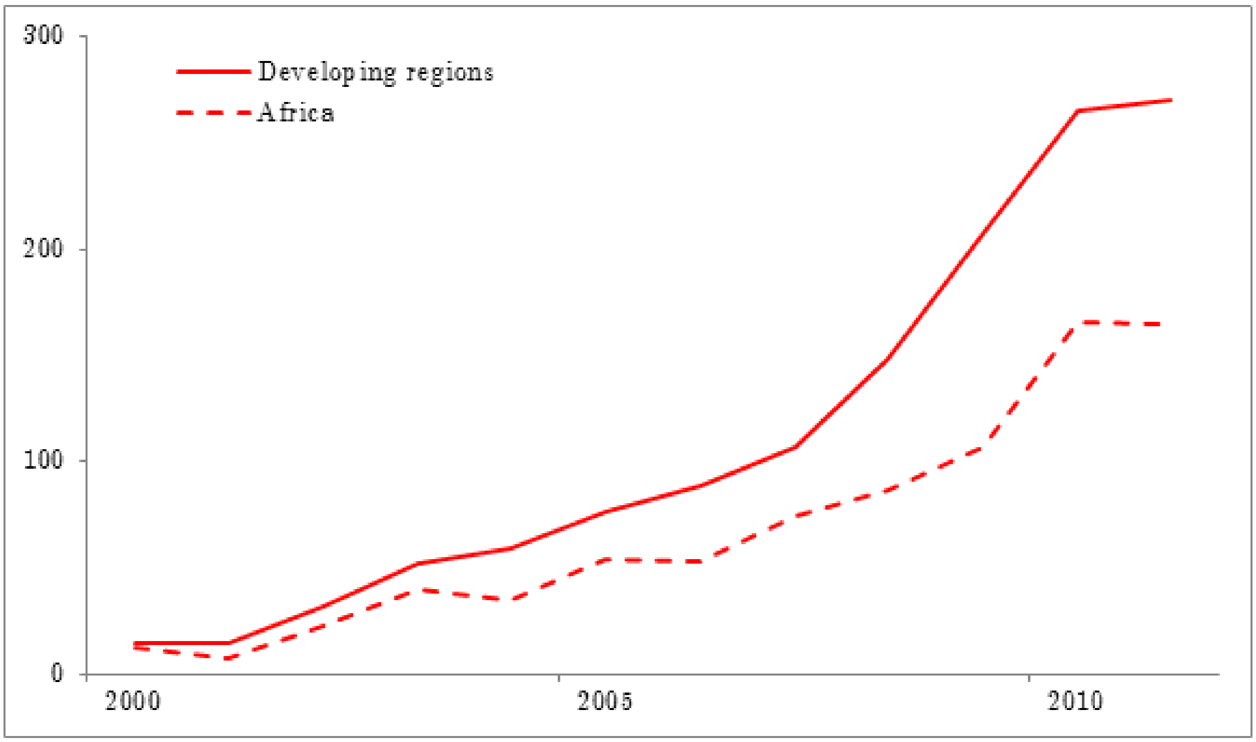
Whitt, S. and Wilson, R. K. (2007). The Dictator Game, Fairness and Ethnicity in Postwar Bosnia. *American Journal of Political Science*, 51(3):655–668.

World Bank (2012). World Development Indicators. Technical report, World Bank, Washington, DC.

Zung, W. W. (1971). A Rating Instrument for Anxiety Disorders. *Psychosomatics: Journal of Consultation Liaison Psychiatry*, 12(6):371–379.

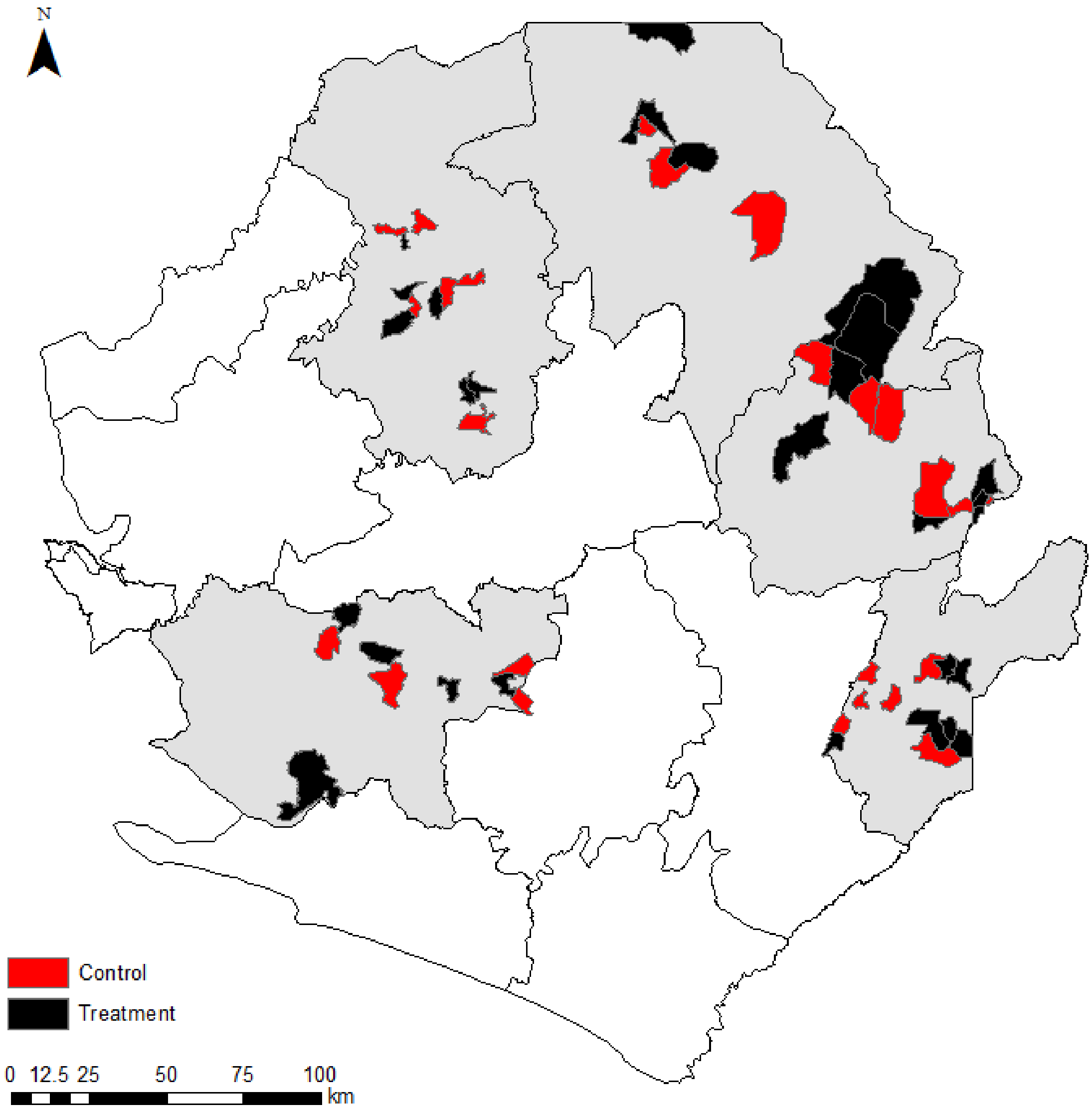
Zwane, A. P., Zinman, J., van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D. S., Hornbeck, R., Giné, X., Duflo, E., Devoto, F., Crepon, B., and Banerjee, A. V. (2011). Being Surveyed Can Change Later Behavior and Related Parameter Estimates. *Proceedings of the National Academy of Sciences of the United States of America*, 108(5):1821–6.

Figure 1: The increasing use of dictator games in developing countries



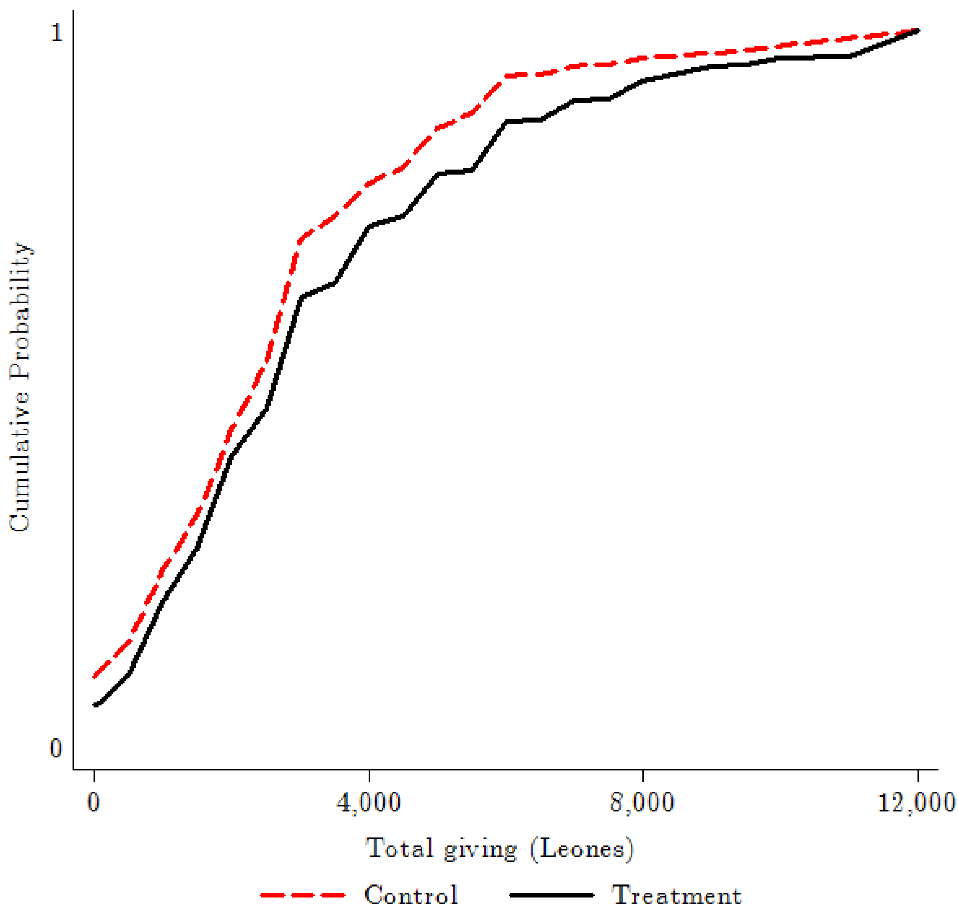
Note. This graph shows the number of mentions of dictator games in developing countries (solid line) and in Africa (dashed line), based on Google Scholar counts.

Figure 2: Treatment and control communities



Note. This figure shows the five districts of Sierra Leone included in the experiment (grey), the 30 treatment sections (black) and the 30 control sections (red). One pre-designated village from each of these sections comprises the treatment and control communities in the study.

Figure 3: CDF of total giving in control and treatment groups



Note. This figure shows the cumulative distribution function of total giving in the control and treatment groups.

Table 1: Key Individual and Village-Level Characteristics

	Treatment	Control	Difference	Std. Err.
Individual				
Years of education	2.06	1.90	0.160	(0.318)
Female	0.55	0.54	0.007	(0.039)
Ethnic majority	0.87	0.91	-0.039	(0.034)
Household asset index	0.00	-0.00	-0.003	(0.232)
Age	41.23	43.43	-2.228*	(1.172)
Customary authority	0.23	0.28	-0.053	(0.032)
Monetary contr. index	0.02	-0.00	0.020	(0.055)
All contr. index	0.04	-0.00	0.037	(0.041)
Village				
NGO aid	0.93	0.87	0.067	(0.076)
Years NGO-owned facilities	10.13	14.23	-4.100	(5.633)
Years NGO activity	11.13	15.70	-4.567	(5.570)
White aid visitors	0.34	0.37	-0.012	(0.116)
White non-aid visitors	0.41	0.37	0.046	(0.127)
Number of households	334.23	234.30	99.933	(94.736)
Market community	0.17	0.10	0.067	(0.087)
Buildings burned during war	6.10	5.43	0.667	(3.559)
Labor gang	0.93	0.97	-0.033	(0.054)
Communal farm	0.27	0.47	-0.200*	(0.106)
Dominant ethnic group				
Fullah	0.03	0.03	-0.000	(0.044)
Kono	0.20	0.20	-0.000	(0.000)
Koranko	0.13	0.17	-0.033	(0.050)
Loko	0.07	0.07	-0.000	(0.055)
Mende	0.40	0.40	0.000	(0.000)
Temne	0.13	0.13	-0.000	(0.055)
Yalunka	0.03	0.00	0.033	(0.032)

Note: Columns 1-2 present treatment and control group means, respectively. Column 3 shows coefficients from a regression of each variable on the white-man treatment indicator. Column 4 displays robust standard errors clustered at the village level. * is significant at the 10% level.

Table 2: White Man Presence and Measured Generosity

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving	(5) Total giving
White-man	537.320*	564.943*	564.681*	555.390*	548.941*
	(312.028)	(296.734)	(297.182)	(293.607)	(292.644)
Replacement			30.307	-40.539	
			(294.568)	(406.934)	
Replacement x white-man				127.623	
				(630.268)	
Control group mean	2943.966	2943.966	2943.966	2943.966	2943.966
Ethnicity fixed effects	No	Yes	Yes	Yes	Yes
Replacements Dropped	No	No	No	No	Yes
Observations	708	708	708	708	653

Note: Robust standard errors clustered at the village level in parentheses. All specifications include district fixed effects. Columns (2) to (5) include ethnicity fixed effects. Replacement equals one if the game participant was a replacement household member. ** is significant at the 5% level and * is significant at the 10% level.

Table 3: Predicting Measured Generosity with Real-World Contributions

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving	(5) Total giving
Monetary contributions index	453.211*				
	(241.460)				
Monetary contributions index x white-man	-429.704*				
	(245.217)				
All contributions index		582.818**			
		(263.763)			
All contributions index x white-man		-633.684**			
		(297.104)			
Public facility			540.712*		
			(272.658)		
Public facility x white-man			-183.594		
			(419.680)		
Brushed road				729.763**	
				(301.932)	
Brushed road x white-man				-367.474	
				(417.499)	
Religious money					0.037**
					(0.016)
Religious money x white-man					-0.013
					(0.025)
White-man	540.930*	542.827*	633.846*	648.127*	601.237**
	(291.159)	(292.115)	(318.936)	(326.096)	(280.947)
Proportion who contribute	—	—	.47	.348	.312
Observations	652	652	647	644	652

Note: Robust standard errors clustered at the village level in parentheses. "Religious money" is the amount of financial contributions to religious groups over the past three months; "Monetary contributions" is an index of financial contributions to all village groups; "Public Facility" indicates whether the respondent has contributed labor or time towards construction or maintenance of facilities; "Brushed road" indicates if someone has brushed a village road the past month; "All contributions" is a broader index of all financial and labor contributions. ** is significant at the 5% level and * is significant at the 10% level.

Table 4: Aid Exposure and Generosity in Control and Treatment Areas

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving	(5) Total giving	(6) Total giving
Years NGO activity	5.433	-20.152**	5.577	-24.879**	4.248	-26.208**
	(6.467)	(9.488)	(7.667)	(11.984)	(6.746)	(12.202)
Met white person 1 to 10 times					-212.215	173.181
					(243.837)	(425.442)
Never met white person					-572.625	-340.905
					(453.383)	(458.827)
Individual controls	No	No	No	No	Yes	Yes
Village controls	No	No	Yes	Yes	Yes	Yes
White-Man Sample	No	Yes	No	Yes	No	Yes
Observations	324	329	324	329	313	324

Note: The odd (even)-numbered columns reports regression coefficients on a restricted sample of the control (treatment) communities. All specifications include district and ethnicity fixed effects. Village-level controls include village size, buildings burned during the war, indicators for whether there is a market, a labor gang, and a communal farm, and their interactions with the white-man treatment. Columns (5) and (6) control for respondents' socio-economic status (asset index and education) and his/her previous exposure to white foreigners. Robust standard errors clustered at the village level in parentheses. ** is significant at the 5% level and * is significant at the 10% level.

Table 5: Aid and the White Man Effect on Generosity

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving	(5) Total giving	(6) Total giving
White-man	794.805**	675.298	730.540	1103.483***	855.183	972.255
	(332.336)	(963.310)	(965.073)	(381.450)	(896.331)	(922.845)
Years NGO activity x white-man	-22.357**	-21.302**	-22.227**	-70.006*	-93.031*	-83.731*
	(9.757)	(9.771)	(10.020)	(35.154)	(46.571)	(42.509)
Years NGO activity	6.057	5.512	4.259	50.097	72.195	61.548
	(6.673)	(7.091)	(6.299)	(31.970)	(45.253)	(41.329)
Met white person 1 to 10 times x white-man			362.405			291.325
			(500.050)			(511.249)
Never met white person x white-man			125.847			-234.626
			(647.737)			(656.409)
Met white person 1 to 10 times			-203.631			-174.842
			(251.762)			(252.847)
Never met white person			-495.643			-189.658
			(429.583)			(454.282)
Individual controls	No	No	Yes	No	No	Yes
Village controls	No	Yes	Yes	No	Yes	Yes
Outliers Dropped	No	No	No	Yes	Yes	Yes
Observations	653	653	637	620	620	604

Note: See table 4 for notes on control variables. Robust standard errors clustered at the village level in parentheses. ** is significant at the 5% level and * is significant at the 10% level.

Table 6: White-man Presence and Participant Beliefs

	Logit	Multinomial Logit		
	(1)	(2)	(3)	(4)
	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research
White-man	0.439**	0.455**	0.392**	0.862
	(0.156)	(0.161)	(0.155)	(0.159)
Observations	664	664	664	664

Notes. Robust standard errors clustered at the village level in parentheses. Columns 1-4 display odds / relative risk ratios for a unit increase in the independent variable. Column 1 displays the odds ratio from a logit regression on the "Aid Test" indicator variable. Columns 2-4 display the relative risk ratios for each pair of choices from a multinomial logit regression on the categorical variable of participant beliefs over the 'aid test', 'give money' and 'research' choices. All specifications include district and ethnicity fixed effects. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.

Table 7: White-man Presence and Beliefs around Aid Testing

	Logit	Multinomial Logit			Logit	Multinomial Logit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research
White-man	0.237***	0.253***	0.189***	0.744	0.662	0.596	0.647	1.086
	(0.091)	(0.010)	(0.077)	(0.168)	(0.450)	(0.415)	(0.572)	(0.637)
Years NGO activity x white-man	1.048**	1.047**	1.051**	1.004	1.060***	1.057***	1.062***	1.004
	(0.020)	(0.020)	(0.022)	(0.006)	(0.018)	(0.018)	(0.021)	(0.006)
Years NGO activity	0.968***	0.970***	0.964***	0.994*	0.972**	0.972***	0.967***	0.994**
	(0.010)	(0.010)	(0.011)	(0.004)	(0.01)	(0.010)	(0.011)	(0.003)
Met white person 1 to 10 times x white-man					1.682	1.71	1.506	0.88
					(1.682)	(0 .939)	(1.023)	(0.363)
Never met white person x white-man					8.01*	7.586*	9.485*	1.21
					(8.505)	(8)	(11.74)	(0.827)
Met white person 1 to 10 times					1.415	1.371	1.619	1.181
					(0.490)	(0.473)	(0.717)	(0.338)
Never met white person					0.604	0.6	0.643	1.072
					(0.540)	(0.531)	(0.674)	(0.554)
Individual-level controls	N	N	N	N	Y	Y	Y	Y
Village-level controls	N	N	N	N	Y	Y	Y	Y
Observations	652	652	652	652	636	636	636	636

Notes. Robust standard errors clustered at the village level in parentheses. See Tables 4 and 6 for additional notes and text for specification and controls. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.

Table 8: Authority Differentials

	OLS	OLS	Logit	Multinomial Logit		
	(1)	(2)	(3)	(4)	(5)	(6)
	Total giving	Total giving	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research
White-man	918.656***	45.131	0.6338	0.57	0.616	1.081
	(314.997)	(1631.26)	(0.940)	(0.815)	(1.203)	(1.364)
Customary authority x white-man	-1,411.335***	-1366.726***	2.445	2.007	4.688*	2.335
	(472.147)	(500.895)	(1.751)	(1.50)	(3.751)	(1.247)
Customary authority	948.297***	800.867**	1.342	1.365	1.23	0.9
	(289.489)	(306.241)	(.471)	(0.470)	(0.584)	(0.299)
Met white person 1 to 10 times x white-man		323.952	1.23	1.248	1.115	0.894
		(484.825)	(0.806)	(0.828)	(0.846)	(0.383)
Never met white person x white-man		153.234	1.23	5.345	5.427	1.015
		(630.370)	(0.806)	(5.784)	(6.561)	(0.664)
Met white person 1 to 10 times		-62.055	5.389	1.261	1.512	1.198
		(248.416)	(5.764)	(0.412)	(0 .610)	(0.325)
Never met white person		351.916	0.469	0.457	0.504	1.103
		(401.087)	(0..397)	(0.384)	(0.498)	(0.537)
Individual-level controls	N	Y	Y	Y	Y	Y
Village-level controls	N	Y	Y	Y	Y	Y
Observations	653	634	634	634	634	634

Notes. Robust standard errors clustered at the village level in parentheses. Customary authority indicates that the respondent is from the household of a chief or secret society leader. Other individual-level controls include respondent age, gender, past exposure to white persons, ethnic majority group dummy, years of education attained, a household asset index, and their respective interactions with the white-man variable. Village-level controls include village size, buildings burned during the war, indicators for whether there is a market, a labor gang, and a communal farm, and their interactions with the white-man treatment. All specifications include district and ethnicity fixed effects. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.

Table 9: Alternative Accounts

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving	(5) Total giving
White-man	555.193	667.584	679.957	786.885	378.466
	(339.984)	(473.190)	(539.930)	(524.731)	(347.408)
Female x white-man	23.836				
	(397.531)				
Female	-738.990***				
	(275.378)				
No. of friends x white-man		-74.942			
		(159.803)			
No. of friends		-57.865			
		(106.043)			
Depression x white-man			-10.045		
			(88.491)		
Depression index			-131.128***		
			(44.922)		
PTSD x white-man				-25.405	
				(54.626)	
PTSD				-14.078	
				(33.143)	
Years of education x white-man					93.191
					(65.370)
Years of education					6.165
					(44.908)
Observations	649	653	636	615	645

Note: Robust standard errors clustered at the village level in parentheses. ** is significant at the 5% level and * is significant at the 10% level.

Supplementary Information Appendix for: ‘White Man’s Burden’? Experimental Evidence on Generosity and Foreigner Presence

Jacobus Cilliers Oeindrila Dube Bilal Siddiqi

October 29, 2014

Tables A.1 and A.2 present descriptive statistics of the key independent and dependent variables used in the analysis.

Our preferred specification is to look at the aggregate giving across the dictator games as we do not have the power to identify different effects across games. Consistent with this, if we look separately at total giving disaggregated by game in Table A.3, we observe coefficients of 224, 119 and 184 for the anonymous own-village, non-anonymous own village and other village games, respectively. Individually, the effect is marginally insignificant for the other village game (with a p-value of .101), where the coefficient indicates an intermediate increase in giving in response to the treatment. The effect appears to be most precisely estimated for the anonymous own village game. However, as discussed further in Table A.6, we are not to statistically distinguish effects across games.

Table A.4 shows that our main results are robust to age and communal farm indicator which appeared to significantly differ in treatment and control areas at the 10% level. Columns (3) also shows that the treatment effect does not vary based on the number of days in between the survey and the game sessions. Columns (4)-(5) show robustness to different ethnicity fixed effects, generated on the basis of broader linguistic categories using parent groups from the third and fourth levels in the Atlantic language hierarchy. The Atlantic language hierarchy is catalogued at http://multitree.org.

Table A.5 shows that our main results are robust to different samples used during the analysis. Column (1) drops one observation where monetary contributions to religious groups was extremely high (with a value of 600,000 which exceeded the 99th percentile of the distribution by a factor of 10). Column (2) drops three village outliers in length of exposure to aid. Column (3) drops the two districts— Koinadugu and Bombali— where more translators were required due to the linguistic diversity of the area. This shows that the results are not sensitive to translator quality. Finally, column (4) shows the robustness of the effect to the sub-sample for which our main individual and village control variables are available.

Table A.6 examines effects in the three dictator games using a pooled specification. The first column examines the average impact on giving across all three games. We include game fixed effects denoted by Game 1 and Game 2 (defined relative to the omitted category, Game 3 – the Anonymous Other Village Game). In this column, the coefficient of 174 on the White-Man variable indicates that giving on average increased by 17 percent above the control group mean of 997.

In the remaining columns of this table, we introduce interaction terms between the white-man indicator and different game types. In columns (2), we consider the effects of both Game 1 and Game 2 relative to Game 3, where giving increased by an intermediate amount in response to treatment. The interactions on Game 2 x White-man and Game 1 x White-man are insignificant, indicating that the treatment effects in these games are not significantly distinguishable from the effect in Game 3.

Next, we try to assess if the effect in any one game differs significantly from the average of the other two. Column (3) examines if the effect in Game 2 significantly differs relative to the average of Games 1 and Game 3, which comprise the omitted category. Column (4) asks if the effect in Game 1 differs significantly from Games 2 and 3, which together comprise the omitted category here. And, column (5) considers if the effect in Game 3 differs significantly from Games 1 and 2 which are the omitted category. The insignificant interaction terms in these specifications indicate that overall, the effects across games are not statistically distinguishable from one another.[1]

Table A.7 shows that the results from Table 4, 5 and 7 hold for different metrics of aid exposure. “NGO-aid” is a dummy variable for whether an NGO either owns a school/clinic or currently provides resources to the local school or clinic, or has contributed to the construction of public facilities. “Years NGO-owned facilities” is the number of years that a NGO has owned either a school or clinic in the village, and thus, has more fine-grained variation than the NGO aid indicator.

Table A.8 shows the robustness of the results in Table 7 to dropping the 5% most aid

[1]These results remain the same if two game effects are put into each specification i.e., if we also include Game 1 dummy variables in columns (3)-(5).

exposed areas, comprising three villages. This table utilizes our comprehensive aid variable — "Years of NGO activityand examines its interaction effect with the white-man treatment. The coefficients in column (1) indicate that those in the top 20% most aid exposed villages are more inclined to believe that the games in which the white man is present are designed to test them for aid.

Finally, Table A.9 shows the results from Tables 3, 5, and 8 disaggregated by game. The coefficients suggest that the effects are broadly consistent across game, though they are estimated with different degrees of precision. For example, the fall in the degree to which real-world contributions predict giving is more precisely estimated for the second game, the Non-anonymous own-village game. On the other hand, the interactions with exposure to aid are more precise for the other two games. Finally, the customary authority interactions are precisely estimated across all three games.

Table A.1: Summary Statistics of Key Individual-level Variables

	Observations	Mean	Std. Dev.	Min.	Max.
Total giving	708	3216.95	2648.11	0.00	12000.00
Game 1 (Anonymous Own-village)	720	1065.28	1032.12	0.00	4000.00
Game 2 (Non-anonymous Own-village)	720	1211.25	995.59	0.00	4000.00
Game 3 (Anonymous Other-village)	708	959.04	1005.16	0.00	4000.00
Giving	2148	1079.19	1015.91	0.00	4000.00
Aid test	719	0.14	0.35	0.00	1.00
Give money	719	0.64	0.48	0.00	1.00
Research	719	0.22	0.41	0.00	1.00
Contr. religious group	665	0.31	0.46	0.00	1.00
Contr. secret society	665	0.07	0.25	0.00	1.00
Contr. village development committee	665	0.08	0.27	0.00	1.00
Contr. women's group	665	0.10	0.30	0.00	1.00
Contr. youth group	665	0.09	0.29	0.00	1.00
Contr. parent-teacher association	665	0.05	0.22	0.00	1.00
Religious money	714	4631.52	26876.92	0.00	600000.00
Brushed road	710	0.35	0.48	0.00	1.00
Public facility	714	0.47	0.50	0.00	1.00
Customary authority	720	0.25	0.44	0.00	1.00
Met white person 1 to 10 times	715	0.46	0.50	0.00	1.00
Never met white person	715	0.10	0.30	0.00	1.00
Age	712	42.33	15.38	12.00	110.00
Years of education	711	1.98	3.61	0.00	13.00
Female	715	0.54	0.50	0.00	1.00
Ethnic majority	714	0.89	0.31	0.00	1.00
Household asset index	706	0.00	1.97	-1.72	20.34

Note: Total giving is the sum of giving in the Anonymous Own-village, Anonymous Other-village, and Non-anonymous Own-village games. Giving is the amount given across the three dictator games when the observations from the games are pooled together. Aid test, Give money and Research are indicators for whether participants believed that researchers wanted to test for aid, distribute money, or find out more about the community, respectively. Customary authority is an indicator of whether the individual comes from the household of a chief, religious leader or secret society leader. The variables starting with "Contr. " indicate whether the respondent contributed financially to the respective group or not; public facility indicates whether respondent has contributed labor or money to a public facility the past 6 months; and "Road brush" if respondent had brushed a village road in the past month.

Table A.2: Summary Statistics of Key Village-level Variables

	Observations	Mean	Std. Dev.	Min.	Max.
NGO aid	720	0.90	0.30	0.00	1.00
Years NGO-owned facilities	720	12.18	22.50	0.00	106.00
Years NGO activity	720	13.42	22.41	0.00	106.00
Number of households	720	284.27	411.02	3.00	2000.00
Market community	720	0.13	0.34	0.00	1.00
Buildings burned during war	720	5.77	15.43	0.00	100.00
Labor gang	720	0.95	0.22	0.00	1.00
Communal farm	720	0.37	0.48	0.00	1.00

Note: NGO-aid is whether an NGO either owns a school/clinicor currently provides resources to the local school or clinic, or has contributed to the construction of public facilities; Years NGO-owned facilities refers to the number of years that a NGO has owned either a school or clinic in the village. Years NGO activity is the number of years a NGO has either owned a school or a clinic, or the number of years since a NGO contributed to the construction of the school or clinic, if a NGO provides current support to the school or clinic.

Table A.3: Giving by Game Type

	(1) Anon. Own-Village	(2) Non-Anon. Own-Village	(3) Anon. Other-Village
White-man	223.585**	118.994	183.908
	(102.443)	(103.127)	(110.449)
Ethnicity fe	Yes	Yes	Yes

Note: All specifications include district and ethnicity fixed effects. Robust standard errors clustered at the village level in parentheses. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.

Table A.4: Robustness to Controls

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving	(5) Total giving
White-man	579.169**	526.681*	957.060*	539.852*	548.941*
	(280.492)	(279.043)	(523.833)	(287.944)	(292.644)
Age	6.064				
	(6.619)				
Communal farm		-116.582			
		(408.673)			
Days from survey			0.461		
			(18.853)		
Days from survey x white-man			-9.728		
			(13.094)		
Linguistic group for ethnicity f.e.s	Primary	Primary	Primary	Parent-3	Parent-4
Observations	646	653	652	653	653

Note: Robust standard errors clustered at the village level in parentheses. All specifications include district and ethnicity fixed effects. Columns (1) to (3) use ethnicity fixed effects based on the primary local language. Columns (4) and (5) use alternative measures of ethnicity constructed from the 3rd (Parent-3) and 4th (Parent-4) highest parent group of the Atlantic language family. "Days from survey" indicates the time passed between conducting the household survey and administering the games. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.

Table A.5: Robustness to Samples

	(1) Total giving	(2) Total giving	(3) Total giving	(4) Total giving
White-man	540.103*	539.252*	551.338*	665.217**
	(293.055)	(306.420)	(297.739)	(291.990)
Ethnicity fe	Yes	Yes	Yes	Yes
Observations	652	620	601	564

Note: Column (1) drops one religious-contribution outlier; column (2) drops three village outliers in exposure to aid; column (3) drops observations that contain missing missing values in the control variables; column (4) drops the two districts— Bombali and Koinadugu — where linguistic diversity meant that more than one translator was used.

Table A.6: Giving Across Games

	(1) Giving	(2) Giving	(3) Giving	(4) Giving	(5) Giving
White-man	174.003*	180.287	204.015*	144.147	170.911*
	(98.905)	(112.248)	(102.948)	(102.087)	(98.164)
Game 1 x White-man		48.818		85.116	
		(67.544)		(57.743)	
Game 2 x White-man		-67.571	-91.253		
		(72.926)	(62.046)		
Game 3 x White-man					9.377
					(61.219)
Game 1		83.127*		-63.032*	
		(44.766)		(31.805)	
Game 2	145.972***	287.293***	245.003***		
	(35.291)	(52.319)	(38.874)		
Game 3	-107.481***				-185.210***
	(34.041)				(45.001)
Omitted Category	Game 1	Game 3	Games 1 and 3	Games 2 and 3	Games 1 and 2
Control group means					
Omitted category	955.556	872.126	914.548	1018.362	1057.639
Overall	997.191	997.191	997.191	997.191	997.191

Note: Game 1 is the Anonymous Own-Village dictator game; Game 2 is the Non-Anonymous Own-village dictator game; and Game 3 is the Anonymous Other-village game. The third-last row shows the comparison group (the omitted category) and the second-last and last rows show the control and treatment mean respective for the comparison group. Robust standard errors clustered at the village level in parentheses. ** is significant at the 5% level and * is significant at the 10% level.

Table A.7: Alternative Aid Meaures

	OLS		Logit	Multinomial Logit			Logit	Multinomial Logit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Total giving	Total giving	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research
Years NGO-owned facilities x white-man	-19.602**		1.055***	1.053***	1.056***	1.003				
	(9.434)		(0.019)	(0.019)	(0.021)	(0.699)				
Years NGO-owned facilities	3.995		0.971***	0.972**	0.965***	0.992				
	(6.209)		(0.11)	(0.010)	(0.012)	(0.003)				
NGO aid x white-man		-2,351.847***					11.135***	7.767***	41.016***	5.280***
		(793.006)					(5.447)	(3.982)	(25.335)	(2.217)
NGO aid		478.874					0.497*	0.552	0.238***	0.431***
		(625.908)					(0.201)	(0.228)	(0.110)	(0 .102)
White-man	654.858	1,488.102**	0.773	0.685	0.808	1.178	0.327	0.339	0.212	(0.423)
	(983.891)	(708.945)	(0.561)	(0.502)	(0.772)	(0.699)	(0.261)	(0.267)	(0 .234)	(0.797)
Individual-level controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Village-level controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	636	636	636	636	636	636	636	636	636	636

Notes. Robust standard errors clustered at the village level in parentheses. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level. See table 4 for notes and text for control variables

Table A.8: Aid exposure and Beliefs: Alternative Sample

	Logit	Multinomial Logit			Logit	Multinomial Logit		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research	Aid Test	Aid Test over Give Money	Aid Test over Research	Give Money over Research
White-man	0.235***	0.250***	0.187***	0.749	0.715	0.612	0.828	1.353
	(0.099)	(0.107)	(0.0813)	(0.178)	(0.490)	(0.434)	(0.721)	(0.797)
Years NGO activity x white-man	1.051**	1.051**	1.054*	1.004	1.068***	1.067**	1.072***	1.004
	(0.029)	(0.029)	(0.031)	(0.011)	(0.028)	(0.029)	(0.028)	(0.014)
Years NGO activity	0.966***	0.968*	0.961*	0.993	0.964*	0.964	0.96**	0.995
	(0.066)	(0.019)	(0.021)	(0.01)	(0.021)	(0.023)	(0.019)	(0.012)
Met white person 1 to 10 times x white-man					6.93	6.707	7.455	1.111
					(7.363)	(7.090)	(9.303)	(0.743)
Never met white person x white-man					7.721*	7.444*	8.511*	0.796
					(8.245)	(7.915)	(10.512)	(0.320)
Met white person 1 to 10 times					1.41	1.319	1.825	1.383
					(0.503)	(0.464)	(0.837)	(0.394)
Never met white person					0.64	0.618	0.747	1.207
					(0.580)	(0.555)	(.787)	(0.589)
Individual-level controls	N	N	N	N	Y	Y	Y	Y
Village-level controls	N	N	N	N	Y	Y	Y	Y
Observations	619	619	619	619	603	603	603	603

Notes. Robust standard errors clustered at the village level in parentheses. Columns 1-8 display the odds ratio (for multinomial logit, the relative risk ratio) of each outcome for a unit increase in the relevant independent variable (calculated as exp(β), where β is the log-odds ratio). Columns 1 and 5 display the odds ratio from a logit regression on the "Aid Test" indicator variable. Columns 2-4 and 6-8 display the relative risk ratios for each pair of choices from a multinomial logit regression on the categorical variable of participant beliefs over the 'aid test', 'give money' and 'research' choices. Individual-level and village-level controls are the same as those listed in Table 6. All specifications include district and ethnicity fixed effects. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.

Table A.9: Giving by Games: Interaction Effects

	Anonymous Own-Village Giving			Non-Anonymous Own-village Giving			Anonymous Other-village Giving		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
White-man	332.647***	335.473***	222.186**	176.940	274.976**	116.770	264.591**	296.306***	181.100
	(114.765)	(108.637)	(101.929)	(112.234)	(104.287)	(102.685)	(126.446)	(110.552)	(110.610)
Years NGO activity x white-man	-9.629***			-4.300			-7.372*		
	(3.265)			(4.723)			(3.905)		
Years NGO activity	2.550			2.614			1.575		
	(2.363)			(2.705)			(2.572)		
Customary authority x white-man		-418.130**			-600.176***			-423.202**	
		(170.315)			(166.569)			(177.366)	
Customary authority		314.049***			367.783***			284.079**	
		(103.427)			(98.633)			(111.262)	
Monetary contr. index			128.027			185.903*			46.475
			(101.682)			(109.239)			(76.871)
Monetary contr. index x white-man			-87.737			-179.137			-33.581
			(109.479)			(112.058)			(81.249)
Observations	720	720	719	720	720	719	708	708	707

Note: All specifications include district and ethnicity fixed effects. Robust standard errors clustered at the village level in parentheses. *** is significant at the 1% level, ** is significant at the 5% level and * is significant at the 10% level.