Lab Experiments in Developing Country Contexts*

Sara Lowes† — *UC San Diego, NBER, and BREAD*

Nathan Nunn‡ — *University of British Columbia and CIFAR*

9 June 2025

Abstract: Lab-in-the-field experiments, in which lab experiments are conducted in more naturalistic settings, are increasingly being implemented in developing country contexts. In this chapter, we outline the conceptual and logistical challenges typically associated with lab-in-the-field experiments in non-Western settings. We describe the importance of worldviews and how researcher preconceptions may inadvertently shape the types of research questions that are asked. We emphasize the importance of increasing diversity of subject pools and researchers. We also suggest a set of best practices when implementing lab-in-the-field experiments in developing countries.

Keywords: lab-in-the-field experiments, development, culture, experimenter demand

JEL Classification: C9, N01, N10, Z1

* Prepared for the *Handbook of Experimental Methodology*, edited by Eric Snowberg and Leeat Yariv. We thank Jeffrey Carpenter, Arkadev Ghosh, Joseph Henrich, Matt Lowe, and Daniel Posner for sharing photos. We thank Carla Colina Cortez, Max Ponce de León, Meghna Sinha Ray, Sebastian Rodriguez Raza, and Asma Tabassum for their excellent research assistance.

† UC San Diego, NBER, and BREAD. Email: slowes@ucsd.edu. Website: www.saralowes.com.

‡ University of British Columbia and CIFAR. Email: nathan.nunn@ubc.ca. Website: https://nathannunn.arts.ubc.ca.

1. Introduction

Lab-in-the-field experiments—lab experiments conducted in more "naturalistic" settings—are increasingly used by researchers in developing country contexts. However, conducting such experiments in these settings presents novel conceptual and logistical challenges. This chapter outlines the common obstacles researchers face when implementing lab-in-the-field experiments in developing countries and offers potential solutions.

We begin by tracing the shift from traditional laboratory experiments—typically conducted with university students in high-income countries—to lab-in-the-field approaches that emerged in the 1990s. This shift expanded the diversity of participants in experimental research and opened new opportunities to study behavior across a wider range of cultural and institutional contexts. In this section, we outline the motivations behind lab-in-the-field experiments, including their potential to improve representation of diverse populations and to explore how behaviors and decision making are shaped by different cultural, economic, and institutional environments. We also discuss the different taxonomies proposed to classify these experiments and provide illustrative examples from developing country contexts, highlighting the range of questions lab-in-the-field experiments have been used to address.

Next, we examine the conceptual and logistical challenges researchers face when conducting lab-in-the-field experiments, particularly in non-WEIRD (Western, Educated, Industrialized, Rich, and Democratic) contexts (Henrich, Heine and Norenzayan, 2010c). [1] We provide examples of how differences in worldviews between researchers and research subjects may lead to experiments that are not designed appropriately or that may miss important questions.

We also outline recommendations for best practices and considerations when implementing lab-in-the-field experiments in developing country contexts. While some of the challenges are common across research settings, many are specific to lab-in-the-field experiments in non-Western settings. We discuss various logistical considerations, such as site and participant selection, and challenges when working in environments with limited infrastructure. We also present strategies for improving the subjects' comprehension of the experimental protocols. Finally, we discuss the

[1] Throughout the chapter, we use terms such as developing countries, non-WEIRD, non-Western, and low-income settings to refer to the contexts in which many lab-in-the-field experiments are conducted. Each of these terms emphasizes a different dimension—economic status, geographic location, institutional development, or cultural distinctiveness—but they broadly overlap in referring to settings outside high-income, Western, industrialized societies. While these terms are not perfectly interchangeable, we use them more or less synonymously for the purposes of this chapter.

importance of experimenter demand effects and how to address them.

The chapter is organized as follows. Section 2 provides a historical overview and taxonomy of lab-in-the-field experiments. Section 3 explores how researcher worldviews shape experimental design. Section 4 addresses logistical challenges and best practices. Section 5 considers experimenter demand effects, and Section 6 concludes.

2. Experiments in Developing Countries

2.1. A Brief History

Lab experiments were initially conducted with university students in campus laboratories in developed countries. In practice, this meant that the populations participating in these experiments were not representative of the broader population. Participants were often young adults with a high level of formal education. Decisions were made in a tightly controlled lab environment—with students seated at separate desks or in cubicles, often using networked computers to make their choices. Additionally, participants had little prior knowledge of, or experience with, the types of decisions they were being asked to make. Until the 1990s, lab experiments were implemented almost exclusively with university students in Western countries. Early exceptions include Binswanger (1980), who measured risk aversion in rural India, and Roth, Prasnikar, Okuno-Fujiwara and Zamir (1991), who implemented the ultimatum game among university students in Israel, Slovenia, Japan, and the United States.

Starting in the mid-1990s, researchers increasingly recognized the need to move beyond university-based subject pools in Western countries and study more diverse populations. Increasing representation of different societies is crucial, since WEIRD cultures are global outliers, both historically and today (Henrich et al., 2010c). WEIRD populations tend to be more individualistic, more trusting of outgroup members, and more reliant on analytic—as opposed to holistic—thinking. Given that these populations are neither typical nor representative of most of the world, it is particularly important to draw on culturally diverse samples.

There are several early examples of bringing the lab to the field. In 1994, Joseph Henrich (2000) implemented the ultimatum game—a two-player bargaining game in which one participant proposes a division of an endowment and the other can accept or reject the offer— among the Machiguenga, a small-scale, pre-industrial, horticultural Indigenous group living in the tropical

forests of the Peruvian Amazon. This line of research was later extended by Henrich and coauthors to include 15 small-scale societies from around the world (Henrich, Boyd, Bowles, Camerer, Gintis, McElreath and Fehr, 2001). A striking finding from this analysis is that the less-developed societies were those that engaged in behavior that was more "rational," defined as behavior consistent with the Nash equilibrium of the game. Unlike the near-equal endowment splits observed among university students, participants in these settings tended to offer smaller amounts, which were often accepted.

Another example of bringing the lab to the field comes from Cardenas, Stranlund and Willis (2000) and Cardenas (2000), who implemented versions of a public goods game in villages across different regions of Colombia during the summer of 1998. The experiments were framed around a natural resource dilemma, in which participants decided how much to extract from a shared forest resource, with higher deforestation reducing water quality. The authors examined how various forms of regulation affected cooperative behavior, and found that external rules could crowd out group-oriented action by undermining intrinsic motivation.

Over the past several decades, the locations, participant populations, and characteristics of traditional lab experiments have moved beyond behavioral tasks conducted in physical labs with university students in industrialized countries. There are now many examples of lab-in-the-field experiments conducted in developing country contexts. While we do not intend to provide a comprehensive overview of this literature, we briefly list some of the key topics that have been studied using such experiments.[2]

Lab-in-the-field experiments have been used to measure a range of outcomes, including intra-household cooperation in couples,[3] including polygamous couples[4] and by kinship structure;[5] learning in the household and from others;[6] time preferences;[7] risk sharing;[8] risk taking and

[2] For literature reviews of lab-in-the-field experiments in developing countries, please refer to Cardenas and Carpenter (2008); for behavioral development economics, see Kremer, Rao and Schilbach (2019). For an overview of lab experiments in historical economics, see Lowes (2021a).

[3] See Ashraf (2009), Iversen, Jackson, Kebede, Munro and Verschoor (2011), Kebede, Tarazona-Gomez, Munro and Verschoor (2013), Munro, Kebede, Tarazona-Gomez and Verschoor (2014), Afzal, d'Adda, Said and Kebede (2022).

[4] Barr, Dekker, Janssens, Kebede and Kramer (2018).

[5] Lowes (2022).

[6] Conlon, Mani, Rao, Ridley and Schilbach (2021, 2022)

[7] Balakrishnan, Haushofer and Jakiela (2020)

[8] Attanasio, Barr, Cardenas, Genicot and Meghir (2012), Barr, Dekker and Fafchamps (2012b), Barr and Genicot (2008).

Figure 1: Examples of Lab-in-Field

(a) Lab-in-field among the Au of Papua New Guinea from Henrich et al. (2005).

(b) Lab-in-field among the Samburu of Kenya from Henrich et al. (2010a).

preferences;[9] adverse selection.[10]

Other topics include trust;[11] respect for earned income;[12] fairness and inequality aversion;[13] redistribution;[14] punishment;[15] group formation; [16] microfinance behavior;[17] and competitiveness.[18] Researchers have also studied in-group preferences,[19] including coethnicity in Africa,[20] co-religion,[21] and caste in India.[22] Table 2 provides examples of lab-in-the-field experiments.

There are several reasons for the growth in lab-in-the-field experiments in developing countries. In some fields, and for some researchers, there is a presumption that the traits being examined in the lab are universal, with little or no heterogeneity across populations. If this assumption

[9] Fafchamps, Kebede and Zizzo (2015), Jakiela and Ozier (2019a), Negussie, van Ierland, Zhu, Kebede and Klopm (2018).

[10] Barboni, Cassar, Rodriguez Trejo and Wydick (2013).

[11] Barr (2003), Carpenter, Daniere and Takahashi (2004a), Cassar, Grosjean and Whitt (2013), Carpenter, Daniere and Takahashi (2004b), Cassar, Healy and Von Kessler (2017).

[12] Jakiela (2011), Jakiela, Miguel and te Velde (2015), Barr, Burns, Miller and Shaw (2015).

[13] Jakiela (2015), Kebede and Zizzo (2015).

[14] Jakiela and Ozier (2016).

[15] Henrich, McElreath, Barr, Ensminger, Barrett, Bolyanatz, Cardenas, Gurven, Gwako, Henrich, Lesorogol, Marlowe, Tracer and Ziker (2006a), Marlowe, Berbesque, Barr, Barrett, Bolyanatz, Cardenas, Ensminger, Gurven, Gwako, Henrich, Henrich, Lesorogol, McElreath, Tracer and Ziker (2008).

[16] Barr, Dekker and Kebede (2012a).

[17] Cassar, Crowley and Wydick (2007), Giné, Jakiela, Karlan and Morduch (2010).

[18] Gneezy, Leonard and List (2009), Flory, Gneezy, Leonard and List (2017), Lowes (2021b), Cassar and Zhang (2022), Cassar, Grosjean, Khan and Lambert (2023).

[19] Bauer, Cassar, Chytilová and Henrich (2014a), Bauer, Blattman, Chytilová, Henrich, Miguel and Mitts (2016), Rao (2019).

[20] Habyarimana, Humphreys, Posner and Weinstein (2007), Berge, Bjorvatn, Galle, Miguel, Posner, Tungodden and Zhang (2020).

[21] Bauer, Chytilová and Ochieng (2024)

[22] Brooks, Hoff and Pandey (2018), Hoff, Kshetramade and Fehr (2011).

holds, then convenience samples—such as university students—are a natural choice. However, there is increasing recognition that the behaviors observed in the lab among university students in industrialized countries are unlikely to generalize across contexts. This growing awareness has motivated efforts to study and understand behavior in a broader range of populations.

This recognition of cross-societal variation has coincided with broader methodological changes in development economics. Over the past several decades, alongside the rise of randomized controlled trials, there has been growing interest in measurement tools beyond traditional surveys, including various types of experiments (Kremer et al., 2019). These developments reflect a shift in the profession toward better capturing behavioral variation and contextual heterogeneity across different populations.

In sum, lab-in-the-field experiments emerged in response to concerns about the external validity of traditional lab studies and growing interest in understanding behavior across a broader range of cultural and institutional settings. They have since become an important tool for studying diverse populations and testing the generalizability of behavioral findings.

2.2. Experiment Taxonomy

Researchers have proposed various classification schemes for lab experiments conducted outside traditional laboratory settings. These taxonomies help clarify how different experimental designs vary in terms of context, participant pools, and their implications for internal and external validity.

Broadly, experiments that do not occur in the lab with university students are referred to as *field experiments* in the economics profession. Experiments using standard laboratory protocols but with non-student populations have been labeled *artefactual field experiments* by Harrison and List (2004), and *extra-laboratory experiments* by Charness, Gneezy and Kuhn (2013). Examples include Binswanger (1980), who elicited measures of risk aversion among farmers in rural India, and Carpenter, Burks and Verhoogen (2005), who compared behavior in the ultimatum game between university students and employees at a Kansas City distribution center. The key distinction in these classifications is that the experimental protocol remains the same, but the subject pool has shifted away from university students.

Beyond the participant pool, experiments conducted outside traditional lab settings—particularly in developing countries—may also differ from standard lab experiments in

terms of the realism or context embedded in the experimental task. These experiments may incorporate actions, stakes, or information that participants naturally encounter in their everyday environments. In some cases, they are explicitly designed around naturally occurring variation relevant to the research question. For example, Rao (2019) develops experiments in Delhi schools that, in addition to using a standard dictator game to measure children's prosocial preferences, include a task where students choose teammates for a relay race —a choice more familiar and natural for children in this context. He also examines how a real-world government policy that integrates poorer and richer students within schools affects prosocial behavior between the groups. This class of experiments, labeled *framed field experiments* by Harrison and List (2008), has grown in prominence—particularly in development economics.

A final class of field experiments involves participants who are unaware they are part of an experiment. An example is the wallet-drop study by Cohn, Marechal, Tannenbaum and Zund (2019), in which 17,303 wallets were dropped in 355 cities across 40 countries to measure civic honesty. The authors recorded the rate at which lost wallets were returned and compared return rates within each country for wallets that did or did not contain money. In these experiments, individuals are observed in their typical environment without knowing their behavior is being studied. Harrison and List (2008) label this type of design a *natural field experiment*.

While these taxonomies are not always intuitive or exhaustive, they provide a useful heuristic for distinguishing between experimental designs along key dimensions such as subject pool, task framing, and participants' awareness. They help clarify the trade-offs between experimental control and contextual realism, and offer a shared language for discussing lab-in-the-field methods.

2.3. Benefits of Lab-in-the-Field Experiments

While we will discuss the challenges and best practices of lab-in-the-field experiments later in the chapter, here we outline their key advantages. These include improved external validity, access to diverse and policy-relevant populations, the ability to study context-specific or otherwise infeasible questions, and opportunities for longitudinal or immersive designs. Conducting experiments outside traditional labs introduces a range of logistical and conceptual difficulties—but these same differences also offer unique benefits.

Lab-in-the-field experiments preserve many advantages of conventional lab studies, including control over the decision environment and the use of precise, often incentivized, measures of

preferences or behaviors. At the same time, they allow researchers to engage more diverse and potentially more policy-relevant populations. In some cases, they also enable sampling strategies that take advantage of naturally occurring variation in the field, while retaining control over the environment. For example, variation in kinship structures informed the design of experiments by Gneezy et al. (2009), Flory et al. (2017), and Lowes (2022), while historical variation in state formation motivated the work of Lowes, Nunn, Robinson and Weigel (2017) and Heldring (2021).

Lab-in-the-field experiments are often essential when researchers aim to study questions that are specific to a particular population or cultural context. For instance, Jakiela and Ozier (2016) examine how social pressure to redistribute income—an issue that is less salient in WEIRD societies—affects investment behavior in rural Kenya. They design an experiment in which individuals can pay to conceal their experimental earnings from others in the community, thereby avoiding redistribution. Their findings show that women are willing to forgo expected earnings to keep their endowment private. In another example, Bursztyn, González and Yanagizawa-Drott (2020) study barriers to female labor force participation in Saudi Arabia, focusing on misperceived social norms. They elicit husbands' beliefs about women working outside the home, as well as their perceptions of what other men believe. Many participants underestimate support for women's employment among peers. When randomly provided with correct information, men become more likely to sign their wives up for a job search platform.

Another example is Berge et al. (2020), who investigate the role of ethnic preferences in Africa. They implemented a series of experimental games—including dictator games, public goods games, and "choose-your-dictator" games—with participants in Nairobi, Kenya, at the Busara Center for Behavioral Economics.[23] The ethnic identity of the paired participants was randomly varied across sessions. They found little evidence of coethnic bias in their sample of 1,300 participants. In this case, the broader location of the experiment—Kenya—was determined by the research question, while the specific site—the Busara lab in Nairobi—was chosen based on existing infrastructure. In all three examples, the research questions were tightly tied to the local context, illustrating how lab-in-the-field experiments are not just a methodological choice, but a necessity for studying certain questions that would be ill-suited to conventional lab settings.

[23] In a dictator game, one participant unilaterally decides how to split a sum of money with another individual. A public goods game involves participants contributing to a common pool that benefits the group, often revealing prosocial or free-riding behavior. In a choose-your-dictator game, participants select who among a set of individuals will be the dictator in a dictator game where the participant is the recipient.

Beyond answering contextually relevant questions, field experiments in developing countries make it possible to test whether behaviors observed in lab settings vary systematically across societies—and, if so, to identify the drivers of that variation. As such, a core motivation behind many of these experiments is to develop a broader and more cross-culturally informed understanding of human behavior. As we will see, this aim was central to some of the earliest lab-in-the-field studies, including Henrich et al. (2001), who implemented the ultimatum game across 15 small-scale societies.

As we discuss in more detail later in the chapter, researchers conducting field experiments typically view the composition of the participant pool as crucial. As a result, they aim to recruit participants who are representative of the population of interest. In our experience, field experiments in developing country contexts can be designed with highly representative samples and extremely low attrition rates. Consequently, such experiments often suffer less from selection bias than conventional lab experiments.

Another often-cited benefit of lab-in-the-field experiments is that some studies may only be logistically or financially feasible in certain settings. Many such experiments address universal topics but strategically select their location or design based on what is practically implementable. For example, Bursztyn, Ferman, Fiorin, Kanz and Rao (2018) study social image concerns and the value of status goods through an experiment on demand for platinum credit cards in Indonesia. The authors collaborated with an Indonesian bank, which made this setting particularly suitable. Similarly, Breza, Kaur and Shamdasani (2018) examine the morale effects of pay inequality by hiring workers at a manufacturing firm in India and varying both pay inequality and the observability of co-worker productivity. This study—while addressing a broadly relevant question—would have been prohibitively expensive or difficult to conduct in many other contexts.

Beyond savings in terms of costs and logistics, another benefit of field experiments is that they allow for data collection over extended time frames—a key advantage when there is meaningful variation in the environment over time. One example is Jakiela and Ozier (2019b), who examine how risk preferences changed before and after the 2007 election in Kenya. They find that post-election instability led to an increase in risk aversion.

Working in the field also allows researchers to conduct experiments in more elaborate, context-specific environments that are intentionally designed or manipulated by the experimenter. These experiments, while controlled, can occur in real-life settings, closely resemble actual decision-

making contexts, and in some cases, participants may not even know they are part of an experiment. One example is Casey, Glennerster and Miguel (2012), who conducted a study in 236 villages in Sierra Leone to evaluate an intervention aimed at altering local institutions. The authors used three experiment-based measures, each presenting participants with artificial but realistic decisions.

To assess each village's ability to engage in collective action, they provided six vouchers potentially worth US $17 each, redeemable only if the community could collectively raise double that amount. The funds had to be used to purchase building materials at a specific store. In the second experiment, a community meeting was convened where villagers were asked to choose between two gifts: a carton of batteries or small bags of iodized salt. The discussion was recorded and analyzed to assess how democratic and inclusive the decision-making process was. In the third experiment, designed to measure elite capture, a large plastic tarpaulin sheet was given to the village, and researchers recorded how it was used when they returned five months later.

Experiments of this kind have become increasingly common and more elaborate. For example, Weigel (2020) studied whether taxation increases government accountability. In the Democratic Republic of the Congo (DRC), participants were given evaluation forms to complete and deposit in a lockbox placed in the city center; these forms were later delivered to the provincial governor. In a second experiment, five town hall meetings were held with the finance minister and the director general of the tax ministry. The behaviors of interest were whether participants submitted evaluation forms or attended the meetings.

In some instances, experiment-like environments occur naturally, and researchers are able to take advantage of them. Although these are arguably not experiments in the strictest sense—since the setting is not created by the researcher—they can nonetheless provide valuable quasi-experimental variation. Hjort (2014) uses data from a flower factory in Kenya that operates with team-based production. Individuals are effectively randomly assigned to teams, generating natural variation in the ethnic composition of workgroups. He exploits this setting to study the impact of ethnic diversity on team productivity.

Midway through the study period, the factory's pay structure shifted to depend more heavily on team output, thereby altering incentives for cooperation. Hjort then estimates how this change influenced the relationship between team diversity and productivity. Conceptually, the key distinction between Breza et al. (2018) and Hjort (2014) is that the former constructed a factory

environment to study their question, while the latter leveraged an existing one.

In summary, lab-in-the-field experiments offer several benefits. They retain a degree of experimental control, as in conventional lab settings, but allow researchers to engage with more diverse and contextually relevant populations than university students. They can also take advantage of naturally occurring variation and quasi-experimental settings. For some researchers, an additional benefit is feasibility: certain experiments that would be prohibitively costly or logistically complex in high-income countries may be implementable in developing country contexts—for example, setting up a factory as an experimental environment.

3. Alternative realities and worldviews

3.1. Researcher assumptions and experimental design

One of the most important issues for researchers to consider when implementing experiments in developing countries is that any experimental design—or data collection exercise—involves a host of assumptions made by the researcher, often implicitly and without conscious recognition. Within the researcher's home-country context, this is less of an issue because their assumptions or ways of thinking are often aligned with those of the participants. This shared worldview can make the experiment run more smoothly and increase the likelihood that participants understand the task as intended.

However, when there are differences in ways of thinking—or in the perceived importance or meaning of various aspects of the experiment (e.g., the wording of protocols, physical setup, or logistical procedures)—there is a greater likelihood that participant behavior will be shaped by factors unknown to the researcher. In short, assumptions that researchers make—often subconsciously—that may hold in developed-country contexts might not translate to developing-country settings. This underscores the importance of having a diverse research team, including members who are deeply familiar with the local context.

To illustrate, one example we encountered early in our survey work in the Democratic Republic of the Congo involves a standard question: "Is success primarily determined by hard work or luck?" Of course, this question presumes that luck is a well-defined and universally understood concept. Yet in many non-Western contexts—including among our Congolese participants—it is not. Instead, the world is often seen as a fully deterministic system in which individuals are

deeply interconnected, and the actions of one part are influenced by all others (Tempels, 1959, Xiaotong, 1992). For instance, Xiaotong (1992) describes Chinese society as "webs woven out of countless personal relationships," with its dynamics "represented by ripples flowing out from the splash of a rock thrown into water." Similar imagery appears in other non-WEIRD contexts: Kanu (2023) describes Igbo society as a "spider's web of which no single thread can be caused to vibrate without shaking the whole network."[24] Thus, within this worldview, the concept of 'luck' or 'randomness' is not natural.

These differing worldviews also affect how researchers should approach the design and interpretation of measurement tools, including standard instruments for mapping social networks. For example, we have found that in the rural DRC, standard questions aimed at measuring a person's network could be challenging to implement because of how individuals conceive of their social relationships. Similarly, we followed protocols developed and implemented in India (Breza, Chandrasekhar, McCormick and Pan, 2020, Breza, Chandrasekhar, Lubold, McCormick and Pan, 2023). However, we encountered the challenge that respondents view everyone in their village as part of their 'network'. If you ask them to list everyone in their network (elicited through a series of questions on who they do certain activities with), then the list is much longer than in more WEIRD cultures. Only by emphasizing that we were interested in individuals with whom they had frequent social interaction were we able to narrow the lists to fewer than 20 names.

Relatedly, Griffith (2022) shows that censored social network elicitation methods—which cap the number of friends a respondent can name, often at five—can attenuate estimated peer effects. This approach is likely more problematic in societies with denser or more interconnected social structures, where such limits are more likely to bind.

Another example comes from the standard ultimatum game. Often, when implementing the UG, researchers determine player 2's action by identifying their minimum acceptable offer. This is done by asking them to state the minimum payment they would accept when playing the role of player 2 (i.e., recipient) in the ultimatum game. This way of eliciting player 2's strategy was used by Güth, Schmittberger and Schwarze (1982), who were the first economists to implement the game (Thaler, 1988). This presumes that one's (latent) desire to accept (rather than reject) is monotonically increasing in the amount of money offered to player 2 by player 1 in the game.

[24] This is conceptually related to variation in analytic versus holistic thinking (Nisbett, 2004). Non-WEIRD cultures tend to see the world as made up of relationships, while WEIRD cultures tend to view it as composed of discrete individuals (Xiaotong, 1992).

While this assumption may be reasonable in WEIRD contexts—such as the original sample of German graduate students—it does not always hold in non-Western settings. In several non-WEIRD contexts, willingness to accept an offer appears to be hump-shaped: participants tend to reject offers that are too low, but also offers that are too high. This suggests that the meaning and interpretation of generous offers may differ cross-culturally, with very high offers possibly perceived as insulting, suspicious, or socially inappropriate.

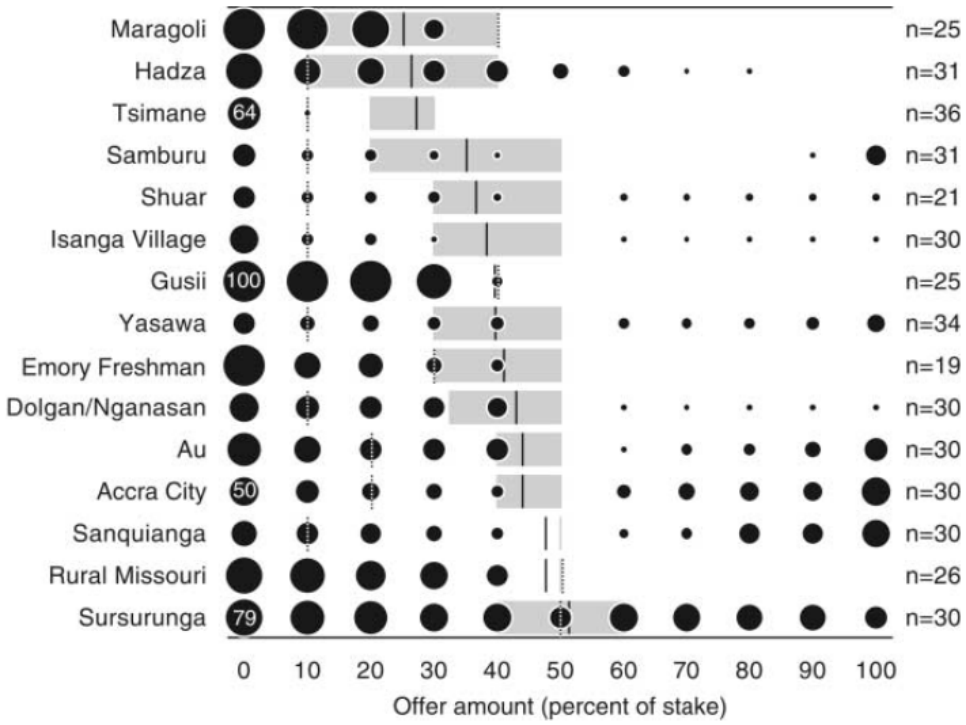
This is evident in Figure 2, which is from Henrich et al. (2006a). The figure presents rejection rates by share offered in the ultimatum game across 15 societies. It shows that for some groups, including the two WEIRD populations (Emory freshman and rural Missouri), rejection rates decline monotonically in the amount offered to player 2. However, for most groups studied, the relationship is non-monotonic. Large offers, such as 90 or 100 percent of the endowment, are commonly rejected. More generally, for offers above 50%, rejection rates tend to rise with the size of the offer, suggesting that extreme generosity can also be perceived negatively in many cultural contexts.

We found similar patterns in the DRC, where we collected ultimatum game data from 570 participants across 10 ethnic groups. Respondents reported their willingness to accept each possible offer from the other player, ranging from 0 CF to 1,000 CF (Congolese Francs). Each participant completed the game three times—with a partner from their own ethnic group, a partner from a different ethnic group, and a randomly selected citizen of Kananga, the city where data collection took place. The mean rejection rates by amount offered, disaggregated by ethnic group, are shown in Figure 3. Rejection rates decline as offers increase up to 500 CF (half the endowment). But, as in Henrich et al. (2006a), larger offers are often rejected. Offers above 500 CF were rejected between 13% and 26% of the time, with rejection rates peaking at 26% when the full endowment was offered to player 2. Notably, without implementing the strategy method to elicit player 2 responses across all possible offers, this pattern would likely go unobserved—especially if large offers are rare in actual play.

Although the reason for this non-monotonicity is unclear—and likely varies by context— exit questions from ultimatum game experiments in the DRC offer some insights. After playing several rounds of the UG, Lowes et al. (2017) asked participants:"How would you feel if the other player offered you 1,000 CF?"[25] While many participants responded that they would be happy

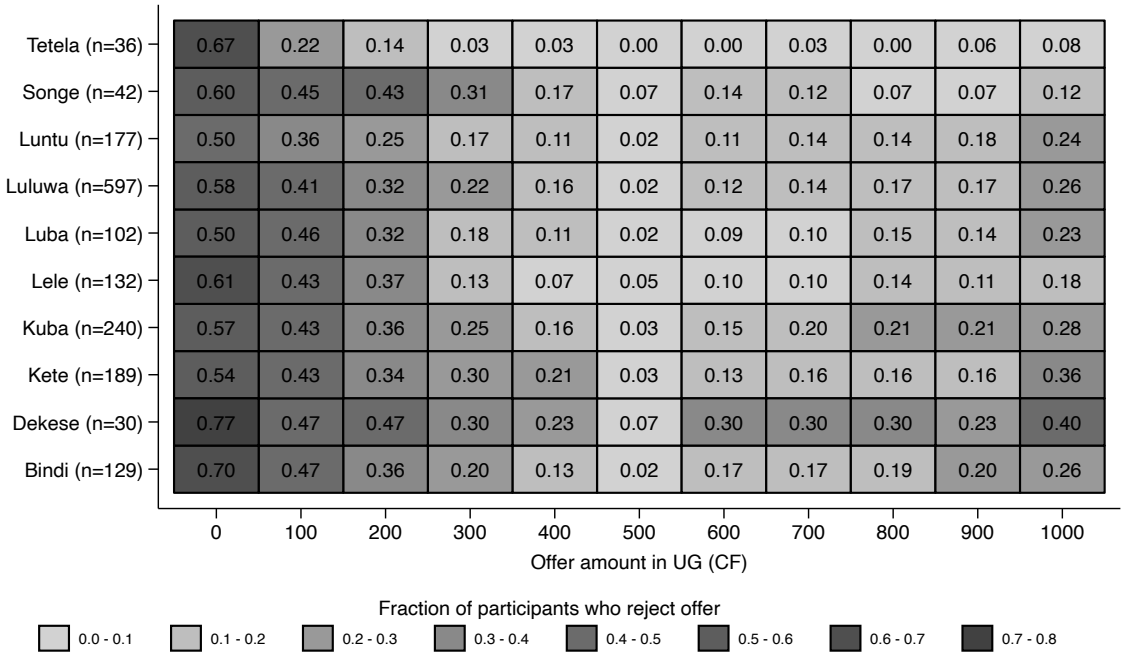
[25] All questions and responses reported here have been translated from either French or Tshiluba to English.

Figure 2: UG rejections across possible offers in Henrich et al. (2006a)



Notes: This is Figure 1 from Henrich et al. (2006a). UG results are displayed as the distributions of rejections across possible offers in the UG, which overlay the mean offers and interquartile. For each population labeled along the vertical axis, the areas of the black bubbles, reading horizontally, show the fraction of the sample of player 2s who were willing to reject that offer. The dashed vertical bars mark the IMO for each population. The solid vertical bars mark the mean offer for each population, with the gray-shaded rectangle highlighting the interquartile of offers. Populations were ordered by their mean offers (from low to high). Counts on the right (n) refer to the number of pairs of players.

Figure 3: UG rejections by offer amount using data from Lowes et al. (2017).



Notes: The figure presents the rate of player 2 rejections for each amount offered by player 1 in the ultimatum game (UG) implemented by Lowes et al. (2017). The UG rejection decisions were recorded using the strategy method – so that a willingness to accept or reject is recorded for each possible amount a participant could be offered.

if this happened, a substantial number reported that they would "not be happy" or would be "worried," "uncomfortable," "sad," "even worse than before," etc. There appear to be multiple reasons for these negative emotions. Some participants expressed discomfort because the other player would receive nothing. Others reported suspicion toward someone offering the entire amount—potentially linked to concerns about witchcraft or excessive social obligations, both of which are salient in the region.

The local context may also shape the types of incentives that individuals respond to in an experiment. For example, Cassar, Wordofa and Zhang (2016) explore the gender gap in willingness to compete. A large literature finds that women are less willing to compete than men, a result that has been replicated across many contexts. These papers generally implement a standard willingness to compete protocol with monetary incentives. However, Cassar et al. (2016) found that when incentives are framed to benefit participants' children rather than themselves, the gender gap in willingness to compete disappears.

This is an example of the broader issue that behavior can be highly contingent on the form that payoffs take—and the nature of this contingency may differ across societies. Medvedev, Davenport, Talhelm et al. (2024) find that the motivating effect of monetary rewards (compared to purely psychological motivators) varies systematically across cultures. Although their study includes only six countries (the US, UK, China, India, Mexico, and South Africa), the results suggest that individuals in less-developed, non-WEIRD societies are less responsive to monetary incentives than to psychological ones. Paradoxically, money—which is relatively scarce in low-income settings—appears to be a weaker motivator in these contexts (Medvedev et al., 2024).

3.2. Research assumptions and the questions we ask

In this section, we highlight how researcher assumptions can shape not only experimental design and implementation but also the research agenda itself. Ever since the seminal paper by Easterly and Levine (1997), inter-ethnic relations have been of great interest to researchers working in Africa. This interest has given rise to a large observational and experimental literature on how ethnic identity affects various forms of behavior, including behavior in lab experiments. In a recent study, Haushofer, Lowes, Musau, Ndetei, Nunn, Poll and Qian (2023) explored how coethnicity and stress influence prosocial behavior across a range of experimental tasks. To mask the study's focus on coethnicity—and to make the decisions feel more realistic to participants—the authors

presented individuals with profiles of their counterpart. These profiles included information on the other player's mother tongue (a proxy for ethnic identity), age group, and gender. The paper finds strong coethnicity effects: individuals behave more prosocially toward coethnics, and this behavior is not mediated by stress.

Haushofer et al. (2023) followed a well-developed literature that focuses on the importance of ethnic identity for behavior and economic development more generally (e.g., Habyarimana et al., 2007, Habyarimana, Humphreys, Posner and Weinstein, 2009, Lowes, Nunn, Robinson and Weigel, 2015, Berge et al., 2020, Blum, Hazlett and Posner, 2021). In some sense, the economic focus on ethnicity stands in contrast with the reality that in most parts of Africa, modern ethnic boundaries are artificially created after European contact. Arguably much more relevant are family, lineage, and other kinship structures (see for example Moscona, Nunn and Robinson, 2017, 2020, Lowes, 2022), as well as cross-cutting associations like age sets (Moscona and Seck, forthcoming).

Because the experiment in Haushofer et al. (2023) also randomly varies other aspects of the matched player's identity, the data allow us to examine how additional identity dimensions shape behavior. In many lineage-based societies, for instance, deference to elders—who often hold authority and power—is a social norm. However, this can be offset in societies with age-set systems, where obligations are directed toward peers in the same age grade, thereby diluting hierarchical deference to older individuals (e.g., Bernardi, 1985).

Drawing on this ethnographic distinction, we use the data from Haushofer et al. (2023) to examine how older individuals are treated in these games. Specifically, we ask whether generosity toward older players varies by the gender of the matched pair, and whether it depends on whether the decision-maker comes from an ethnic group with an age-set tradition. Among the four ethnic groups in the sample, two—Kamba and Kikuyu—traditionally have age sets, while the Luo and Luhya do not.

We report these estimates in Table 1. As reported in panel A, we find that individuals are more prosocial towards older partners. This is despite the fact that participants do not feel closer to older individuals (as measured by survey questions). This effect is found clearly for observations where a male decision-maker is playing with a matched player man (panel B) but is not found when a female decision-maker is playing with a matched woman.

Further analysis (not reported here) indicates that this pattern is driven by the Luo and Luhya,

the two ethnic groups in the sample without traditional age sets, where lineage structures and deference to elders remain particularly salient. Notably, the magnitude of the "older partner" effect is similar to that of the coethnicity effect. Yet while coethnicity has received extensive scholarly attention, the influence of gerontocracy and age-based hierarchies remains largely unexamined—likely because foreign researchers have historically prioritized ethnic identity over other culturally relevant social structures.

Another example is highlighted in the study by Godoy, Karlan and Zinman (2021), where a field experiment was implemented among 1,094 individuals (from 61 villages) belonging to a small-scale horticultural group called the Tsimane living in the Bolivian Amazon. They randomized the allocation of small wooden savings lockboxes with and without keys (to the treatment groups) and plates of a similar value (to the control group). The design was motivated by the researchers' priors, which were based on studies that took place among populations that were much less small-scale and remotely located and a presumption that outcomes could be improved if savings rates were increased.

From a WEIRD perspective characterized by analytic thinking, where the world is (subconsciously) viewed as being comprised of primarily disconnected agents, a natural potential explanation for the poverty of a person or group is that their savings may be too low. Within this framework, it's reasonable to hypothesize that alcohol may be a temptation good that is an alternative to saving for the future. Thus, within the typical WEIRD framework, the experiment makes perfect sense. However, as Godoy et al. (2021) show, the reality of the situation is very different. The world of the Tsimane is centered around relationships and connected individuals. In Tsimane society, one of the few things that is purchased with cash on the market is hard alcohol, which is consumed collectively and provides prestige to the person providing it to the group. The study finds that consistent with this reality, but against the priors of the researchers who were initially only familiar with the society through written ethnographic sources, the only effects of the lockboxes (beyond an increase in cash holdings at home) was an increase in alcohol consumption and (high) blood pressure. They found no effect on physical assets, agricultural investments, or changes in income.

As the authors explain, the history of their study provides a telling example of how implicit assumptions from a WEIRD society cause researchers to choose hypotheses and research questions that do not fit the reality of the society being studied. For the researcher considering a

Table 1: Examining deference to the elderly using the data from Haushofer et al. (2023).

	Choose Your Dictator	Dictator Game	Trust Game P1	Trust Game P2	Social Proximity	Likely to be friends	Trust	Closeness
				Panel A. Full Sample: All Observations				
Matched player is older	0.0442	0.0197***	0.0229***	0.0102**	-0.0343***	-0.2135***	-0.0687**	-0.1625***
	(0.030)	(0.005)	(0.006)	(0.005)	(0.007)	(0.031)	(0.032)	(0.038)
Coethnic	0.4015***	0.0187***	0.0293***	0.0016	0.0821***	0.3085***	0.3632***	0.3920***
	(0.039)	(0.004)	(0.004)	(0.002)	(0.005)	(0.025)	(0.027)	(0.031)
Observations	21,408	10,704	10,704	53,520	7,136	7,136	7,136	7,136
				Panel B. Restricted Sample: Men Playing with Men				
Matched player is older	0.0421	0.0198**	0.0333***	0.0094	-0.0259**	-0.1937***	-0.0409	-0.0949
	(0.057)	(0.010)	(0.011)	(0.009)	(0.012)	(0.057)	(0.058)	(0.070)
Coethnic	0.3842***	0.0278***	0.0494***	0.0030	0.0847***	0.3047***	0.3518***	0.4498***
	(0.068)	(0.009)	(0.010)	(0.007)	(0.012)	(0.057)	(0.061)	(0.067)
Observations	5,780	2,900	2,880	14,325	1,940	1,940	1,940	1,940
				Panel C. Restricted Sample: Women Playing with Women				
Matched player is older	-0.0291	0.0115	0.0164	-0.0052	-0.0192	-0.1248*	0.0113	-0.1458*
	(0.064)	(0.010)	(0.012)	(0.009)	(0.014)	(0.064)	(0.069)	(0.082)
Coethnic	0.4364***	0.0168*	0.0367***	0.0047	0.0841***	0.2942***	0.4133***	0.3778***
	(0.073)	(0.009)	(0.011)	(0.008)	(0.013)	(0.060)	(0.064)	(0.075)
Observations	4,925	2,482	2,483	12,365	1,690	1,690	1,690	1,690

Notes: The unit of observation is a decision in a round of a game. All regressions include an indicator for the gender of the decision maker being the same as the matched player; decision maker ethnicity, gender, and age group fixed effects; and matched player gender and ethnicity fixed effects.

research question and a particular location, it is important to consider whether the setting is appropriate. As the title of Godoy et al. (2021) suggests ("Randomization for Causality, Ethnography for Mechanisms: Illiquid Savings for Liquor in an Autarkic Society"), the procedure they employed was to implement the experiment first and then, after the estimates were obtained, used ethnographic methods to try to understand them. However, our view is that it is more efficient, ethically more defensible, and intellectually optimal for researchers to first learn about the setting through first-hand experience (e.g., living in the location, focus groups, qualitative research, etc.) and to use the knowledge gained to develop the research question and experimental design. An important part of this being successful will be for researchers unfamiliar with the setting to question the implicit assumptions they might be making and to collaborate with coauthors who are more familiar with the local context.

These examples underscore a broader point: the assumptions that researchers bring to their work—often shaped by their own cultural backgrounds—can influence not only how experiments are implemented but also which questions are considered worth asking. In cross-cultural settings, particularly in developing countries, such assumptions may misrepresent local priorities, social structures, or motivations. Recognizing this, researchers must engage deeply with the context they study—not just to avoid design flaws, but to ensure that their work meaningfully captures local realities.

4. Logistical considerations

All lab-in-the-field experiments require well-organized logistics to ensure smooth implementation. However, novel issues arise when implementing lab experiments in developing countries. We discuss representation and site selection; participant selection; infrastructure; ensuring comprehension; other logistical challenges researchers may face; and ethical considerations.

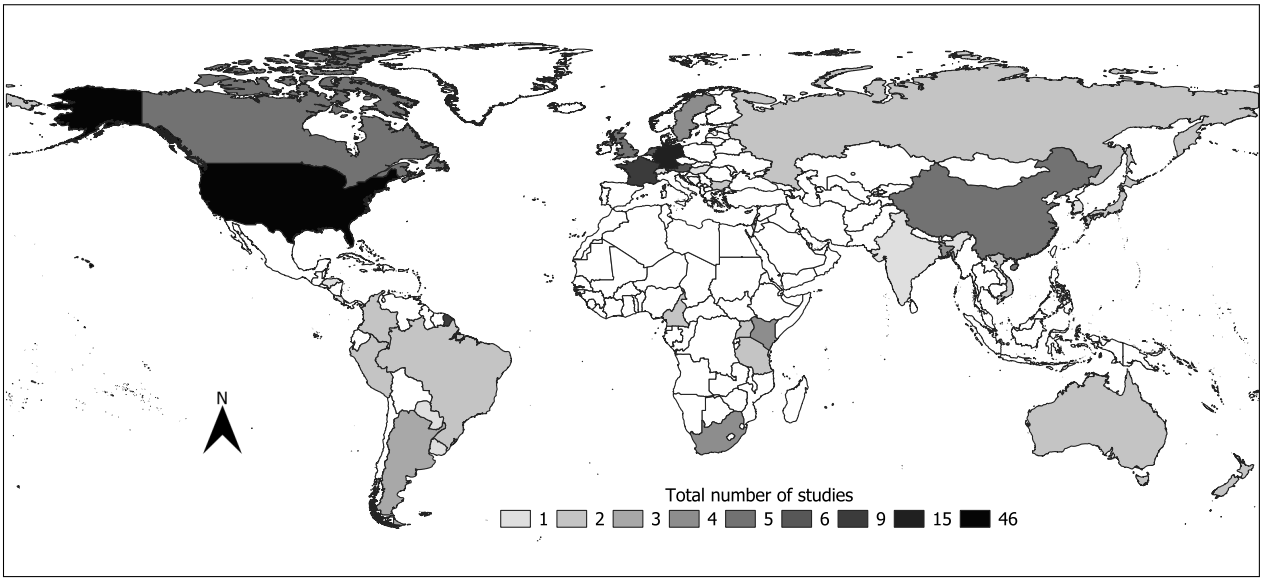
4.1. Site selection

The limited presence of existing lab and data collection infrastructure has important implications for where scholars undertake research. Consequently, there is a bias towards conducting research in locations with pre-existing data collection infrastructure. For instance, many randomized controlled trials are conducted in Kenya, Uganda, or India because of the infrastructure that has been set up there through JPAL or IPA. One of the most well-used lab spaces to implement lab experiments in Africa is the Busara Center for Behavioral Economics, with headquarters in Nairobi, Kenya. The Busara Center has lab space in Nairobi and a participant pool recruited from nearby communities. In addition, they work across sites in Kenya, Ethiopia, Nigeria, and Uganda (Busara Lab, 2024).

A similar issue arises with lab-in-the-field experiments more generally. Figure 4 presents the total number of trust-game experiments (Berg, Dickhaut and McCabe, 1995) implemented in each country in the meta-analysis sample of Johnson and Mislin (2011). As of this date, there are no published trust experiments implemented in much of the Middle East and Africa. Within Africa, the experiments took place in the less representative but easier-to-work-in countries, such as South Africa and Kenya, while the continent's largest countries, like Nigeria and the DRC, were overlooked because they are more difficult to work in. The largest number of experiments occurred in Western Europe and its offshoots, particularly the USA and Canada.

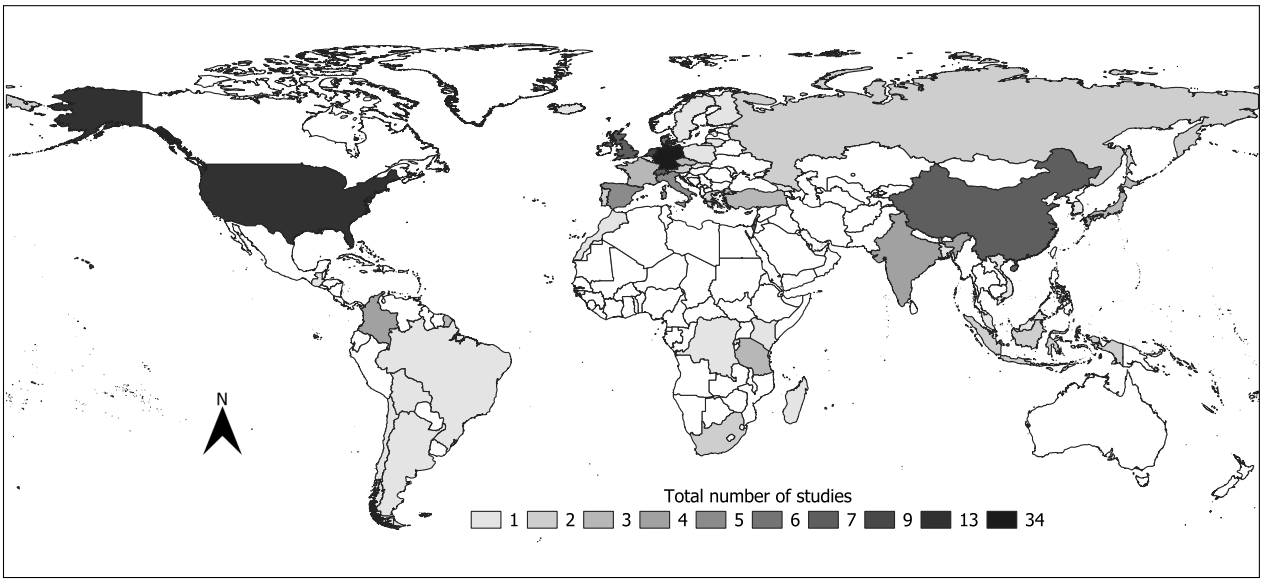
A more recent example is the meta-analysis of Abeler, Nosenzo and Raymond (2019) looking at cheating experiments. Despite being a more recent compilation of experiments, the geographic coverage shows little improvement relative to the older compilation in Johnson and Mislin (2011). There is a clear bias towards locations in more industrialized nations and away from developing countries, particularly in Africa and the Middle East.

Figure 4: Total number of trust-game experiments implemented in each country in the meta-analysis sample of Johnson and Mislin (2011)



Notes: The figure shows the number of studies with populations from each country for the sample of experiments contained in the trust game meta-analysis of Johnson and Mislin (2011).

Figure 5: Number of cheating experiments in each country in the meta-analysis sample of Abeler et al. (2019)



Notes: The figure shows the number of studies with populations from each country for the sample of experiments contained in their cheating game meta analysis of Abeler et al. (2019).

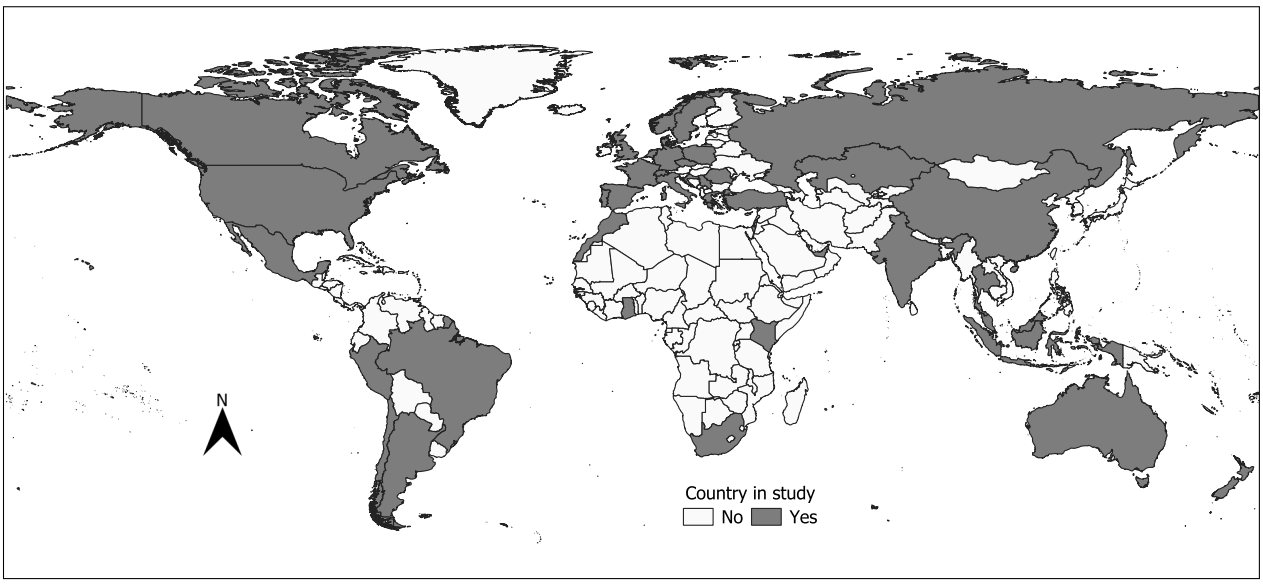
One of the experiments with an extensive sample from a large number of countries is the study of civic honesty by Cohn et al. (2019), where 17,303 wallets (some with money and some without) were dropped in 355 cities in 40 countries around the world. However, even in this study, where the authors went to great lengths to have as inclusive a sample as possible, there is still a clear bias in which countries are not included. The countries included are shown in Figure 6. Again, we see that most of Africa is not included, and the countries included are those that are logistically easier to work in, such as South Africa, Kenya, and Ghana. The pattern observed in Figure 6 is very similar to that in Figures 4 and 5.

The bias in participant samples is not specific to experiments and is also found in non-experimental surveys (Porteous, 2022, Stacy, Kitzmuller, Wang, Mahler and Serajuddin, 2024). Despite being one of the most thorough and inclusive surveys, the Global Preferences Survey (GPS), which uses a set of experimentally validated survey questions, still does not include many developing countries, particularly in sub-Saharan Africa. The sample of countries is shown in Figure 7. Similarly, the World Values Survey (WVS) also misses much of sub-Saharan Africa. Further, even when surveys are implemented in more challenging countries, the surveys tend to collect data in urban and easily accessible locations, limiting how representative the data are.

The non-representativeness of study participants globally limits our collective understanding of human behavior. While incomplete coverage of countries is not itself particularly problematic, the fact that there is a clear bias in which countries are included is more troubling, particularly since those that are included – WEIRD societies – appear to be at the tail of the distribution for many behavioral and cultural traits (Henrich et al., 2010c). Beyond this, extrapolating to the omitted countries is often difficult. The patterns we see for previously unstudied populations are often contrary to priors. For example, before the research by Henrich et al. (2001), most in the profession would not have guessed that smaller-scale societies would behave less prosocially with actions closer to the Nash equilibrium in the ultimatum game. More recently, the meta-analysis of Abeler et al. (2019) shows that cheating in cheating experiments is not systematically more prevalent in poorer or more corrupt countries. Thus, in general, behavioral patterns across societies are often surprising and difficult to predict ex-ante.

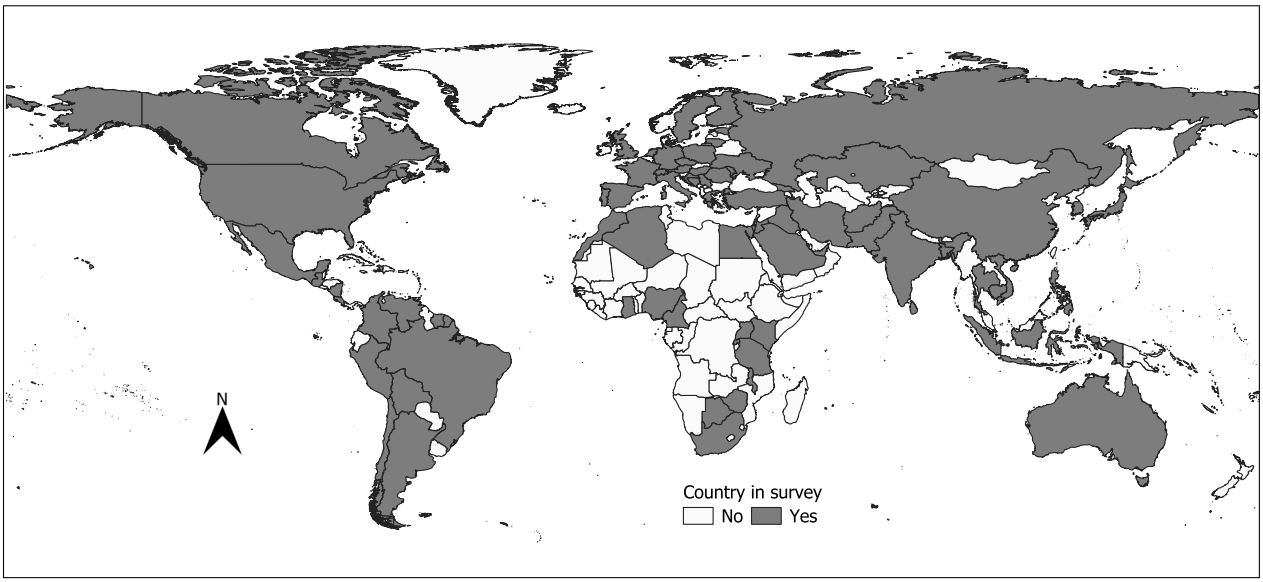
While costs, logistics, and convenience may discourage researchers from implementing experiments in new locations, there can be a compelling benefit: the opportunity to combine experiments with naturally occurring variation in developing countries. This combination adds a

Figure 6: Countries included in the wallet drop experiments of Cohn et al. (2019).



Notes: The figure shows the 40 countries included in the study of civic honesty across the globe by Cohn et al. (2019) where civic honesty was measured using a wallet drop experiment.

Figure 7: Countries included in the Global Preferences Survey of Falk et al. (2018).



Notes: The figure shows the 76 countries included in the Global Preferences Survey (GPS) by Falk et al. (2018). The GPS comprises data from survey questions meant to capture preferences related to altruism, trust, positive and negative reciprocity, time discounting, and risk. The survey questions were experimentally validated using a sample of participants from Germany.

crucial dimension to the experiments and can often drive site selection. Studies have considered populations that, due to the historical context, provide naturally occurring variation between groups. Lowes et al. (2017) implemented behavioral games among those whose ancestors who lived inside and outside of the Kuba Kingdom in the DRC. This allowed them to ask how a history of exposure to a state affects norms of rule-following. Similarly, Heldring (2021) examines the norms of rule-following among individuals living in villages just inside and outside of the Nyinga Kingdom in Rwanda. Lowes (2022) studied the cooperative behavior of spouses belonging to ethnic groups living near the border of the matrilineal belt in the DRC, leveraging ethnic group-level variation in matrilineal and patrilineal kinship structure. Lowes and Montero (2021) asked how a history of exposure to colonial concessions affects altruism and respect for property rights, recruiting individuals whose ancestors come from just inside or outside the former concession boundaries. Karaja and Rubin (2022) implemented trust games on both sides of the historical border between the Ottoman and Habsburg Empires in contemporary northeastern Romania. Rustagi (Forthcoming) studied the long-run consequences of Medieval participatory self-governance by implementing public goods games across cantons in Switzerland.[26] These are all cases where researchers went to a particular location because of rich naturally occurring variation, and the location and its history guided the choice and design of the experiments implemented.

An important consideration when selecting a research site is "saturation" – i.e., the extent to which the site or local population has been exposed to previous research studies. Given the logistical challenges of working in new places, it is natural that studies are implemented in the same locations once the data collection infrastructure has been established. This provides some benefits but also has drawbacks. Regarding benefits, it is often logistically much easier to work in places where prior studies have already been implemented. Additionally, local populations may be more trusting or welcoming of researchers if they have had past experiences. Another potential benefit is that past studies, like RCTs, can generate treatment variation that can potentially be leveraged in a follow-up study. For example, Haushofer, Larreboure, Lowes and Mait (2024) build off a previously implemented cash transfer program (Haushofer, Mudida and Shapiro, 2020) to ask how cash transfers shape the prosocial preferences of children, which had not been

[26] See Lowes (2021a) for an overview of the literature combining lab-in-the-field experiments with historical variation.

an outcome of interest in the original study. They collect lab-in-the-field data with children whose parents were part of the original cash transfer study. The study was feasible only because of the pre-existing experimental variation.

Potential drawbacks of working in "saturated" environments include participant fatigue because they have participated in many past studies. Additionally, in cases where particular populations have been involved in past studies, it may be necessary to control for their past exposure, especially if past interventions may influence the outcome of interest.

4.2. Choosing participants

Lab experiments are often implemented using an existing participant pool. A university lab, for example, will maintain an active database of potential participants and have the means to recruit new participants. In most developing countries, this type of database is not available, and screening surveys must be implemented to identify potential participants from the target participant pool. Generally, researchers want a participant pool to represent the full population as much as possible. A strategy that achieves this is randomly choosing individuals from the full population, which is difficult in any context, but particularly in less-developed countries where censuses are unavailable, and individuals may be more mobile seasonally or year-to-year.

In the absence of census-type data, one strategy to obtain a representative sample is to rely on satellite imagery to obtain a "census" of homes and then to choose homes to survey randomly. This has the benefit of providing a more structured sampling frame that allows the researcher to choose how to weigh different sampling units. This was employed, for example, in Lowes et al. (2017) and Le Rossignol, Lowes and Nunn (2024). There are many different ways to implement the randomization, but the studies followed a two-step procedure: first, they randomly selected polygons constructed from satellite imagery, and second they had enumerators follow a predetermined skip pattern among all houses within the polygon. Figure 8a and Figure 8b show the satellite images and polygons used for sample selection for these studies in the DRC.

In other settings, like smaller rural villages where tree or cloud cover obscures satellite imagery, other forms of randomization are possible. For example, enumerators can be instructed to start in the village center. From the village center, they can report how many paths there are and then randomize which direction to take and then follow a skip pattern along that street.

Figure 8: Satellite imagery used to obtain a random sample

(a) Satellite imagery used in Lowes et al. (2015, 2017).

(b) Satellite imagery used in Le Rossignol et al. (2024).

In certain studies, a particular subset of the population was targeted for an experiment. For example, Rustagi and Kroell (2022) study milk sellers in the outskirts of Delhi, India, and Lowes et al. (2017) focus on individuals from the Kuba, Lele, and other neighboring ethnicities living in the city of Kananga. One strategy to ensure as representative a sample as possible for the group of interest is to still take a random sample in the location of interest and implement a screening survey. The screening survey can be used to omit those individuals who do not belong to the group of interest. This strategy, which was followed by Lowes et al. (2017), helps ensure that the final sample is as representative as possible of the broader population of interest. An alternative strategy, which is more cost-effective, but potentially delivers a less representative participant set, is snowball sampling, where existing study participants are asked to help recruit future participants among those they know.

During the duration of the recruitment, screening, and implementation of experiments, there will be some attrition. Individuals sometimes choose to participate in a screening survey but decline to participate in a lab experiment. Or they consent to participate in the experiment but ultimately decline participation once they understand the full set of experimental choices. Thus, during the recruitment process, it is also important to maintain comprehensive records of all individuals selected for the experiment. If possible, keep data on all individuals targeted for recruitment into the study, to understand what types of individuals, if any, prefer not to

participate in the study. This will help you understand if you are capturing the intended subject pool and its representativeness of the broader population.

4.3. Infrastructural limits and other logistical challenges

In some settings, there may be limited access to reliable internet and electricity. If data collection is being done on tablets, it may be necessary to prepare solutions for charging this equipment, such as use of a generator or solar panels and rechargeable batteries. Pretest the enumerator workflow and build in redundancies (e.g., ability to access both a generator and a charged battery) to avoid any delays with data collection.

Relatedly, internet access may be slow or unreliable. Particularly for electronic data collection, it may be necessary to build in time for enumerators to access the internet to upload data. Longer periods without uploading the data increase the risk of data loss and also prevent the researcher from reviewing the data to provide any feedback.

While traditional experiments with university students in developed countries are typically designed to be communicated and implemented using computers (or analogs) in many other contexts, this may not be possible for various reasons. For example, in the DRC, we found that due to overheating, it was not possible to use laptop computers. Thus, when implementing common behavioral games, like the ultimatum or dictator games, we have had participants play with actual money, making allocations by placing money in envelopes. For example, Lowes et al. (2017) implemented the dictator game and ultimatum games using bills and envelopes. In the dictator game, participants were given ten 100 CF bills and were then asked how much they would like to give to a player 2. They made the allocation decision by putting the money to give to the other player in an envelope that they then sealed and returned. Player 1's decision in the ultimatum game was implemented using the same set of bills and two envelopes. The proposed amount offered to the other player was put in one envelope and the amount to keep was put in another envelope. Both were returned to the enumerator. The envelopes, which were marked, were brought back to the office and compared to the accept/reject decisions of the matched player 2. Based on this, payments were made to the players in a subsequent visit.

Implementing games using physical money rather than virtual money on a computer is logistically more challenging. The money and envelopes must be prepared prior to the experiment. In addition, the common strategy of only actually paying out one round of a set of games is often

much more difficult. For example, with the version of the dictator game described above where the participant keeps in their possession the money they do not give to the player 2, it would be logistically difficult and culturally awkward to ask for the kept money back because that round of that game was not selected to be paid out. However, a benefit of using real rather than virtual money is that the decision might feel more tangible and realistic.

From a logistical perspective, an important decision is whether mobile experimental equipment is brought to the participants or the participants come to a fixed experimental infrastructure. The downside of a fixed lab is that it can be exceptional. In many settings, being in a large structure with a consistent supply of electricity, a concrete floor, a tin roof, multiple rooms, electronic equipment, and multiple staff is not a common experience. Thus, experiments in a fixed lab may result in a particularly novel environment for the participants. In addition, the logistics of having participants come to a fixed lab, rather than moving the lab to them, typically means that appointments must be made in advance, that it may be necessary to facilitate participants' transport and to send them reminders. All of these have the potential to raise anticipation and to make the experiment particularly exceptional or novel. An additional concern is that these logistics also typically lead to higher rates of attrition.

Having a fixed lab offers the benefits of a more secure supply of electricity and better control over the external environment, e.g., better temperature control, removal of unwanted external stimuli, increased ability to ensure complete privacy, etc. Additionally, existing fixed labs usually have the benefit of an existing participant pool, which reduces the logistical burden of running the experiment. There are often requirements for inclusion in the participant pool, such as having a cell phone. This, along with other forms of attrition, could lead to selection if the researcher does not create their own participant pool. In a helpful analysis, Haushofer, Collins, de Giusti, Njoroge, Odero, Onyango, Vancel, Jang, Kuruvilla and Hughes (2014) compare the characteristics of the Busara lab with the general population of Kenya. They show that the population of the participant pool is broadly comparable to that of Kenya in general. Deviations that existed were understandable. In the Busara pool, women are slightly over-represented.[27] However, this was driven by women who had fewer children. For example, the share of women in the Busara pool who had five or more children was 13.33% which is lower than the same figure for Kenya, 24.9%. The Busara subject pool also showed a slight over-representation of people in their twenties and

[27] In the Busara pool, 55.4% were women (compared to 48.9 in Nairobi and 50.2 in Kenya)

an under-representation of older participants. These disparities likely stem from factors related to the convenience and accessibility of the Busara lab.

Some of the larger deviations were for education, particularly for women. For example, in the Busara sample, 1.93% of men had no education, while for Nairobi, the figure was 5.53%, and Kenya was 6.70%. For women, only 0.68% of women had no education in the Busara sample, but for the broader population, the fraction was 5.57 (Nairobi) and 16.2 (Kenya). Those with very high levels of education also appear to be under-represented. In the Busara pool, 14.51% of men had college/university education, while in Nairobi, this number is 18.35.

Haushofer et al. (2014) effectively quantify the trade-offs researchers face when deciding between using an existing participant pool and creating their own set of participants. The optimal choice depends on the importance of having a sample representative of the broader population of interest, weighed against the convenience of using existing participant pools.

In pilots in the DRC, we found that when experiments were run out of our fixed labs, participants tended to view everything with greater astonishment, and the whole experience was particularly novel to them. We observed that this often affected participants' decisions. For example, in an experiment that asked participants to choose between a fixed sum of money and participating in a gamble that yielded a higher payoff with some probability and a lower payoff with some probability, participants always chose the gamble rather than the certain payoff. The primary reason was that, in their mind, if they received a fixed payoff, the experience would be over. If they chose the gamble, then it would continue to a second stage of activity.

Additionally, in more traditional (fixed) lab settings, lab experiments can be implemented with networked computers. Participants can make choices and respond to the choices of other participants in "real-time". Participant payoffs can be calculated immediately based on their choices and the choices of others present in the experiment. In a setting without a lab and without networked devices, alternative strategies will need to be developed for experiments that involve choices made by multiple players. For example, it may be possible to build a database of other player choices that can be used to calculate participant payoffs immediately.

Figures 9–11 present photos of the lab settings from various studies. We group labs into three types: fixed labs, mobile labs, and non-lab (more naturalistic) setups. As shown, the nature of the 'lab' can vary widely, ranging from a mat on the floor, portable tents outside a person's home, a cricket pitch, a working factory, or networked computers similar to traditional lab experiments in

developed-country university labs.

4.4. Ensuring comprehension

A significant capacity constraint relevant to implementing lab-in-the-field experiments in many parts of the world is the prevalence of illiteracy among potential participants. This presents unique challenges, as protocols designed for university undergraduates in industrialized countries are often unsuitable for illiterate populations. In such settings, researchers must adapt their methodologies to ensure meaningful participation and comprehension.

To address these challenges, researchers can implement several strategies. Experiments should be designed so that responses can be provided, ideally in private, by illiterate participants. For instance, instructions can be read orally to maintain comprehension and allow each participant to learn at their own pace. Alternatively, experiments can be designed such that decisions do not require literacy. This may involve using verbal instructions coupled with touch screens displaying images rather than written words or employing physical objects (e.g., money) for decision-making.

Even for when the population is literate in general, if one wants a participant pool that is truly representative of the broader population, then protocols must account varying levels of literacy, slower comprehension, and/or language issues. The common practice of screening out individuals with weak comprehension due to these or other factors does not yield a representative participant pool. An alternative strategy is to pair written instructions with oral instructions and images that facilitate comprehension. This, combined with comprehension checks and the ability for instructions to be repeated to participants, helps to ensure all individuals, even those with slower comprehension, can participate effectively.

Similar considerations are important in studies that focus on populations for which comprehension might be particularly low. In work studying the effects of cash transfers on pro-social preferences with children, it was important to design experimental protocols that were accessible to children as young as six (Haushofer et al., 2024). To make the choices more contextually appropriate, the games were designed so that children could allocate fruits to a basket labeled for themselves and another child. An enumerator sat with the child as they made their allocation decisions. The child would accumulate tokens, that could then be used to purchase different prizes with different age-appropriate items.

Figure 9: Experimental Set Ups: Fixed Labs

(a) Fixed lab in Fiji from Henrich and Henrich (2010).

(b) Fixed lab in Fiji from Henrich et al. (2006b, 2010a), Ensminger and Henrich (2014).

(c) Fixed lab in Mulago-Kyebando, Uganda from Habyarimana et al. (2007, 2009).

(d) Fixed lab in Toyama Bay, Japan from Carpenter and Seki (2011).

(e) Fixed lab in the Republic of Georgia from Bauer et al. (2014b).

(f) Fixed lab in Nairobi, Kenya from Haushofer et al. (2023).

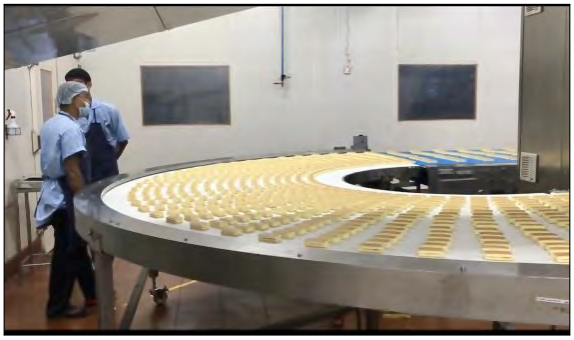
Figure 10: Experimental Set Ups: Mobile Labs

(a) Mobile lab in Kananga, DRC from Lowes et al. (2017).

(b) Mobile lab in Gemena, DRC from Le Rossignol et al. (2024).

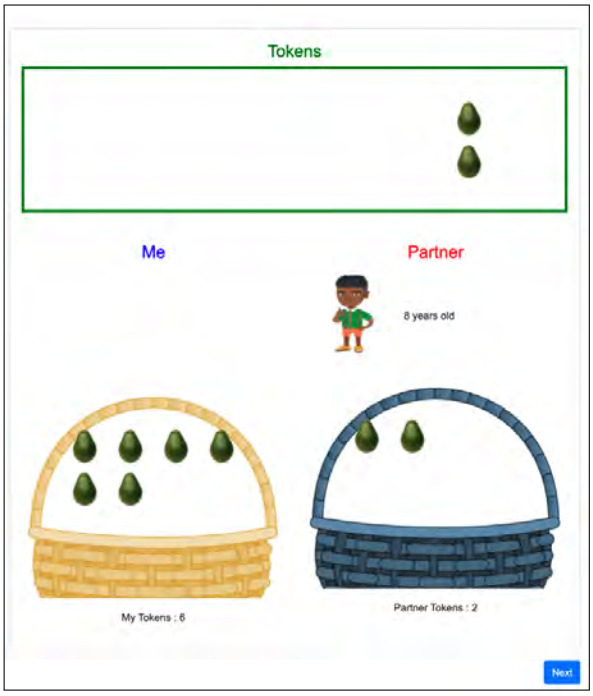
Figure 11: Experimental Set Ups: Non-Lab Setups

(a) Non-lab experiment in West Bengal, India from Ghosh (2024).

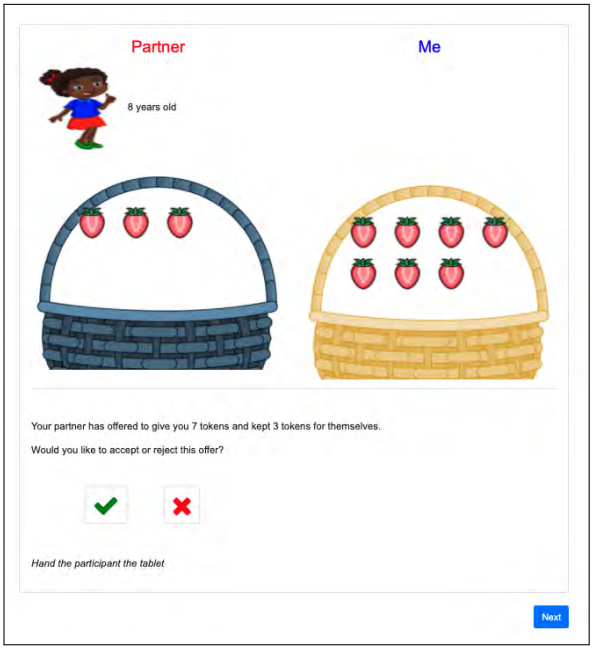


(b) Non-lab experiments in Uttar Pradesh, India from Lowe (2021).

Figure 12: Examples of the Dictator Game and the Ultimatum Game



(a) Dictator game



(b) Second player of ultimatum game

Notes: Screenshots from experiments implemented in Nakuru County, Kenya, with 4,022 children aged 6 to 17 (Haushofer et al., 2024).

Practical recommendations

Several best practices can enhance comprehension of experimental tasks, particularly in low-literacy settings. First, participants should be paired with an enumerator who reads the experimental protocol with them in the local language. If subjects do not have experience with touchscreen tablets, it is advisable for the enumerator to handle all data entry.

Protocols may need to be made "less abstract" by including props to illustrate the various choices being made in the experiment. For example, for choices that involve the allocation of money, the enumerator may demonstrate proposed allocations using actual currency. Protocols should include detailed examples of the types of choices that the participant will make during the experiment, clearly illustrating how the choices affect the final payoffs that a participant will receive. To avoid priming subjects to make specific decisions, particularly with the potential for anchoring effects, it is recommended to randomly assign examples to participants and to randomize their order.

One example of tailoring the protocols to the local context comes from Lowes et al. (2015) who implemented an Implicit Association Test (IAT) in the DRC. The IAT was intended to measure implicit bias towards different ethnic groups. We made several modifications to the protocol to ease the implementation of the IAT. First, we used the single-target IAT, in which individuals sort a "target" (such as "man" or "coethnic") with images representing good or bad. This is in contrast to the standard IAT, in which individuals sort two targets during the same IAT round (e.g. man and woman; or coethnic and non-coethnic). The second modification we made was to use headphones to present the target information to the respondent. Normally, a participant reads a particular stimulus or sees an image on a screen. Instead, we chose to have participants wear headphones and hear the target words. This overcame any challenges related to literacy.

In addition, it is helpful to include a series of test questions that participants complete before undertaking the experimental task. These questions should cover the basic incentive structure of the task and the types of choices the participant will make. The researcher may consider having participants meet a certain threshold of correct responses before participating in the experiment. If the participant does not meet the threshold of correct responses, the enumerator can re-explain the experimental protocol to the individual. Moreover, it is helpful to record the correct response rate to the test questions for robustness analysis.

For experimental tasks involving the allocation of money between different participants, the

researcher may consider having the participant make the allocation choices with real money. This makes the choice very tangible and salient. For example, for a dictator game in which the respondent allocates money between themselves and another player, the participant can receive the endowment in cash, and then place the money for themselves in an envelope and for the other player in a different envelope.

Translation (and back-translation)

An important element of implementing lab-in-the-field experiments is that the protocols are accessible to and easily understood by participants. To facilitate a reasonable level of comprehension, protocols must be translated into the local language, and scholars typically must work with local survey enumeration teams who are fluent in the languages/dialects, as well as being knowledgeable about local practices and conventions. To ensure quality translations, it is recommended that the survey be translated from the researchers' language into the local language and then independently back-translated by another individual into the original language. The initial text should then be compared to the back-translated text by another independent third party. Any discrepancies during this process indicate issues with the translation and should be resolved.

4.5. Ethics and IRB Approval

Studying human subjects involves ethical considerations, particularly when a researcher is experimentally manipulating some treatment. There are ethical considerations that are specific to or heightened for research in developing country contexts. Researchers sometimes choose to work in developing countries because research is cheaper, and the stakes that you can experimentally vary (relative to average income) can be much higher. This has been described by scholars as a potential benefit of lab-in-the-field experiments in developing countries (see Harrison and List, 2004, Charness et al., 2013). However, the larger relative stakes mean that there is more on the line for participants and that there is greater potential for effects that spill over outside the experiment. For example, in the study by Casey et al. (2012), since the experimental games were high stakes and embedded in real life, there is the potential that they may have sizable effects on the well-being and social cohesion of villagers.

Since field experiments involve human subjects, it is necessary to seek Institutional Review Board (IRB) approval prior to conducting the research. Generally, each university has its own IRB that will review a research protocol and grant approval if all human subjects' protections are met. This process can be time-consuming and potentially involve multiple rounds of submissions. It may also be necessary to obtain IRB approval from the study site. Whether this is necessary will depend on local regulations and the presence of a local IRB. The researcher's university's IRB approval may be contingent on receiving local IRB approval.

Additionally, depending on the country, researchers may also need to obtain a research license – which gives the individual permission to undertake research in the country. These licenses may stipulate, for example, that the researcher has an affiliation with a local accredited institution that is able to monitor the research project.

In general, and consistent with the requirements of human subject protections, researchers should keep in mind the following principles, which are similar to those for any field experiment, when designing lab-in-the-field experiments.[28] First, it is important to consider any potential risks for subjects that participate in the experiment, as well as for those implementing the experiment. These risks may include the potential for any physical, psychological, economic, social, or legal harm. Any potential for such risks should be considered and steps must be taken to minimize these risks. Additionally, researchers should build in feedback mechanisms so that if any risks or negative effects arise, they are able to report those outcomes to their IRB and effectively prevent them in the future. Second, researchers should consider what the benefits are of their study, for either the individual participants or for society more broadly. If the study offers limited benefits, then the authors should carefully reconsider its value before proceeding with the project. Third, understanding the relative risks and benefits will require knowledge of the local cultural context. Researchers should consult knowledgeable locals to identify any potential risks and to develop strategies to mitigate them.

[28] For an overview and recommendations of ethical considerations for RCTs in development, see Evans (2023).

5. Experimenter demand effects

5.1. What is the experiment really about?

Within developing country contexts, standard experimenter demand effects are expected to be a concern. These refer to cases where participants' behaviors are affected by their perceptions of appropriate behavior – e.g., what the experimenters 'demand' from them (Zizzo, 2010). While this concern is present in all experiments, it is likely to be heightened for experiments in developing country settings for several reasons delineated below.

First, lab experiments are particularly novel in many developing country contexts. To academics, the fact that researchers travel to remote parts of the world and spend tens or hundreds of thousands of dollars to collect information through surveys or lab experiments is natural and commonplace. However, for many individuals in developing countries, this is difficult to believe. During the course of our research, respondents, participants, and even enumerators expressed confusion and curiosity about our endeavor. Why would individuals in wealthy countries, like the United States or Canada, come to the DRC just to have people in rural villages play strange games? From their perspective, it is puzzling, and the idea that all of that effort is solely for the purpose of learning does not seem credible. It is very common for individuals to believe that this is not the truth. Within the areas in which we work, it is common for researchers to face suspicion, and even be accused of being spies. In some sense, this is not far from the truth since both collect information. The key difference between a spy and a researcher is the reason for the data collection.

Second, experiments in developing country contexts typically involve sizable payoffs, often equivalent to multiple days' wages. Thus, participants have strong incentives (and likely much stronger than in traditional lab experiments) to ensure the research is successful, knowing that successful research may lead to future studies and additional opportunities. Thus, while the monetary incentives in the games are stronger, the incentives for the participants to please the experimenters are also more pronounced. Furthermore, to the extent that experimenter demand is driven by perceptions of appropriate behavior in the experiment (Zizzo, 2010), the fact that developing countries (and non-WEIRD countries more generally) tend to have tighter social norms (Jackson, Gelfand and Ember, 2020, Gelfand, Raver, Nishii, Leslie, Lun, Lim, Aycan et al., 2011), means that these effects are stronger in developing country contexts. This is consistent with the

findings from Medvedev et al. (2024), which show that the importance of non-monetary changes to the experimental protocols (what they call 'psychological incentives') relative to monetary incentives was higher among less-developed non-WEIRD cultures.

One recommended practice to mitigate these effects is to have a set of exit questions that ask individuals about their choices in the experiment and what drove their decisions. For example, researchers might ask respondents how they decided on their choices or if the decision reminded them of decisions in their daily life. These exit questions may help interpret results that are unexpected or confirm that the participants indeed understood the choices they were making.

5.2. Assessment

A primary concern is that participant behavior is driven, to an important extent, by what participants feel that the experimenters want them to do or by what they think experimenters are looking for. Since experiments provide participants with a significant amount of money, they have a vested interest in ensuring that the experiment is 'successful' from the experimenter's point of view. Such experimenter demand effects are a potential concern in all experiments but are particularly important for experiments that take place outside universities in industrialized countries.

Scholars have designed strategies to assess and address the presence of experimenter demand effects. For example, de Quidt, Haushofer and Roth (2018) propose a strategy researchers can employ to assess the extent to which participants' behavior in an experiment is influenced by experimenter demand effects. They do this by communicating to participants: "We expect that participants who are shown these instructions will..." They also implement a stronger version of this: "You will do us a favor if you..." They argue that comparing the estimate of interest to the baseline protocols with the statement included provides one with evidence on the strength of the experimenter demand effects and allows one to bound the estimates of interest. In their setting of 19,000 online participants (mostly recruited from MTurk) who undertake 11 common experimental tasks, they find modest effects for the weaker treatment, most of which are close to zero. They find larger effects for the stronger treatment. More recently, Mummolo and Peterson (2019) reach a similar conclusion within the context of online survey experiments (using MTurk or an online survey company). Beyond (randomly) providing information about the experimenter's intent, they also provide financial incentives. They find that both interventions have small effects.

An important caveat of the evidence from the two above studies is that the participant pool is very different from those in developing country contexts. Thus, it is unclear if these findings apply to participants not drawn from industrialized countries with access to the internet. Existing evidence suggests that in developing country contexts, experimenter demand effects are likely important. For example, in Sierra Leone, it was found that having a white foreigner present affected behavior in games (Cilliers, Dube and Siddiqi, 2015). Individuals contributed more in a dictator game when a white foreign was present. Further, even the ethnicity of non-foreigners may also matter. Adida, Ferree, Posner and Robinson (2016) found that within 14 African countries, the ethnicity of the surveyor matters – especially in cases where the ethnicities of the respondent and/or surveyor have a history of political competition. Likewise, Fiala and di Maio (n.d.) randomly assign enumerators to respondents in Uganda. They found that while enumerator identity does not affect responses in less sensitive questions, there are large effects for sensitive questions – such as support of opposition political parties. Lowes et al. (2015), who study ethnic preferences, found that for their survey-based measures, the identity of the enumerator matters. Respondents report feeling closer to and having more positive perceptions of an ethnic group if the enumerator was from that group.

5.3. Addressing Experimenter Demand Effects

Given the potential importance of experimenter demand effects, which are heightened in developing-country contexts, maximizing the external validity of experimental findings (i.e., ensuring the behavior is most likely to hold outside of the experiment), requires making the specific aims, goals, and hypotheses of the experiment opaque. A belief in the purpose of the experiment, even if incorrect, can significantly influence participant behavior.

List experiments shroud a person's own beliefs by asking participants how many statements they agree with and then randomizing the presence or not of the statement of interest. Methods like the IAT or the conjunction fallacy provide ways of measuring implicit associations that do not rely on survey questions. This is particularly important since participants are often not aware of their implicit associations. Lowes et al. (2015) showed that implicit and explicit ethnic preferences and stereotypes can be very different. The study also demonstrated that while explicit preferences are prone to demand effects (the ethnicity of the enumerator affects the measure), this is not the case for the single-target ethnicity IAT that they implement. Consistent with the ST-IAT being

less prone to demand effects because it captures implicit associations, the study found no effect of enumerator ethnicity on the ST-IAT measure of ethnic bias.

Cheating games, like the resource allocation game (RAG), follow the same logic. In the version implemented by (Lowes et al., 2017), participants roll a die with three sides that are white and three sides that are black. Before they roll the die, the participant chooses the color to associate with them winning a fixed amount of money. They then roll the die and see if they win. If they do, they take the money. This setup, which maximizes hideability, makes it impossible to determine if participants are truthful about any individual roll since the choice of which color indicates a win is made internally. But with larger samples, one can see the extent to which the distribution of wins deviates from the expected 50%.

Another strategy is to strive to make the experiment's goals as opaque as possible. While this is possible for many more complex experiments, it is more challenging for simpler experiments, which often are best suited for field experiments. Take, for example, the following scenario. A participant has a single coin. They can give the coin to one of two people. One person is a coethnic, and another person is a non-coethnic. Who do they give the money to? If money is disproportionately given to coethnics, then this is evidence of ethnic bias in the population. This is an example of the situation and behavior that was of interest to Habyarimana et al. (2007, 2009) during their pioneering studies that measured ethnic preference in Uganda using behavioral experiments. Aware of problems related to the transparency of the research question, they implemented a slightly different version of the experiment. Rather than having one coin, participants had two. The two coins were to be allocated across three players: oneself, a coethnic, and a non-coethnic. Most players gave the first coin to themselves, and the second was then allocated between one of the two other players, effectively leaving the same decision but with much less transparency about the reason for the experiment. Beyond this, the authors also reported additional characteristics of the players, and they randomized the players that appeared in the pairs so that it wasn't always one coethnic and one non-coethnic. Subsequent work has provided evidence that whether or not one finds ethnic preferences may depend on the method of measurement and whether the measurement can be manipulated by the participants (Hazlett, Hamilton and Posner, 2024, Blum et al., 2021).

The rationale behind these strategies is that if the participants do not know the aims or even focus of the experiment, they are less likely to attempt to please the experimenter in a systematic

way that generates the expected results. Some researchers have used strategies that assess the extent to which the purpose of the experiment was made sufficiently opaque to the participants. For example, Siddique, Vlassopoulos and Zenou (2024) at the end of their experiments, which were aimed at assessing the effects of providing video information about outgroup members (Santals) Bengalis in Bangladesh, the authors ask participants: "If you had to guess, what would you say was the purpose of this study? You will only get one guess, and if you guess it correctly, you will get 50 Taka." They were then given eight options. They found that within their sample, only 3 percent correctly guessed the research purpose.

Although not studied to our knowledge, it is plausible that other issues that commonly contaminate experiments could be stronger in developing country contexts. For example, it is known that peoples' behaviors can change just from the knowledge that they are being watched, a phenomenon known as the Hawthorne effect (Mayo, 1933, Roethlisberger and Dickson, 1939). This underscores the importance of providing privacy to participants when making decisions in behavioral experiments, such as the factory work experiment developed by Breza et al. (2018) or the town hall meetings and evaluation form experiments designed by Weigel (2020) described earlier in the chapter.

6. Conclusion

Behavioral and lab-in-the-field experiments are becoming increasingly used by scholars, particularly those aiming to better understand human behavior and economic development in non-WEIRD contexts. In this chapter, we outlined the practical and conceptual challenges researchers may face implementing lab-in-the-field experiments in non-Western settings. While some of the issues we raised are common to all experiments, many are specific to or more pronounced in developing country contexts. These include logistical and capacity constraints, worldviews and assumptions held by the participants that are potentially different from those designing the experiments, and heightened experimenter demand effects. In response to these challenges, we have outlined suggestions for best practices and considerations when implementing experiments in developing country contexts.

Table 2: Examples of Lab-in-the-Field Experiments

Citation	Location	Research aim	Lab experiments
Afzal et al. (2022)	Pakistan	Test for knowledge of partner's consumption preferences and demand for decision-making agency within household.	Consumption bundle choice experiment
Ashraf (2009)	Philippines	Identify how information and communication affect decision-making within households and financial behavior.	Intrahousehold decision-making experiment
Attanasio et al. (2012)	Colombia	Study the effects of pre-existing social networks on risk-sharing behavior and group formation.	Risk pooling experiment
Balakrishnan et al. (2020)	Kenya	Test whether experimental payment timing (immediate payment or close of day) affects present bias.	Convex time budget experiment
Barboni et al. (2013)	Bolivia	Test for the presence of adverse selection and the effects of joint-liability lending on moral hazard.	Investment experiment
Barr (2003)	Zimbabwe	Test whether trusting behavior is driven by trustworthiness of others.	Trust game
Barr and Genicot (2008)	Zimbabwe	Test the effects of changes in the level of commitment and information available to individuals when entering into risk-sharing arrangements.	Risk sharing experiment
Barr et al. (2012a)	Zimbabwe	Study assortative matching by gender when economically useful groups are being formed.	Risk sharing experiment
Barr et al. (2012b)	Zimbabwe	Test effects of external enforcement and social punishment on risk-sharing.	Risk sharing experiment; Trust game
Barr et al. (2015)	UK, South Africa	Effect of endowment being earned or randomly assigned on allocations to others in modified dictator games.	Modified dictator game
Barr et al. (2018)	Nigeria	Compare the willingness of husbands and wives to contribute to household public goods across monogamous and polygynous households.	Public goods game
Bauer et al. (2014a)	Georgia, Sierra Leone	Assess how the experience of war shapes people's social preferences, particularly fairness and altruism.	Dictator game; Sharing game; Envy game
Berge et al. (2020)	Kenya	Assess the degree of coethnic bias in preferences for redistribution and cooperation.	Dictator game; Public goods game; Choose your dictator game; Implicit association test
Binswanger (1980)	India	Test whether variation in farmer behavior across different wealth levels are due to differences in risk aversion or limitations in access to credit and modern inputs.	Risk preference elicitation

Citation	Location	Research aim	Lab experiments
Blum et al. (2021)	Kenya	Use misattribution-based approaches and standard experiments to measure coethnic bias.	Dictator game; Public goods game; Choose your dictator game; Affect misattribution procedure; Face anger attribution task; Weapon misidentification task
Breza et al. (2018)	India	Predict when pay inequality could generate adverse effects. Are fairness norms violated if pay levels are unequal or does fairness require that pay differences reflect productivity differences across workers.	Randomized controlled trial; Treatment: different pay structures
Brooks et al. (2018)	India	Investigate the effects of culture (i.e. high and low caste) on individuals' tendency to coordinate in repeated social dilemmas.	Repeated coordination game
Bursztyn et al. (2018)	Indonesia	The authors provide field-experimental evidence of the existence of status goods; test for the associated positional externalities; and study how self-image interacts with social image in explaining the demand for status.	Randomized controlled trial; Three experiments: (1) Randomize status of credit card; (2) Randomized information of credit card product; (3) Randomized into self-affirmation tasks
Bursztyn et al. (2020)	Saudi Arabia	Test the effects of randomly updating the social norm about female participation in the labor force among Saudi men.	Randomized controlled trial; Treatment: revelation of social norm
Cardenas (2000)	Colombia	Study how individuals try to solve the commons dilemma through self-governance.	Public goods game
Cardenas et al. (2000)	Colombia	Examine the effects of external institutions (rules) on collective decision-making in common-pool resource settings.	Public goods game
Carpenter and Seki (2011)	Japan	Test whether student measures of social preferences predict real-world cooperation among fishermen.	Public goods game
Carpenter et al. (2004a)	Thailand, Vietnam	Search for links between behavioral social capital and economic outcomes in a cross-country comparison.	Public goods game
Carpenter et al. (2004b)	Thailand, Vietnam	Conduct a comparative analysis of social capital in slums using experiments on cooperation and public goods provision.	Public goods game
Carpenter et al. (2005)	United States	Explore the external validity of experimental results by replicating well-known lab studies in the field.	Ultimatum game; Dictator game
Casey et al. (2012)	Sierra Leone	Evaluation of program that seeks to strengthen institutions and representation.	Randomized controlled trial; Evaluation of The GoBifo project

Citation	Location	Research aim	Lab experiments
Cassar and Zhang (2022)	Togo, Sierra Leone, Bosnia, Colombia, China	Propose a theoretical framework rooted in evolutionary psychology to explain the occurrence of the gender gap in competitiveness frequently found in economic experiments and test it with data.	Competition game
Cassar et al. (2007)	South Africa, Armenia	Estimate the importance of social capital to the performance of microfinance institutions.	Trust game; Microfinance game
Cassar et al. (2013)	Tajikistan	Study the effect of individual exposure to civil conflict on trust and preferences for market participation.	Trust game; Dictator game; Ultimatum game
Cassar et al. (2016)	China	Study the gender gap in willingness to compete using experimental data.	Competition game
Cassar et al. (2017)	Thailand	Test whether a natural disaster can have systematic effects on individuals' long-term preferences relating to trust, risk, and time discounting.	Trust game; Risk preference elicitation; Time discounting preferences elicitation
Cassar et al. (2023)	Sierra Leone	Study the possibility that women and men have distinct behavioral reactions to victimization.	Competition game; Cooperation game
Cilliers et al. (2015)	Sierra Leone	Study the effect of researcher race and nationality on experimental outcomes in dictator games.	Dictator game
Cohn et al. (2019)	Argentina, Australia, Brazil, Canada, Chile, China	Examine the trade-off between honesty and self-interest.	Natural field experiment: honesty experiment (lost wallets)
Conlon et al. (2021)	India	Measure how well spouses learn from each other about risk preferences over the course of a series of interactions.	Social learning task
Conlon et al. (2022)	India	Test the standard assumption in economics that equivalent pieces of information are weighed equally regardless of their source.	Social learning task
Ensminger and Henrich (2014)	Ghana, Papua New Guinea, Russia, Kenya, Tanzania, Ecuador, Mongolia	Examine levels of fairness, cooperation, and norm enforcement in small-scale societies around the world.	Dictator game; Ultimatum game
Fafchamps et al. (2015)	Ethiopia, UK	Test whether redistributive actions hinder the formation of Pareto-improving groups.	Risk taking experiment
Flory et al. (2017)	Malawi	Investigate how women's preferences over competition and risk-taking are affected by cultural norms.	Competition game
Giné et al. (2010)	Peru	Explore the impact of a variety of individual and community characteristics on microfinance repayment behavior.	Microfinance games
Gneezy et al. (2009)	Tanzania, India	Study whether there are gender differences in selecting into competitive environments across two distinct societies: the Maasaiin Tanzania and the Khasi in India.	Competition game

Citation	Location	Research aim	Lab experiments
Godoy et al. (2021)	Bolivian Amazon	Identify whether and how the introduction of a financial institution (lock box) changes savings, borrowing, and investing behavior.	Randomized controlled trial. Treatment: lockboxes
Habyarimana et al. (2007)	Uganda	Understand why ethnic diversity undermines public goods provision using laboratory experiments.	Public goods game
Habyarimana et al. (2009)	Uganda	Understand why ethnic diversity undermines public goods provision using laboratory experiments.	Dictator game
Haushofer et al. (2023)	Kenya	Examine the effects of stress on coethnic preferences in economic decision-making tasks.	Dictator game; Choose your dictator game; Trust game; Social proximity survey
Haushofer et al. (2024)	Kenya	Study the effects of an unconditional cash transfer program on social preferences of children.	Dictator game; Ultimatum game; Third party dictator game; Joy of destruction game; Public goods game
Heldring (2021)	Rwanda	Test the hypothesis that centralized states affect the development of norms of obedience to political authority to develop.	Resource allocation game
Henrich (2000)	Peruvian Amazon	Examine whether humans everywhere deploy the same economic behaviors or if these behaviors are shaped by cultural norms.	Ultimatum game
Henrich et al. (2001)	Peru, Tanzania, Bolivia, Ecuador, Mongolia, Chile	Test whether deviations from the canonical model of self-regarding behavior are due to cultural variations.	Ultimatum game; Public goods game; Dictator game
Henrich et al. (2006a)	Ghana, Kenya, Tanzania, US, Colombia, Ecuador	Explicitly measure costly punishment via a strategy method in small-scale societies.	Ultimatum game; Third party punishment game
Henrich, Ensminger, McElreath, Barr, Barrett, Bolyanatz, Cardenas, Gurven, Gwako, Henrich, Lesorogol, Marlowe, Tracer and Ziker (2010b)	Ghana, Papua New Guinea, Russia, Kenya, Tanzania, Ecuador, Mongolia	Test their theory that market norms may be more prevalent in societies with greater market integration.	Dictator game; Ultimatum game; Third party punishment game
Hjort (2014)	Kenya	Investigate the productivity effects of ethnic diversity and the role of ethnic quotas in teamwork settings.	Natural experiment; Treatment: ethnical composition of team
Hoff et al. (2011)	India	Investigate how the social structure of a society affects prosocial behavior and third-party punishment.	Trust game
Iversen et al. (2011)	Uganda	Experimentally test whether and to what extent gender norms influence contributions to household efficiency.	Public goods game

Citation	Location	Research aim	Lab experiments
Jakiela (2011)	US, Kenya	Measure internalized informal institutions, such as norms of sharing, in different cultural contexts.	Dictator game variants
Jakiela (2015)	Kenya	Test the extent to which individuals living in communal societies make redistributive choices based on social pressure.	Dictator game
Jakiela and Ozier (2016)	Kenya	Report the results of an experiment designed to measure social pressure to share income with relatives and neighbors.	Investment game with possibility to hide endowment
Jakiela and Ozier (2019a)	Kenya	Estimate the impact of a specific episode of civil conflict on risk preferences.	Risk preference elicitation
Jakiela et al. (2015)	Kenya	Measure the causal impact of human capital on respect for earned property rights.	Dictator game variant
Karaja and Rubin (2022)	Romania	Understand the degree to which norms are transmitted intergenerationally.	Trust game
Kebede and Zizzo (2015)	Ethiopia	Explore the existence and variations of money burning behavior across different groups.	Money burning games
Kebede et al. (2013)	Ethiopia	Conduct a comprehensive examination of the effect of pro-social norms on contributions to public goods.	Public goods game
Le Rossignol et al. (2024)	Democratic Republic of the Congo	Examine how those with traditional religious beliefs are treated.	Dictator game; Choose your dictator game; Joy of destruction game; Norms elicitation
Lowes (2021b)	Democratic Republic of the Congo	Test how kinship structure (i.e. matrilineal relative to patrilineal kinship) affects willingness to compete.	Competition task; Stress measurement
Lowes (2022)	Democratic Republic of the Congo	Test how kinship structure (i.e. matrilineal relative to patrilineal kinship) affects spousal cooperation.	Public goods game; Dictator game
Lowes and Montero (2021)	Democratic Republic of the Congo	Examine the effects of the rubber concessions on social preferences.	Dictator game; Reverse dictator game
Lowes et al. (2015)	Democratic Republic of the Congo	Test whether coethnicity affects implicit and explicit bias.	Single-target implicit association test
Lowes et al. (2017)	Democratic Republic of the Congo	Examine the long-term impact of institutions on cultural norms.	Resource allocation game; Ultimatum game
Marlowe et al. (2008)	Ghana, Kenya, Tanzania, US, Colombia, Ecuador	Test the hypothesis that population size (and contact) predicts punishment norms.	Dictator; Ultimatum game; Third party punishment game
Medvedev et al. (2024)	India	Systematically compare the effectiveness of monetary versus non-monetary incentives in promoting pro-social behavior.	Social norms experiment
Munro et al. (2014)	India	Examine the implications of the regional variation in female autonomy for the efficiency of family decision-making.	Modified voluntary contribution games

Citation	Location	Research aim	Lab experiments
Negussie et al. (2018)	Ethiopia	Examine how behavioral factors contribute to smallholder's decisions on adapting to climate change through the adoption of climate innovations.	Lottery game; Time preference game
Rao (2019)	India	Study how being integrated with poor students affects the social behaviors and academic outcomes of rich students.	Dictator game
Roth et al. (1991)	Israel, Japan, US, Yugoslavia	Compare behavior in related bargaining and market environments.	Ultimatum game
Rustagi and Kroell (2022)	India	Investigate whether sellers differ in innate honesty (incur private cost to provide good quality) and whether this explains the variation in product quality.	Dictator game
Weigel (2020)	Democratic Republic of the Congo	Provide evidence from a fragile state that citizens demand more of a voice in the government when it tries to tax them.	Randomized controlled trial; Randomized property tax collection

Notes: The table summarizes findings from lab-in-the-field experiments conducted between 1991 and 2024, covering a total of 85 published and working papers.

References

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, July 2019, *87* (4), 1115–1153.

Adida, Claire L., Karen E. Ferree, Daniel N. Posner, and Amanda Lea Robinson, "Who's Asking? Interviewer Coethnicity Effects in African Survey Data," *Comparative Political Studies*, 2016, *49* (12), 1–31.

Afzal, Uzma, Giovanna d'Adda, Farah Said, and Bereket Kebede, "Intrahousehold Consumption Allocation and Demand for Agency: A Triple Experimental Investigation," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 400–444. with Online Appendix.

Ashraf, Nava, "Spousal Control and Intra-household Decision Making: An Experimental Study in the Philippines," *American Economic Review*, 2009, *99* (4), 1245–1277.

Attanasio, Orazio, Abigail Barr, Juan Camilo Cardenas, Garance Genicot, and Costas Meghir, "Risk Pooling, Risk Preferences, and Social Networks," *American Economic Journal: Applied Economics*, 2012, *4* (2), 134–167.

Balakrishnan, Uttara, Johannes Haushofer, and Pamela Jakiela, "How soon is now? Evidence of present bias from convex time budget experiments," *Experimental Economics*, 2020, *23* (2), 294–321.

Barboni, Giorgia, Alessandra Cassar, Arturo Rodriguez Trejo, and Bruce Wydick, "Adverse selection and moral hazard in joint-liability loan contracts: evidence from an artefactual field experiment," *Journal of Economics and Management*, 2013, *9* (2), 153–184.

Barr, Abigail, "Trust and expected trustworthiness: experimental evidence from Zimbabwean villages," *The Economic Journal*, 2003, *113* (489), 614–630.

_ **and Garance Genicot**, "Risk sharing, commitment, and information: an experimental analysis," *Journal of the European Economic Association*, 2008, *6* (6), 1151–1185.

_ **, Justine Burns, Luis Miller, and Ingrid Shaw**, "Economic status and acknowledgement of earned entitlement," *Journal of Economic Behavior & Organization*, 2015, *118*, 55–68.

_ **, Marleen Dekker, and Bereket Kebede**, "Bridging the gender divide: An experimental analysis of group formation in African villages," *World Development*, October 2012, *40* (10), 2063–2077.

_ **,** _ **, and Marcel Fafchamps**, "Who shares risk with whom under different enforcement mechanisms?," *Economic Development and Cultural Change*, 2012, *60* (4), 677–706.

_ **,** _ **, Wendy Janssens, Bereket Kebede, and Berber Kramer**, "Cooperation in Polygynous Households," *American Economic Journal: Applied Economics*, 2018, *11* (2), 266–283.

Bauer, Michal, Alessandra Cassar, Julie Chytilová, and Joseph Henrich, "War's Enduring Effects on the Development of Egalitarian Motivations and In-Group Biases," *Psychological Science*, 2014, *25* (1), 47–57.

_ **,** _ **,** _ **, and** _ , "War's Enduring Effects on the Development of Egalitarian Motivations and In-Group Biases," *Psychological Science*, 2014, *25* (1), 47–57.

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

_ , **Julia Chytilová, and Eric Ochieng**, "Preference Transmission within Churches: Religious Leaders and Clusters of (In)Tolerance," September 2024.

Berg, Joyce, John Dickhaut, and Kevin McCabe, "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 1995, *10* (1), 122–142.

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

Bernardi, Bernardo, *Age Class Systems: Social Institutions and Polities Based on Age*, New York: Cambridge University Press, 1985.

Binswanger, Hans P., "Attitudes Toward Risk: Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 1980, *62*, 395–407.

Blum, Ashley, Chad Hazlett, and Daniel Posner, "Measuring Ethnic Bias: Can Misattribution-Based Tools from Social Psychology Reveal Group Biases that Economics Games Cannot?," *Political Analysis*, 2021, *29*, 1–20.

Breza, Emily, Arun G. Chandrasekhar, Shane Lubold, Tyler H. McCormick, and Mengjie Pan, "Consistently estimating network statistics using aggregated relational data," *Proceedings of the National Academy of Sciences*, 2023, *120* (21).

_ , _ , **Tyler H. McCormick, and Mengjie Pan**, "Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data," *American Economic Review*, 2020, *110* (8), 2454–2484.

_ , **Supreet Kaur, and Yogita Shamdasani**, "The Morale Effects of Pay Inequality," *Quarterly Journal of Economics*, 2018, *133* (2), 611–663.

Brooks, Benjamin, Karla Hoff, and Priyanka Pandey, "Cultural Impediments to Learning to Cooperate: An Experimental Study of High- and Low-Caste Men in Rural India," *Proceedings of the National Academy of Sciences*, 2018, *115* (45), 11385–11392.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

_ , **Bruno Ferman, Stefano Fiorin, Martin Kanz, and Gautam Rao**, "Status Goods: Experimental Evidence from Platinum Credit Cards," *Quarterly Journal of Economics*, 2018, *133* (3), 1561–1595.

Busara Lab, "https://www.busara.global/," August 9 2024.

Cardenas, Juan C., "How Do Groups Solve Local Commons Dilemmas? Lessons from Experimental Economics in the Field," *Environment, Development and Sustainability*, 2000, *2* (3-4), 305–322.

Cardenas, Juan Camilo and Jeffrey Carpenter, "Behavioural Development Economics: Lessons from Field Labs in the Developing World," *Journal of Development Studies*, 2008, *44* (3), 311–338.

_ , **John Stranlund, and Cleve Willis**, "Local Environmental Control and Institutional Crowding-Out," *World Development*, 2000, *28* (10), 1719–1733.

Carpenter, Jeffrey and Erika Seki, "Do Social Preferences Increase Productivity? Field Experimental Evidence from Fishermen in Toyama Bay," *Economic Inquiry*, 2011, *49* (2), 612–630.

Carpenter, Jeffrey P, Amrita G Daniere, and Lois M Takahashi, "Cooperation, trust, and social capital in Southeast Asian urban slums," *Journal of Economic Behavior & Organization*, 2004, *55* (4), 533–551.

_ , _ , and _ , "Social capital and trust in South-East Asian cities," *Urban Studies*, 2004, *41* (4), 853–874.

_ , Stephen Burks, and Eric Verhoogen, "Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games," in "Field Experiments in Economics," Vol. 10, Emerald Group Publishing Limited, 2005, pp. 261–289.

Casey, Katherine, Rachel Glennerster, and Eduard Miguel, "Reshaping Institutions: Evidence on Aid Impacts Using an Preanalysis Plan," *Quarterly Journal of Economics*, 2012, *127* (4), 1755–1812.

Cassar, Alessandra and Y. Jane Zhang, "The competitive woman: Evolutionary insights and cross-cultural evidence into finding the Femina Economica," *Journal of Economic Behavior & Organization*, May 2022, *197*, 447–471.

_ , Andrew Healy, and Carl Von Kessler, "Trust, risk, and time preferences after a natural disaster: experimental evidence from Thailand," *World Development*, June 2017, *94*, 90–105.

_ , Feven Wordofa, and Y. Jane Zhang, "Competing for the benefit of offspring eliminates the gender gap in competitiveness," *Proceedings of the National Academy of Sciences*, 2016, *113* (19), 5201–5205.

_ , Luke Crowley, and Bruce Wydick, "The effect of social capital on group loan repayment: evidence from field experiments," *The Economic Journal*, February 2007, *117* (517), F85–F106.

_ , Pauline Grosjean, and Sam Whitt, "Legacies of violence: trust and market development," *Journal of Economic Growth*, June 2013, *18*, 285–318.

_ , _ , Fatima Jamal Khan, and Miranda Lambert, "Mothers, fathers, and others: Competition and cooperation in the aftermath of conflict," *Journal of Economic Behavior & Organization*, November 2023, *215*, 207–223.

Charness, Gary, Uri Gneezy, and Michael A. Kuhn, "Experimental methods: Extra-Laboratory Experiments-Extending the Reach of Experimental Economics," *Journal of Economic Behavior & Organization*, 2013, *91*, 93–100.

Cilliers, Jacobus, Oeindrila Dube, and Bilal Siddiqi, "The White-Man Effect: How Foreigner Presence Affects Behavior in Experiments," *Journal of Economic Behavior & Organization*, 2015, *118*, 397–414.

Cohn, Alain, Michel Andre Marechal, David Tannenbaum, and Christian Lukas Zund, "Civic Honesty Around the Globe," *Science*, 2019, *365*, 70–73.

Conlon, John J, Malavika Mani, Gautam Rao, Matthew W Ridley, and Frank Schilbach, "Learning in the Household," 2021.

_ , _ , _ , _ , and _ , "Not Learning from Others," 2022.

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

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

Ensminger, Jean and Joseph Henrich, *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*, New York: Russell Sage Foundation Press, 2014.

Evans, David, "Towards improved and more transparent ethics in randomised controlled trials in development social science," *Journal of Development Effectiveness*, 2023.

Fafchamps, Marcel, Bereket Kebede, and Daniel John Zizzo, "Keep up with the winners: Experimental evidence on risk taking, asset integration, and peer effects," *European Economic Review*, 2015, *79*, 59–79.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

Fiala, Nathan and Michele di Maio, "Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of Enumerator Effects," *World Bank Economic Review.*

Flory, Jeffrey A., Uri Gneezy, Kenneth L. Leonard, and John A. List, "Gender, Age, and Competition: A Disappearing Gap?," July 2017.

Gelfand, M. J., J. L. Raver, L. Nishii, L. M. Leslie, J. Lun, B. C. Lim, Z. Aycan et al., "Differences Between Tight and Loose Cultures: A 33-Nation Study," *Science*, 2011, *332* (6033), 1100–1104.

Ghosh, Arkadev, "Religious Divisions and Production Technology: Experimental Evidence from India," 2024. Working paper.

Giné, Xavier, Pamela Jakiela, Dean Karlan, and Jonathan Morduch, "Microfinance Games," *American Economic Journal: Applied Economics*, 2010, *2* (3), 60–95.

Gneezy, Uri, Kenneth L. Leonard, and John A. List, "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *Econometrica*, 2009.

Godoy, Ricardo, Dean Karlan, and Jonathan Zinman, "Randomization for Causality, Ethnography for Mechanisms: Illiquid Savings for Liquor in an Autarkic Society," December 2021. NBER Working Paper No. 29566.

Griffith, Alan, "Name Your Friends, but Only Five? The Importance of Censoring in Peer Effects Estimates Using Social Network Data," *Journal of Labor Economics*, October 2022, *40* (4), 779–1091.

Güth, Werner, R. Wayne Schmittberger, and Bernd Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 1982, *3*, 367–388.

Habyarimana, James, Macartan Humphreys, Daniel Posner, and Jeremy Weinstein, "Why does ethnic diversity undermine public goods provision?," *American Political Science Review*, 2007, *101* (4), 709–725.

_ , _ , _ , and _ , *Coethnicity: Diversity and the Dilemmas of Collective Action*, New York, NY: Russell Sage Foundation Press, 2009.

Harrison, Glenn W. and John A. List, "Field Experiments," *Journal of Economic Literature*, December 2004, *42*, 1009–1055.

_ **and** _ , "Naturally Occurring Markets and Exogenous Laboratory Experiments: A Case Study of the Winner's Curse," *Economic Journal*, 2008, *118* (528), 822–843.

Haushofer, Johannes, Magdalena Larreboure, Sara Lowes, and Leon Mait, "Cash Transfers and Social Preferences of Children," 2024.

_ **, Marie Collins, Giovanna de Giusti, Joseph Njoroge, Amos Odero, Cynthia Onyango, James Vancel, Chaning Jang, Maneesh Kuruvilla, and Conor Hughes**, "A Methodology for Laboratory Experiments in Developing Countries: Examples from the Busara Center," February 2014. Working Paper, First posted: 1 Oct 2012.

_ **, Robert Mudida, and Jeremy P Shapiro**, "The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being," 2020.

_ **, Sara Lowes, Abednego Musau, David Ndetei, Nathan Nunn, Moritz Poll, and Nancy Qian**, "Stress, Ethnicity, and Prosocial Behavior," *Journal of Political Economy Microeconomics*, May 2023, *1* (2).

Hazlett, Chad, Jennifer Hamilton, and Dan Posner, "Uncovering Ethnic Discrimination: Using Distraction to DetectHidden Bias in Economics Games," 2024. Working Paper, UCLA.

Heldring, Leander, "The Origins of Violence in Rwanda," *Review of Economic Studies*, 2021, *88* (2), 730–763.

Henrich, Joseph, "Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon," *American Economic Review*, 2000, *90* (4), 973–979.

_ **and Natalie Henrich**, "The Evolution of Cultural Adaptations: Fijian Taboos during Pregnancy and Lactation Protect against Marine Toxins," *Proceedings of the Royal Society B: Biological Sciences*, 2010, *277* (1701), 3715–3724.

_ **, Jean Ensminger, Richard McElreath, Abigail Barr, H Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank W Marlowe, David Tracer, and John Ziker**, "Markets, Religion, Community Size and the Evolution of Fairness and Punishment," *Science*, 2010, *327* (5972), 1480–1484.

_ , _ , _ , _ **, H. Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank W. Marlowe, David Tracer, and John Ziker**, "Markets, religion, community size and the evolution of fairness and punishment," *Science*, 2010, *327* (5972), 1480–1484.

_ **, R. Boyd, S. Bowles, H. Gintis, E. Fehr, C. Camerer, R. McElreath, M. Gurven, K. Hill, A. Barr, J. Ensminger, D. Tracer, F. Marlow, J. Patton, M. Alvard, F. Gil-White, and N. Henrich**, "'Economic Man' in Cross-Cultural Perspective: Ethnography and Experiments from 15 Small-Scale Societies," *Behavioral and Brain Sciences*, 2005, *28*, 795–855.

_ **, Richard McElreath, Abigail Barr, Jean Ensminger, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank Marlowe, David Tracer, and John Ziker**, "Costly Punishment Across Human Societies," *Science*, 2006, *312*, 1767–1770.

_ , _ , _ , _ **, H Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, Carolyn Lesorogol, Frank W Marlowe, David Tracer, and John Ziker**, "Costly punishment across human societies," *Science*, 2006, *312* (5781), 1767–1770.

_ , **Robert Boyd, Sam Bowles, Colin Camerer, Herbert Gintis, Richard McElreath, and Ernst Fehr**, "In Search of Homo Economicus: Experiments in 15 Small-Scale Societies," *American Economic Review*, 2001, *91* (2), 73–79.

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

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

Hoff, Karla, Mayuresh Kshetramade, and Ernst Fehr, "Caste and Punishment: The Legacy of Caste Culture in Norm Enforcement," *The Economic Journal*, November 2011, *121* (556), F449–F475.

Iversen, Vegard, Cecile Jackson, Bereket Kebede, Alice Munro, and Arjan Verschoor, "Do Spouses Realise Cooperative Gains? Experimental Evidence from Rural Uganda," *World Development*, 2011, *39* (4), 569–578.

Jackson, Conrad, Michele Gelfand, and Carol R. Ember, "A global analysis of cultural tightness in non-industrial societies," *Proceedings of the Royal Society B: Biological Sciences*, 2020, *287* (1930).

Jakiela, Pamela, "Social Preferences and Fairness Norms as Informal Institutions: Experimental Evidence," *American Economic Review*, 2011, *101* (3), 509–513.

_ , "How fair shares compare: Experimental evidence from two cultures," *Journal of Economic Behavior & Organization*, 2015, *118*, 40–54.

_ **and Owen Ozier**, "Does Africa Need a Rotten Kin Theorem? Experimental Evidence from Village Economies," *Review of Economic Studies*, 2016, *83* (1), 231–268.

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

_ **and** _ , "The Impact of Violence on Individual Risk Preferences: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 2019, *101* (3), 547–549.

_ , **Edward Miguel, and Vera L. te Velde**, "You've earned it: estimating the impact of human capital on social preferences," *Experimental Economics*, 2015, *18*, 385–407.

Johnson, Noel and Alexandra Mislin, "Trust Games: A Meta-Analysis," *Journal of Economy Psychology*, 2011, *32*, 865–889.

Kanu, Ikechukwu Anthony, "The Ecological Value of Igbo Spirituality," *Harvard Divinity Bulletin*, 2023, (Spring/Summer), 47–54.

Karaja, Elira and Jared Rubin, "The Cultural Transmission of Trust Norms: Evidence from a Lab in the Field on a Natural Experiment," *Journal of Comparative Economics*, 2022, *50* (1), 1–19.

Kebede, Bereket and Daniel John Zizzo, "Social Preferences and Agricultural Innovation: An Experimental Case Study from Ethiopia," *World Development*, 2015, *67*, 267–280.

_ , **Maria Tarazona-Gomez, Alistair Munro, and Arjan Verschoor**, "Intra-Household Efficiency: An Experimental Study from Ethiopia," *Journal of African Economies*, 2013, *23* (1), 105–150.

Kremer, Michael, Gautam Rao, and Frank Schilbach, "Behavioral Development Economics," in "Handbook of Behavioral Economics: Foundations and Applications 2," Vol. 2, Elsevier, 2019, pp. 345–458.

Le Rossignol, Etienne, Sara Lowes, and Nathan Nunn, "The Social Consequences of Traditional Religion in Contemporary Africa," 2024.

Lowe, Matt, "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration," *American Economic Review*, 2021, *111* (6), 1807–1844.

Lowes, Sara, "Ethnographic and Field Data in Historical Economics," in Alberto Bisin and Giovanni Federico, eds., *The Handbook of Historical Economics*, Elsevier, 2021, chapter 6.

_ , "Kinship Structure, Stress, and the Gender Gap in Competition," *Journal of Economic Behavior and Organization*, December 2021, *192*, 36–57.

_ , "Matrilineal Kinship and the Family: Evidence from the Matrilineal Belt," 2022. Mimeo.

_ **and Eduardo Montero**, "Concessions, Violence, and Indirect Rule: Evidence from the Congo Free State," *Quarterly Journal of Economics*, 2021, *136* (4), 2047–2091.

_ **, Nathan Nunn, James A. Robinson, and Jonathan L. Weigel**, "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom," *Econometrica*, 2017, *85* (4), 1065–1091.

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

Marlowe, Frank, J. Colette Berbesque, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Jean Ensminger, Michael Gurven, Edwins Gwako, Joseph Henrich, Natalie Henrich, Carolyn Lesorogol, Richard McElreath, David Tracer, and John Ziker, "More 'altruistic' punishment in larger societies," *Proceedings of the Royal Society B: Biological Sciences*, 2008, *275* (1634), 587–590.

Mayo, Elton, *The Human Problems of an Industrial Civilization*, New York: Macmillan, 1933.

Medvedev, Dmitry, David Davenport, Thomas Talhelm et al., "The Motivating Effect of Monetary over Psychological Encentives is Stronger in WEIRD Cultures," *Nature Human Behaviour*, 2024, *8*, 456–470.

Moscona, Jacob and Awa Ambra Seck, "Age Set vs. Kin: Culture and Financial Ties in East Africa," *American Economic Review*, forthcoming.

_ **, Nathan Nunn, and James A. Robinson**, "Keeping it in the family: Lineage organization and the scope of trust in Sub-Saharan Africa," *American Economic Review: Papers and Proceedings*, 2017, *107* (5), 565–571.

_ **,** _ **, and James A. Robinson**, "Social Structure and Conflict: Evidence from Sub-Saharan Africa," *Econometrica*, 2020, *88* (5).

Mummolo, Jonathan and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," *American Political Science Review*, 2019, *113* (2), 517–529.

Munro, Alistair, Bereket Kebede, Maria Tarazona-Gomez, and Arjan Verschoor, "Autonomy and Efficiency: An Experiment on Household Decisions in Two Regions of India," *Journal of the Japanese and International Economies*, 2014, *33*, 114–133.

Negussie, Yared Molla, Ekko C. van Ierland, Xueqin Zhu, Bereket Kebede, and Jan Klopm, "Implications of smallholders' risk preferences for climate innovation adoption: Evidence from field experimental games," 2018. In preparation.

Nisbett, Richard E., *The Geography of Thought: How Asians and Westerners Think Differently...and Why*, New York: Free Press, 2004.

Porteous, Obie, "Research Deserts and Oases: Evidence from 27 Thousand Economics Journal Articles on Africa," *Oxford Bulletin of Economics and Statistics*, 2022, *84* (6), 1235–1258.

Rao, Gautam, "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools," *American Economic Review*, 2019, *109* (3), 774–809.

Roethlisberger, F. J. and W. J. Dickson, *Management and the Worker*, Cambridge, MA: Harvard University Press, 1939.

Roth, Alvin, Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study," *American Economic Review*, 1991, *81*, 1068–1095.

Rustagi, Devesh, "Historical Self-Governance and Norms of Cooperation," *Econometrica*, Forthcoming.

_ **and Markus Kroell**, "Measuring Honesty and Explaining Adulteration in Naturally Occurring Markets," *Journal of Development Economics*, 2022, *156*, 102819.

Siddique, Abu, Michael Vlassopoulos, and Yves Zenou, "Leveraging Edutainment and Social Networks to Foster Interethnic Harmony," 2024. Working paper.

Stacy, Brian, Lucas Kitzmuller, Xiaoyu Wang, Daniel Gerszon Mahler, and Umar Serajuddin, "Missing Evidence: Tracking Academic Data Use around the World," 2024. World Bank Policy Research Working Paper 10673.

Tempels, Placid, *Bantu Philosophy*, Paris: Présence Africaine, 1959.

Thaler, Richard H., "Anomalies: The Ultimatum Game," *Journal of Economic Perspectives*, 1988, *2* (4), 195–206.

Weigel, Jonathan, "The Participation Dividend of Taxation: How Citizens in Congo Engage More with the State when it Tries to Tax Them," *Quarterly Journal of Economics*, 2020, *135* (4), 1849–1903.

Xiaotong, Fei, *From the Soil: The Foundations of Chinese Society*, Oakland: University of California Press, 1992.

Zizzo, Daniel. J., "Experimenter Demand Effects in Economic Experiments," *Experimental Economics*, 2010, *13* (1), 75–98.