DISCUSSION PAPER SERIES

Using Survey Questions to Measure Preferences: Lessons from an Experimental Validation in Kenya

Michal Bauer
Julie Chytilová
Edward Miguel

Using Survey Questions to Measure Preferences: Lessons from an Experimental Validation in Kenya

Michal Bauer
Charles University ,CERGE-EI and IZA

Julie Chytilová
Charles University and CERGE-EI

Edward Miguel
University of California, Berkeley

ABSTRACT

Using Survey Questions to Measure Preferences: Lessons from an Experimental Validation in Kenya*

Can a short survey instrument reliably measure a range of fundamental economic preferences across diverse settings? We focus on survey questions that systematically predict behavior in incentivized experimental tasks among German university students (Becker et al. 2016) and were implemented among representative samples across the globe (Falk et al. 2018). This paper presents results of an experimental validation conducted among low-income individuals in Nairobi, Kenya. We find that *quantitative* survey measures - hypothetical versions of experimental tasks - of time preference, attitude to risk and altruism are good predictors of choices in incentivized experiments, suggesting these measures are broadly experimentally valid. At the same time, we find that *qualitative* questions - self-assessments - do not correlate with the experimental measures of preferences in the Kenyan sample. Thus, caution is needed before treating self-assessments as proxies of preferences in new contexts.

JEL Classification:	C83, D90
Keywords:	preference measurement, experiment, survey, validation

Corresponding author:
Michal Bauer
Charles University
Politických vězňů 7
111 21 Prague 1
Czech Republic
E-mail: bauer@cerge-ei.cz

Introduction

Fundamental preferences in the economic domain, such as time discounting and risk preferences, and in the social domain, such as altruism, reciprocity and spitefulness, constitute key elements of individual decision-making. Figuring out ways to accurately measure these preferences among large samples in the field holds considerable promise since doing so may shed light on the sources of vast differences in preferences observed across individuals and societies, and their role in fundamental economic choices and societal trajectories. While measuring preferences using incentivized tasks is generally considered the gold standard,[1] implementing incentivized tasks among large samples outside of the controlled environment of an experimental laboratory is often infeasible, given that they are relatively expensive and time consuming. Consequently, a potentially attractive alternative is to employ survey questions instead of incentivized experiments, but there has long been widespread concern that non-incentivized self-reported survey measures of preferences may not reliably capture real life choices.

To tackle this important methodological trade-off, Falk et al. (2018) have recently developed a short (7-8 minutes) innovative survey module, designed to measure a wide range of economic preferences. It has been implemented among representative samples of subjects in more than seventy countries (Falk et al., 2018), creating the most comprehensive global data set with comparable measures of preferences, namely, the Global Preference Survey (GPS). Measures of preferences in each domain are constructed as a weighted average based on one objective quantitative item -- a hypothetical version of an experimental task -- and one subjective qualitative item that measures self-reported willingness to act in a certain way.

To establish the validity of the survey preference measures, Becker, Dohmen, Huffman, Falk, & Sunde (2016) perform a careful experimental validation of the survey questions, and document that survey measures of preferences do predict choices in incentivized decisions. The validation was conducted among students at University of Bonn, Germany. Given the wide coverage of the existing GPS data set and the convenience of the survey module in terms of implementation,[2] it has the potential to become a widely adopted instrument for (i) studying differences in preferences across societies and their relationships with economic outcomes, (ii) employing preference measures as control variables when identifying causal

---

[1] Experimental measures of preferences have been shown to predict a wide range of real-life behavior (e.g., Ashraf, Karlan, & Yin, 2006; Burks, Carpenter, Goette, & Rustichini, 2009; Meier & Sprenger, 2010; Rustagi, Engel, & Kosfeld, 2010; Sutter, Kocher, Glätzle-Rützler, & Trautmann, 2013).

[2] There is a laudable public good element in the GPS project. The global data set, as well as the survey instrument - and its 116 versions for 70 countries and 78 different languages - are readily available to researchers at https://www.briq-institute.org/global-preferences/home. Our validation experiment benefited greatly from this transparency, as we build on the Swahili translation of the survey module for Kenya.

effects of other factors correlated with preferences, and (iii) as outcome variables in new randomized controlled trials aiming to uncover the effects of various interventions on individual preferences.[3]

This paper adds to these efforts and aims to be useful in three ways. First, we test the experimental validity of the survey questions outside of a sample of university students from a rich country, by focusing on a sample from the other end of the global distribution of income and education. Our experimental subjects are residents of working class neighborhoods (sometimes referred to as "slums") in Nairobi, Kenya, a setting with different set of institutions and economic constraints. The participants are aged between 20-46, with average income of around USD 3 per day, and 54% are unemployed. Establishing the experimental validity of the measures among this subject pool is important for several reasons. Most of humanity lives in low and middle income countries, outside of Western, Educated, Industrialized, Rich and Democratic societies (Henrich et al. 2010), in which the original GPS validation was conducted. Next the GPS module is particularly suitable to be integrated into large-scale follow-up surveys in randomized control trials, which are routinely implemented by development economists (Banerjee and Duflo 2012), often in Africa, and thus knowledge of whether the survey preference measures predict incentivized behavior among low-income individuals in Kenya is a useful input for scholars considering the adoption of these measures.[4]

Second, comparing the results of analogous validations conducted in Germany and Kenya is methodologically interesting, because measures of economic preferences in GPS are derived from both objective quantitative tasks as well as subjective qualitative questions, based on self-assessments.[5] There is a legitimate concern that subjective self-assessments might be understood and interpreted in different ways across countries, which can attenuate their ability to uncover personality traits and complicate cross-country comparisons. For example, the Big Five measures of personality traits, the most widely-used method to measure and classify personality traits in psychology, are based on self-assessments, and recent attempts to validate the Big Five measures have failed to reliably predict the intended personality traits in low- or

---

[3] To date, the GPS measures have been used to explore global variations of preferences and their relationships with country-level and individual-level characteristics (Falk et al., 2018), deep historical origins of variation of preferences (Becker, Enke, and Falk 2018) and the relationships between economic development and gender differences in preferences (Falk & Hermle, 2018).

[4] Indeed, this experimental validation itself took place as a part of a larger project that aims to estimate the long-term effects on individual preferences of a randomized public health intervention (a school-based deworming program) which took place in Western Kenya in the late 1990s (Baird et al. 2016; Miguel and Kremer 2004). We used lessons from the current validation exercise in the design of a preference survey module that is integrated into the most recent round of follow-up data collection (Kenyan Life Panel Survey, KLPS, round 4).

[5] An example of a qualitative question from GPS would be "Please tell me, in general, how willing or unwilling you are to take risks, using a scale from 0 to 10", or "How willing are you to give to a charity without expecting anything in return?"

middle-income countries, in contrast to samples from the wealthy countries for which they were originally developed (Laajaj et al. 2019; Gurven et al. 2013). An advantage of GPS is that, besides self-assessments, it also contains quantitative questions that are arguably less subject to this issue, because they directly define the parameters and nature of the decision and more closely mirror the incentivized experimental task. Thus, we can test whether quantitative questions are relatively more robust predictors of actual incentivized behavior across two diverse settings, as compared to qualitative self-assessments.

Third, we place additional emphasis on the types of preferences that are likely to be especially important in settings with low social capital and a history of inter-group conflict, issues that are particularly pressing in low-income countries (Blattman and Miguel 2010). While pro-social preferences, such as altruism and positive reciprocity, help to establish and maintain cooperative and fair group outcomes even in situations with limited scope for reputation-building (Bowles 2006; Fehr and Fischbacher 2003), anti-social preferences (such as spitefulness and aggressive competitiveness) can contribute to the deterioration of co-operation (Falk, Fehr, & Fischbacher, 2005; Herrmann, Thoni, & Gachter, 2008).[6] Furthermore, ethnic biases in social preferences -- in-group favoritism and out-group hostility -- create fertile ground for violent inter-group conflict. While the GPS focuses on measuring preferences relevant for explaining positive aspects of human social behavior, such as generalized altruism and reciprocity, we also assess the experimental validity of survey questions designed to measure the dark side of human social behavior. Specifically, we test the validity of questions designed to uncover anti-social preferences, such as spite, and distinguish between generalized, in-group, and out-group preferences, along both prosocial and anti-social dimensions.

Experimental design

The sample in our study are 123 subjects from the Kibera neighborhood in Nairobi, Kenya. The participants come from a low-income environment, are between 20 and 46 years of age, more than half are unemployed, half are women and, on average, they have two children (Table A1). The experiments were implemented in a state-of-the-art experimental economics laboratory in the Busara Center for Behavioral Economics (Haushofer et al. 2014).

---

[6] Anti-social preferences – malevolent willingness to harm others at a cost to self – have been shown to be relatively widespread in numerous settings in both high and low income settings (Abbink and Sadrieh 2009; Fehr, Hoff, and Kshetramade 2008; Herrmann, Thoni, and Gachter 2008; Prediger, Vollan, and Herrmann 2014; Bauer, Cahlíková, Chytilová, et al. 2018; Bauer, Cahlíková, Katreniak, et al. 2018).

Subjects were invited to the lab twice, for visits one week apart, where the time gap was introduced in order to minimize spillovers between the survey and experimental measures. During one visit, they made choices in a set of incentivized experiments, while during the other, they answered non-incentivized survey questions. The order of experiments/survey was randomized on an individual level. We elicited measures of the following types of preferences: (i) time discounting, (ii) risk preference, (iii) ambiguity aversion, (iv) altruism (generalized, in-group, and out-group), (v) anti-social behavior (generalized, in-group, and out-group), and (vi) positive reciprocity.

The experimental choices involved high stakes, in order to capture decision situations with substantial financial consequences for the subjects. Specifically, each subject received a show-up fee (KSh 450 for the survey part and KSh 250 for the experimental part, where 100 KSh was roughly equal to 1 USD during the study period) and a payoff determined by one randomly selected choice made in the experimental part. The average payoff from experiments was KSh 820, i.e., the equivalent of approximately 2.5 days' typical earnings. Each type of preference was elicited using one experimental task. The full experimental protocol is available in the Online Appendix C and D.

For time discounting, subjects made 25 binary choices between an immediate payment or a larger payment with a three-month delay, which was increased by a fixed amount in each subsequent binary choice, using a multiple price list. Similarly, when eliciting risk preference, subjects made 21 binary choices between a lottery that yielded a positive amount or zero with equal probability, and a safe payment option that increased in each subsequent binary choice. Ambiguity aversion was measured by a binary choice between two bags – one with a known and one with an unknown composition of differently colored balls, with the payoff determined by drawing a ball of a specific color.

In the experiments focusing on the social domain, altruism was measured by the choice of how much of an endowment the participant decided to donate to a charity. One choice measured donations to a charity which helps people in Kenya (generalized altruism), the second choice elicited donations to a charity which helps people from the participant's ancestral home area (in-group altruism), and the third elicited donations to a charity which helps people in Kenya outside of the subject's own ancestral home area (out-group altruism). Anti-social behavior was measured using a binary choice in which subjects could decide to reduce the payoff of another person by sacrificing a part of their own payoff. Again, we implemented three versions, using the same wording as above to indicate generalized, in-group, and out-group versions of the task. Finally, positive reciprocity was measured by the amount of money given to a person who had been kind to the participant. This person was an anonymous participant in a different, earlier experiment in the lab who decided to leave a gift (a bag of sugar, which is a popular gift item in the setting we study) for a future visitor of the lab (i.e., decision-maker of our study), instead of keeping all the sugar for him or

herself. As an alternative measure of reciprocity, we used the difference in the amount donated to this (kind) person and to another (unkind) person who had decided not to give any sugar.

In the survey part, we elicited one objective quantitative measure and one or two subjective qualitative measures for each type of preference.[7] The quantitative questions presented a hypothetical scenario that mimicked the experimental task. For time and risk preferences, instead of asking the full set of questions as in the experiment, we used the "staircase" or "unfolding brackets" procedure, in which each participant answers a sub-set of five binary choices chosen based on their answer to the previous question. The qualitative questions measure self-reported willingness to act in a certain way on a 0-10 scale. Specifically, respondents rate their own willingness to give up something that is beneficial today in order to benefit more in the future (time discounting), to take risks (risk preference), to give to a charity and to share with others (altruism), to cause trouble for other people and to do harm to other people (anti-social preferences), and to return a favor (reciprocity).

Note that the experimental validation in Nairobi is closely comparable to, but not strictly identical to, the preference measure validation conducted in Bonn (Becker et al. 2016). Some of the experimental tasks had to be simplified, reflecting the differences in average schooling between the Kenyan and German subject pools. We also slightly adjusted the wording in some of the GPS survey questions, based on feedback from piloting and focus-group discussions, in order to improve comprehension in the Kenyan context. In terms of procedure and data analysis, we use a similar approach as Becker et al. (2016). Please see the Online Appendix for details of each experimental task, questions used and the comparison of the Kenyan and German validation exercises.

To start, we observe that the elicited preference measures have several desirable properties (see Online Appendix Table A1 for summary statistics). First, there is substantial variation in all our measures of preferences, both survey and experimental, attenuating concerns that a failure to identify relationships between variables of interest could be mechanically driven by a lack of variation. Second, behavior in the experiments is largely comparable to previous studies. For example, in the generalized version of the dictator game (altruism measure), we observe that subjects allocate around 20% of their endowment to charity. We also find that subjects are significantly more willing to give to a charity that helps their own ethnic group, as compared to a charity that helps out-group members. Similarly, slightly fewer than 20% reduce another person's income at a cost to themselves, which is comparable to the prevalence of anti-

---

[7] The only exception is ambiguity aversion, for which there is only one quantitative survey measure.

6

social behavior in other settings (Abbink and Sadrieh 2009; Prediger, Vollan, and Herrman 2014), and subjects are significantly more destructive towards out-group members.

Results

We begin by describing the predictive power of objective quantitative survey measures. For each survey item, Table 1 displays an OLS coefficient from a regression of the standardized experimental measure on the standardized survey item (column 1) and the Spearman correlation between the survey item and a respective experimental incentivized preference measure (column 2). We find that the quantitative survey measures of time preference, attitude to risk, generalized altruism, altruism towards one's own ethnic group, and altruism towards out-group members are strongly positively correlated with experimental measures, and the observed relationships are statistically significant. The quantitative survey measure of ambiguity aversion and all three measures of anti-social behavior correlate weakly with the experimental measure: the correlations for all are relatively small in magnitude and none is significant at traditional levels.

Specifically, in terms of magnitudes, the correlations are 0.40 for time discounting, 0.25 for risk preference, 0.29 for positive reciprocity, 0.41 for generalized altruism, 0.36 for in-group altruism and 0.38 for out-group altruism. These correlations are slightly lower than, though comparable to the correlation generated in the validation of the same set of survey preference measures in Germany (Becker et al. 2016), reported for comparison in column 3, in which the corresponding correlations were found to be 0.55 (time discounting), 0.34 (risk taking), 0.35 (positive reciprocity) and 0.39 (generalized altruism). We speculate that the somewhat smaller correlations in Kenya may potentially reflect greater measurement error in the elicitation of preferences among a subject pool with lower average schooling levels.

The observed patterns are robust to controlling for the level of understanding, based on direct cross-check questions, and violations of monotonicity (in tasks eliciting time and risk preferences, which use multiple price lists), an indirect proxy of understanding. The correlations are also similar for different order of the survey and experimental tasks (whether they were elicited during the first or second week), and robust to controlling for a set of basic individual characteristics (i.e., age, gender, being unemployed, and the number of children); the results of these robustness checks are presented in Online Appendix Table A2.

Further, we address a concern which is inherent in this type of experimental validation, namely, that subjects may remember their choices from the previous week and choose the same options in the second week in order to appear consistent over time. To address this, we included an independent task to measure a subject's memory. Specifically, in the first week, the participants were shown a set of ten letters on a

screen for twenty seconds and were incentivized to remember those letters for a short period. In the second week, they were asked to recall these ten letters, again in an incentive-compatible way. We show that the correlations observed between experimental and survey measures of preferences are not driven by subjects with more accurate recall (those remembering above the median number of letters), with the exception of the time preference measure (Table A3).

Next, we explore the predictive power of the subjective survey self-assessments. In contrast to the objective survey measures, qualitative survey measures are rather poor predictors of the experimental measures of preferences (Table 2). None of the correlations reaches statistical significance at conventional levels when we use the Spearman correlation (column 2), and this is unlikely to be due to lack of power. The estimated coefficients are close to zero and in many cases do not have the expected sign: nine estimated correlations have expected signs, while seven have an opposite sign than predicted. In terms of magnitudes, none of the estimated 16 correlations is larger than 0.15. We arrive at a similar conclusion when using OLS regressions (column 1), with the exception of measures of positive reciprocity and out-group altruism, for which we find positive coefficients (0.21 and 0.18, resp.), significant at the 5% level.

The overall pattern of (non)results differs sharply from the patterns observed in the German validation exercise, where subjective self-assessments reliably predict behavior in experimental tasks (column 3): all estimated coefficients are statistical significant, have the expected sign, and the magnitudes are at least 0.3, similar to the quantitative survey measures.

Concluding remarks

An experimental validation of survey preference measures in Kenya has several noteworthy implications. First, our results should boost confidence in the ability of *objective quantitative* GPS survey measures of preferences, based on hypothetical tasks, to predict high-stakes incentivized behavior in experiments designed to measure range of preferences. We show that quantitative survey measures of time preference, attitude to risk and generalized, in-group, and out-group altruism are good predictors of choices in incentivized experiments among residents of a working class Nairobi neighborhood. This finding reinforces the findings from a similar validation exercise performed among a sample of university students in Germany (Becker et al. 2016), and thus, together, the two studies document the experimental validity of these measures across culturally diverse settings at opposite ends of the global income and education distribution.

At the same time, we find that the predictive power of *subjective qualitative* questions on preferences do not meaningfully correlate with the experimental measures in the Kenyan sample, in contrast

to the German sample. Thus, although in rich (mostly German) settings qualitative survey questions have been shown to do a good job of predicting behavior in incentivized experiments (Dohmen et al. 2011; Becker et al. 2016) and a range of real-life behaviors (Barasinska, Schaefer, and Stephan 2012; Bauernschuster et al. 2014; Bonin et al. 2007; Fouarge, Kriechel, and Dohmen 2014; Jaeger et al. 2010; Dohmen et al. 2011), caution is needed before interpreting these measures as proxies of preferences in all contexts, especially low-income settings. Based on our evidence alone, it is unclear how widely the validation results from Germany will apply, and assessing this should be an important focus of future work. Since our findings suggest that the experimental validity of subjective self-assessments are likely to be economically and culturally specific[8], we hope they will motivate implementation of series of comparable validation exercises in other diverse settings across the globe, in order to better understand the characteristics of societies for which the qualitative self-assessment are informative. Future research may also need to determine whether alternative formulations of qualitative questions are more reliable than current self-assessments.

Finally, this study also tested the experimental validity of survey preference measures in a new domain, anti-social preferences, which is arguably most prone to social desirability biases. We document that survey measures of anti-social preferences only weakly predict incentivized behavior, which strengthens the case for investing resources into gathering incentivized measures in this domain.

References

Abbink, Klaus, and Abdolkarim Sadrieh. 2009. "The Pleasure of Being Nasty." *Economics Letters* 105 (3): 306–8.

Ashraf, N., D. Karlan, and W. Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics* 121 (2): 635–672.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2016. "Worms at Work: Long-Run Impacts of a Child Health Investment." *Quarterly Journal of Economics* 131 (4): 1637–80.

Banerjee, Abhijit V., and Esther Duflo. 2012. *Poor Economics*. Public Affairs.

Barasinska, N., D. Schaefer, and A. Stephan. 2012. "Individual Risk Attitudes and the Composition of Nancial Portfolios: Evidence from German Household Portfolios." *Quarterly Review of Economics and Finance* 52 (1): 1–14.

---

[8] In contrast to rich country settings, validation studies conducted in low income countries are still rare and typically focus on measures of a single preference type, specifically on risk preference. Following up on Dohmen et al. (2011), who conducted a validation experiment in Germany, positive correlations between survey and experimental measures of risk preference were documented in rural Thailand (Hardeweg, Menkhoff, & Waibel 2013) and among Chinese students (Ding, Hartog, and Sun 2010). A recent cross-cultural study on risk-taking from 30 countries (Vieider et al. 2015) documents that qualitative survey measures are positively correlated with choices in incentivized experiments in a majority of cases, but the magnitude and statistical significance of the correlation varies substantively across countries, which also suggests that the experimental validity of qualitative survey measures may be context specific.

Bauer, Michal, Jana Cahlíková, Julie Chytilová, and Tomáš Želinský. 2018. "Social Contagion of Ethnic Hostility." *Proceedings of the National Academy of Sciences* 115 (19): 4881–86.

Bauer, Michal, Jana Cahlíková, Dagmara Celik Katreniak, Julie Chytilová, Lubomír Cingl, and Tomáš Želinský. 2018. "Antisocial Behavior in Groups." CEPR Discussion Paper 13315.

Bauernschuster, S., O. Falck, S. Heblich, J. Suedekum, and A. Lameli. 2014. "Why Are Educated and Risk-Loving Persons More Mobile across Regions?" *Journal of Economic Behavior and Organization* 98: 56–69.

Becker, Anke, Thomas Dohmen, David Huffman, Armin Falk, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Time, Risk, and Social Preferences." IZA Discussion Paper 9674.

Becker, Anke, Benjamin Enke, and Armin Falk. 2018. "Ancient Origins of the Global Variation in Economic Preferences." Working Paper.

Blattman, C., and E. Miguel. 2010. "Civil War." *Journal of Economic Literature* 48 (1): 3–57.

Bonin, H., Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. 2007. "Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes." *Labour Economics* 14 (6): 926–37.

Bowles, Samuel. 2006. *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press Princeton, NJ.

Burks, S. V, J. P Carpenter, L. Goette, and A. Rustichini. 2009. "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment." *Proceedings of the National Academy of Sciences* 106 (19): 7745.

Ding, X., J. Hartog, and Y. Sun. 2010. "Can We Measure Risk Attitudes in a Survey?" *IZA Discussion Paper 4807*.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, J. Schupp, and G. Wagner. 2011. "Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences." *Journal of the European Economic Association* 9 (3): 522–50.

Falk, A., E. Fehr, and U. Fischbacher. 2005. "Driving Forces behind Informal Sanctions." *Econometrica* 73 (6): 2017–2030.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *Quarterly Journal of Economics* 133 (4): 1645–92.

Falk, Armin, and Johannes Hermle. 2018. "Relationship of Gender Differences in Preferences to Economic Development and Gender Equality." *Science* 362 (6412).

Fehr, E., and U. Fischbacher. 2003. "The Nature of Human Altruism: Proximate Patterns and Evolutionary Origins." *Nature* 425: 785–91.

Fehr, E., K. Hoff, and M. Kshetramade. 2008. "Spite and Development." *American Economic Review* 98 (2): 494–499.

Fouarge, D., B. Kriechel, and Thomas Dohmen. 2014. "Occupational Sorting of School Graduates: The Role of Economic Preferences." *Journal of Economic Behavior and Organization* 106: 335–51.

Gurven, Michael, Christopher von Rueden, Maxim Massenkoff, Hillard Kaplan, and Marino Lero Vie. 2013. "How Universal Is the Big Five? Testing the Five-Factor Model of Personality Variation Among Forager–Farmers in the Bolivian Amazo." *Journal of Personality and Social Psychology* 104 (2): 354–70.

Hardeweg, B., L. Menkhoff, and H. Waibel. 2013. "Experimentally- Validated Survey Evidence on Individual Risk Attitudes in Rural Thailand." *Economic Development and Cultural Change* 61 (4): 859–88.

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

Henrich, J., S.J. Heine, A. Norenzayan, and others. 2010. "The Weirdest People in the World." *Behavioral and Brain Sciences* 33 (2–3): 61–83.

Herrmann, B., C. Thoni, and S. Gachter. 2008. "Antisocial Punishment across Societies." *Science* 319 (5868): 1362.

Jaeger, D., Thomas Dohmen, Armin Falk, David Huffman, Uwe Sunde, and H. Bonin. 2010. "Direct Evidence on Risk Attitudes and Migration." *Review of Economics and Statistics* 92 (3): 684–89.

Laajaj, Rachid, Karen Macours, Daniel Alejandro Pinzon Hernandez, Omar Arias, Samuel Gosling, Jeff Potter, Marta Rubio-Codina, and Renos Vakis. 2019. "Challenges to Capture the Big Five Personality Traits in Non-WEIRD Population." *Scientific Advances*, no. forthcoming.

Meier, S., and C. Sprenger. 2010. "Present-Biased Preferences and Credit Card Borrowing." *American Economic Journal: Applied Economics* 2 (1): 193–210.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.

Prediger, Sebastian, Bjorn Vollan, and Benedikt Herrman. 2014. "Resource Scarcity and Antisocial Behavior." *Journal of Public Economics*, mimeo, 119: 1–9.

Prediger, Sebastian, Bjorn Vollan, and Benedikt Herrmann. 2014. "Resource Scarcity and Antisocial Behavior." *Journal of Public Economics* 119: 1–9.

Rustagi, Devesh, Stefanie Engel, and Michael Kosfeld. 2010. "Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management." *Science* 330 (6006): 961–65.

Sutter, Matthias, Martin G. Kocher, Daniela Glätzle-Rützler, and Stefan T. Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *The American Economic Review* 103 (1): 510–31.

Vieider, F., M. Lefebre, R. Bouchouicha, T. Chmura, R. Hakimov, M. Krawczyk, and P. Martinsson. 2015. "Common Components of Risk and Uncertainty Attitudes across Contexts and Domains: Evidence from 30 Countries." *Journal of the European Economic Association* 13 (1): 421–52.

Table 1: Correlations between quantitative survey measures and experimental measures

Preference		Quantitative survey item	Kenya: Kibera residents		Germany: Bonn students	
			OLS Coefficient	Correlation	Correlation	Measures
			(1)	(2)	(3)	(4)
Time		Staircase measure: 5 interdependent choices between an early and delayed amount of money	0.33***	0.40***	0.55***	comparable
Risk		Staircase measure: 5 interdependent choices between a lottery and varying safe options	0.21**	0.25***	0.34***	comparable
Ambiguity aversion		Hypothetical choice between a bag with known and unknown number of balls of different color	0.13	0.13	n.a.	
Reciprocity		Hypothetical choice of the amount of a gift given to a stranger who provided help	0.12	0.29***	0.35***	experimental different; survey comparable
Reciprocity (diff)		Hypothetical choice of the amount of a gift given to a stranger who provided help	0.06	0.19**	n.a.	
Altruism	generalized	Hypothetical choice of the amount donated to a charity (out of Ksh3200)	0.41***	0.41***	0.39***	comparable
	in-group	Hypothetical choice of the amount donated to a charity that helps people in ancestral home area (out of Ksh3200)	0.33***	0.36***	n.a.	
	out-group	Hypothetical choice of the amount donated to a charity that helps people in other parts of Kenya than ancestral home area (out of Ksh3200)	0.40***	0.38***	n.a.	

Anti-social behavior	generalized	Hypothetical decision between (3200, 3200) or (3150, 1600) for self and for another person	0.05	0.05	n.a.
	in-group	Hypothetical decision between (3200, 3200) or (3150, 1600) for self and for a person from ancestral home area	0.07	0.07	n.a.
	out-group	Hypothetical decision between (3200, 3200) or (3150, 1600) for self and for a person from other parts of Kenya than ancestral home area	0.14	0.14	n.a.

Notes: Column 1 is an OLS coefficient from a regression of the standardized experimental measure on the standardized survey item. Column 2 displays Spearman correlations between the survey item and the respective experimental measure (one for each preference type, except for reciprocity, where we use two experimental measures). ***, **, and * denote significance at the 1-, 5-, and 10-percent level, respectively. Column 3 displays the correlation between experimental and quantitative survey measures from the validation study of Becker et al. (2016) among university students in Germany. Column 4 indicates to what extent measures from our study in Kenya and measures from the German study are comparable.

Table 2: Correlations between qualitative survey measures and experimental measures

Preference		Qualitative survey item	Kenya: Kibera residents		Germany: Bonn students	
			OLS Coefficient	Correlation	Correlation	Measures
			(1)	(2)	(3)	(4)
Time		Willingness to give up something that is beneficial today in order to benefit more in the future	0.04	0.06	-0.41***	comparable
Risk		Willingness to take risks	0.01	-0.02	0.35***	comparable
Reciprocity		Willingness to return a favor	0.11	0.06	0.30***	experimental
Reciprocity (diff)		Willingness to return a favor	0.21**	0.14		different; survey comparable
Altruism	generalized, measure 1	Willingness to give to a charity	0.03	0.07	0.38***	comparable
	generalized, measure 2	Willingness to share with others	-0.06	-0.02	0.23***	comparable
	in-group, measure 1	Willingness to give to a charity that helps people in ancestral home area	-0.03	-0.09	n.a.	
	in-group, measure 2	Willingness to share with others from ancestral home area	-0.05	-0.05	n.a.	
	out-group, measure 1	Willingness to give to a charity that helps people in other parts of Kenya than ancestral home area	0.18**	0.12	n.a.	
	out-group, measure 2	Willingness to share with people from other parts of Kenya than ancestral home area	0.12	0.13	n.a.	
Anti-social behavior	generalized, measure 1	Willingness to cause troubles to other people	-0.1	-0.05	n.a.	
	generalized, measure 2	Willingness to make harm to other people	0.01	0.05	n.a.	
	in-group, measure 1	Willingness to cause troubles to people in ancestral home area	-0.02	-0.003	n.a.	
	in-group, measure 2	Willingness to make harm to people in ancestral home area	0.11	0.15	n.a.	

14

out-group, measure 1	Willingness to cause troubles to people from other parts of Kenya than ancestral home area	-0.01	0.02	n.a.
out-group, measure 2	Willingness to make harm to people from other parts of Kenya than ancestral home area	0.01	0.03	n.a.

Notes: Column 1 displays OLS coefficients in a regression of the standardized experimental measure on the standardized module items. Column 2 displays Spearman correlations between the survey item and the respective experimental measure (one for each preference type, except for reciprocity, where we use two experimental measures). ***, **, and * denote significance at the 1-, 5-, and 10-percent level, respectively. Column 3 displays the correlation between experimental and qualitative survey measure from the validation study of Becker et al. (2016) among university students in Germany. Column 4 indicates to what extent measures from our study in Kenya and measures from the German study are comparable.

ONLINE APPENDIX

APPENDIX A

Additional Tables

Table A1: Summary statistics

	Mean	Std. Dev.	Min	Max	N
	(1)	(2)	(3)	(4)	(5)
Panel A: Experimental measures					
Time preference	9.15	9.11	1	26	123
Risk preference	6.82	6.05	1	22	123
Ambiguity aversion	0.24	0.43	0	1	123
Reciprocity	257.33	259.37	0	1000	123
Reciprocity (diff)	179.8	260.73	-400	1000	123
Alturism - generalized	205.45	213.7	0	1000	123
Altruism - in-group	213.13	221.65	0	1000	123
Altruism - out-group	165.71	183.31	0	1000	123
Anti-social behavior - generalized	0.2	0.4	0	1	123
Anti-social behavior - in-group	0.23	0.43	0	1	115
Anti-social behavior - out-group	0.38	0.49	0	1	123
Panel B: Survey quantitative measures					
Time preference	7.58	11.92	1	32	123
Risk preference	11.93	11.11	1	32	123
Ambiguity aversion	0.24	0.43	0	1	123
Reciprocity	269.15	403.87	0	2000	123
Alturism - generalized	645.45	689	0	3200	123
Altruism - in-group	708.82	712.68	0	3200	123
Altruism - out-group	663.65	724.59	0	3200	123
Anti-social behavior - generalized	0.17	0.38	0	1	123
Anti-social behavior - in-group	0.19	0.39	0	1	123
Anti-social behavior - out-group	0.24	0.43	0	1	123
Panel C: Survey qualitative measures					
Time preference	7.54	3.03	0	10	123
Risk preference	6.93	2.94	0	10	123
Reciprocity	8.92	2.22	0	10	123
Alturism - generalized, measure 1	6.89	3.34	0	10	123
Altruism - in-group, measure 1	7.63	3.02	0	10	123
Altruism - out-group, measure 1	6.81	3.21	0	10	123
Altruism - generalized, measure 2	7.72	3.04	0	10	123
Altruism - in-group, measure 2	7.32	3.29	0	10	123
Altruism - out-group, measure 2	7.02	3.37	0	10	123
Anti-social behavior - generalized, measure 1	2.37	3.56	0	10	123
Anti-social behavior - in-group, measure 1	2.37	3.69	0	10	123

Anti-social behavior - out-group, measure 1	2.17	3.39	0	10	123
Anti-social behavior - generalized, measure 2	1.98	3.47	0	10	123
Anti-social behavior - in-group, measure 2	2.24	3.7	0	10	123
Anti-social behavior - out-group, measure 2	1.97	3.42	0	10	123
Panel D - Observable characteristics					
Age	29.79	4.82	20	46	119
Female	0.54	0.5	0	1	123
Unemployed	0.54	0.5	0	1	123
Number of children	1.92	1.62	0	9	123

Table A2: Robustness checks – correlation between experimental and quantitative survey measures

Preference type (dependent variable - experimental measure, explanatory variable - quantitative survey measure)	Controlling for:					Sub-sample	
	No controls	Qualitative measure(s)	Understanding; and consistency (where applicable)	Age, gender, being unemployed and number of children	Order of tasks and order of survey/experiments; and understanding and consistency (where applicable)	Survey questions in the first week	Experimental choices in the first week
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Time preference	0.33***	0.33***	0.31***	0.35***	0.33***	0.31***	0.39***
	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.10)	(0.14)
Risk preference	0.21**	0.22**	0.25***	0.19**	0.25***	0.32**	0.12
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.13)	(0.13)
Ambiguity aversion	0.13		0.14	0.09	0.12	0.18	0.08
	(0.09)		(0.09)	(0.09)	(0.09)	(0.13)	(0.13)
Reciprocity	0.12	0.10		0.15	0.11	0.20	0.07
	(0.09)	(0.09)		(0.09)	(0.09)	(0.13)	(0.13)
Reciprocity (diff)	0.06	0.04		0.07	0.06	0.13	0.03
	(0.09)	(0.09)		(0.09)	(0.09)	(0.13)	(0.13)
Generalized altruism	0.41***	0.43***		0.44***	0.44***	0.33***	0.67***
	(0.08)	(0.08)		(0.09)	(0.09)	(0.10)	(0.16)
In-group altruism	0.33***	0.36***		0.32***	0.36***	0.26**	0.54***
	(0.09)	(0.09)		(0.09)	(0.09)	(0.10)	(0.17)
Out-group altruism	0.40***	0.38***		0.35***	0.39***	0.27**	0.54***
	(0.08)	(0.09)		(0.09)	(0.09)	(0.12)	(0.12)
Generalized antisocial	0.05	0.06	0.04	0.05	0.05	-0.02	0.09
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.12)	(0.14)
In-group antisocial	0.07	0.07		0.06	0.09	0.08	0.08
	(0.10)	(0.10)		(0.10)	(0.10)	(0.12)	(0.16)
Out-group antisocial	0.14	0.14		0.14	0.13	0.24*	0.03
	(0.09)	(0.09)		(0.09)	(0.09)	(0.13)	(0.13)

18

Notes: OLS, standard errors in parentheses. Each cell provides a coefficient from a separate regression, in which the dependent variable is the experimental measure of a given preference type, and the explanatory variable is a quantitative survey measures of the same preference type. All measures of preferences are standardized. *** p<0.01, ** p<0.05, * p<0.1

Table A3: The role of memory

Dependent variable	Time preference (1)	Risk preference (2)	Reiprocity (3)	Reciprocity (diff) (4)	Ambiguity aversion (5)	Generalized altruism (6)	Ingroup altruism (7)	Outgroup altruism (8)
Quantitative measure	0.15	0.18	0.18	0.13	0.25*	0.44***	0.38***	0.51***
	(0.12)	(0.14)	(0.12)	(0.12)	(0.15)	(0.10)	(0.11)	(0.12)
Good memory*Quantitative measure	0.40**	0.09	-0.20	-0.21	-0.16	-0.00	-0.03	-0.29
	(0.17)	(0.19)	(0.18)	(0.18)	(0.19)	(0.20)	(0.19)	(0.18)
Qualitative measure	0.27**	-0.04	0.14	0.25**		0.12	-0.10	-0.07
	(0.13)	(0.14)	(0.11)	(0.11)		(0.14)	(0.17)	(0.18)
Good memory*Qualitative measure	-0.36**	0.17	0.04	-0.10		-0.23	-0.08	0.10
	(0.17)	(0.19)	(0.19)	(0.20)		(0.18)	(0.22)	(0.22)
Qualitative measure 2						-0.10	0.10	0.20
						(0.13)	(0.16)	(0.16)
Good memory*Qualitative measure2						-0.05	-0.20	-0.30
						(0.18)	(0.21)	(0.21)
Good memory (above median)	0.15	-0.20	-0.59***	-0.24	-0.26	-0.03	-0.40**	-0.33*
	(0.17)	(0.18)	(0.18)	(0.18)	(0.18)	(0.17)	(0.17)	(0.17)
Constant	-0.09	0.11	0.31**	0.13	0.15	0.01	0.20	0.19
	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)	(0.12)	(0.12)	(0.13)
Observations	123	123	123	123	123	123	123	123
R-squared	0.190	0.064	0.117	0.073	0.039	0.201	0.178	0.224

Notes: OLS, standard errors in parentheses. All measures of preferences are standardized. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

APPENDIX B

Design of the validation experiment

Below, we describe in detail how we elicited each of the experimental and survey measures. The full experimental and survey protocols in English and Swahili are available in Appendix C and D. Where relevant, for convenience we also carefully compare the similarities and differences between our approach and the original validation experiment implemented by Becker, Dohmen, Huffman, Falk, & Sunde (2016) - BDHFS.

Sample

The sample in our study are 123 participants from the Kibera neighborhood in Nairobi, Kenya during August 2018. The participants come from a low income environment, are between 20 and 46 years of age, more than a half are unemployed, half of them are women and on average they have two children. The average monthly earnings among those who reported this measure (N=57) is approximately USD 96.

Experimental measures

We conducted a set of incentivized choice experiments in which each type of preferences is elicited in one experimental task. Specifically, we implemented ten experiments in total, focusing on the following types of preferences: time preference, risk preference, reciprocity, ambiguity aversion, generalized altruism, in-group altruism, out-group altruism, generalized anti-social behavior, in-group antisocial behavior, and out-group antisocial behavior.

BDHFS also elicited measures of time preference, risk preference, altruism, and reciprocity. We will focus on comparison of the experiments in these four domains where there is an overlap between the two studies. In addition, BDHFS elicited measures of trust and negative reciprocity, and they implemented nine experiments in total.

Time preference

We conducted an experiment that involved 25 binary choices between a payment "today" and a higher payment that would be received in 3 months in the future. The delayed payment in each subsequent binary choice increased such that the implied 3-months return from waiting would rise in steps of 5 percentage points from 0 percent in the first binary choice to 120 percent in the 25th binary choice. The payment today was in all 25 binary choices KSh 600, while the payment in 3 months increased from KSh 600 in the first binary choice to KSh 1320 in the last binary choice. (The exchange rate during the study period was approximately 100 KSh to 1 USD.) If a choice in this experiment was selected to be payoff relevant, the money was sent to participant's mobile phone via M-PESA, either on the day of the experiment, or 3 months later. The row in which a participant switched from preferring the earlier payment to the larger delayed payment provides a measure of time preference.

BDHFS conducted two experiments to elicit measures of time preference with 25 binary choices each. In both price lists, participants had to trade-off a payment of 400 points "today" and a higher payment that would be received 12 months in the future. In one price list, the delayed amount was increased such that

the implied annual return from waiting would rise in steps of 2.5 percentage points from 0 percent in the first row to 60 percent in the 25th row. In the second price list the delayed payments were perturbed by adding or subtracting an amount of up to 0.6 points. The payments were sent by regular mail. In both experiments, the row in which a participant switched from preferring the earlier payment to the larger delayed payment provides a measure of impatience. The ultimate measure of time preference was constructed by averaging the switching rows in the two discounting experiments.

Our experiment closely follows the first experiment implemented by BDHFS. It differs in terms of (i) the amount of the payment "today", (ii) the delay (3 months instead of 12 months) and (iii) the annual return from waiting. We changed the delay and the interest rate based on a pilot study which suggested that there would be limited variation in the measure of time preference if we used exactly the same parameters as in BDHFS.

Risk preference

We conducted an experiment that involved 21 binary choices between a safe payment and a lottery that yielded with equal probability KSh 0 or KSh 2,000. The lottery was the same in all binary choices, while the safe payment was increased in steps of KSh 100 from KSh 0 in the first binary choice to KSh 2,000 in the 21st binary choice. Also, here we follow closely the approach of BDHFS who implemented this experiment, the only difference being the specific amounts. In their case, the lottery was between 0 and 1,000 points, while the safe payment was increased in steps of 50 points from 0 to 1,000. In addition, BDHFS implemented a second price list in which they perturbed the safe payments by adding or subtracting up to five points to each safe payment alternative. The row in which a participant switched from preferring the lottery to preferring the safe payment is a measure of risk preference. The measure in BDHFS is constructed by averaging the switching rows in the two experiments.

Altruism

Each participant was endowed with KSh 1,000 and had to decide how much of that amount to donate to a charity that helps people in Kenya. The task is very similar to BDHFS, with small differences in the wording. In BDHFS, participants were endowed with 300 points and made a decision how many points to assign to a charitable organization (by choosing a specific organization from a provided list or by naming a different one). The amount donated to a charity is a measure of altruism.

Reciprocity

Each participant was endowed with KSh 1,000 and had to decide how much of that amount to give to two other people who visited the lab in the past. The participants were informed that each of these people received two bags of sugar and could decide to leave one of the bags for a future visitor of the lab. One of them decided to give a bag of sugar while the other one not. We use two measures of reciprocity – the amount assigned to the "kind" person who left a bag of sugar for the participants, and the difference in the amounts assigned to the "kind" person and to the other person who did not leave a bag of sugar for the participant. Our measure of reciprocity differs from that of BDHFS who elicit the measure of positive reciprocity from second mover behavior in the Trust game.

We elicited further measures of preferences which were not included in the BDHFS study.

Ambiguity aversion

We conducted an experiment in which participants made a choice whether to draw a ball from one or from another jar. In jar 1, there were ten balls, out of which four were green and six were yellow. In jar 2, there were also ten balls, but the number of green and yellow balls was unknown. If jar 1 was chosen, the participant needed to draw a green ball to win KSh 1,000. If jar 2 was chosen, the participant needed to choose a color and draw a ball of that color to win KSh 1,000. The choice of jar 1 is our measure of ambiguity aversion.

Anti-social behavior

In the task related to anti-social behavior, each participant was matched with an anonymous person from Kenya and both of them received an endowment of KSh 1,000. The participant made a choice between two options. The first one was to keep KSh 1,000 for self and KSh 1,000 for the other person. The second one was to lower the amount of the other person by KSh 500, but this cost the participant KSh 20, and thus the participant would receive KSh 980 and the other person KSh 500. The choice of the second option is our measure of anti-social behavior.

In-group and out-group measures of altruism and anti-social behavior

In total, we elicited three measures of altruism and three measures of anti-social behavior. Besides the generalized measures described above, we elicited a measure of behavior towards members of participants' in-group and towards out-group members. In the in-group version of the experiment on altruism, the participants were informed that they could donate a part of their endowment to a charity that helped people in their ancestral home area. Similarly, in the in-group version of the experiment on anti-social behavior, they were informed they were matched with an unknown person from their ancestral home area. In the out-group version of the experiment on altruism, they were informed that the charity helped people from other parts of Kenya than their ancestral home area. In the out-group version of the experiment on anti-social behavior they were matched with an unknown person from Kenya, but not from their ancestral home area. The formulation "from your home area" was carefully selected from a list of possible information indicating one's own ethnic group based on a detailed conversation with the Innovations for Poverty Action (IPA) field team in Kenya.

Survey measures

To measure each type of preference, we use one quantitative survey measure and one or two qualitative survey measures (one for time preference, risk preference and reciprocity; two for altruism and anti-social behavior). The only exception is ambiguity aversion for which there is only one quantitative survey measure. The quantitative questions present a hypothetical scenario that mimics closely the experimental task. The qualitative questions measure willingness to act in a certain way on 0-10 scale. In total, the participants answered 25 survey questions.

In their validation experiment, BDHFS used a larger number of survey questions to measure each type of preference. In total, they included 199 questions. Then, the researchers identified the best linear combination of items for measuring a particular preference type and these items were selected to be included in GPS. For some measures, the researchers, when developing this streamlined version of the survey module, used survey items which have slightly lower predictive power but are simpler and faster to implement. In some cases, they also adjusted the wording such that it can be used in different cultures. For example, instead of asking the whole set of binary choices on time/risk preference which is time-consuming, only a sub-set of five interdependent questions was included in GPS. In our study, the survey questions on

time preference, risk preference, reciprocity and altruism follow very closely the GPS questions. We use exactly the same or slightly adjusted wording of the questions included in the Swahili version of GPS for Kenya. In addition, we designed new questions on other types of preferences we are interested in (ambiguity aversion, anti-social behavior) and include the in-group and out-group versions of questions on altruism and anti-social behavior.

Below, we provide description of the questions used in our study, as well as – where relevant – a comparison to GPS-Kenya questions, and questions used in the validation study of BDHFS for which correlations with experimental measures are displayed in Tables 1 and 2.

Time preference

Quantitative measure. Since the experiment involved 25 binary choices and making decision in all of these is rather time-consuming, for the quantitative survey measure we use the "staircase" or "unfolding brackets" procedure where each participant answered a sub-set of five binary choices. First, they made a choice between KSh 300 today or KSh 461 in 12 months. In the second and all subsequent binary choices, the immediate payment remained the same, but the delayed payment changed based on the previous decision. If the participant had chosen the immediate payment, the delayed payment in the subsequent binary choice was increased. If the participant had chosen the delayed payment, the delayed payment in the subsequent binary choice was decreased. In total, there were 31 binary choices out of which each participants faced five. The measure of time preference takes the values between 1 (preference of the immediate payment in a situation with the highest delayed payment, specifically KSh 644) and 32 (preference of delayed payment in a situation with the lowest delayed payment, specifically KSh 309).

Qualitative measure. In the qualitative survey item, the participants were asked "How willing are you to give up something that is beneficial for you today in order to benefit from that in the future?" and indicated their answer on a scale from 0 to 10 where 0 means "completely unwilling to do so", and 10 means "very willing to do so". Both the quantitative and qualitative survey measure are identical with the GPS-Kenya questions.

Risk preference

Quantitative measure. Similarly to time preference, to elicit quantitative survey measure of risk preference, we use the "staircase" method. First, the participants made a choice between a draw with a 50-percent chance of receiving KSh 900 and the same 50-percent chance of receiving nothing, or the amount of KSh 480 as a sure payment. In the second and all subsequent decisions, the lottery remained the same. If the participant had chosen the safe option, the safe option in the subsequent question was smaller. If the participant had chosen the lottery, the safe option in the subsequent question was larger. In total, there were 31 binary choices out of which each participants faced five. The measure of risk preference takes the values between 1 (preference of sure payment in a situation with the lowest sure payment, specifically KSh 30) and 32 (preference of lottery in a situation with the highest sure payment, specifically KSh 930). The quantitative survey measure of risk preference is identical with the GPS-Kenya questions.

Qualitative measure. In the qualitative survey item, the participants were asked "Please tell me, in general, how willing or unwilling you are to take risks. Let me explain what I mean by risk. Imagine you are going to start a business. You are going to take risk because you do not know if the business will succeed or if it will fail." and indicated their answer on a scale from 0 to 10 where 0 means "completely unwilling to take risks", and 10 means "very willing to take risks". The qualitative survey measure of risk preference is similar to the GPS-Kenya question, but based on a pilot session which revealed that the participants had

hard times to understand the term "risks", we expanded the wording adding an explanation what we mean by risk. The original GPS question asks "Please tell me, in general, how willing or unwilling you are to take risks." The question used in the validation experiment by BDHFS is "Generally speaking, are you a person who is willing to take risks or do you try to avoid risks?".

Altruism

Quantitative measure. The participants were asked "Imagine the following situation: Today you unexpectedly received KSh 3,200. How much of this amount would you donate to a charity?". This question closely follows the GPS-Kenya question, including the specific amount. The difference is that the GPS asks about donating "to a good cause", while we ask about donating "to a charity".

Qualitative measures. We use two qualitative survey measures of altruism. One asks respondents "How willing are you to give to a charity without expecting anything in return?". The participants rate their willingness on 0-10 scale. Again, with the exception of the formulation "to a charity" instead of "to good causes", the question is the same as in GPS. Nevertheless, the validation in the study by BDHFS was based on a survey question with included the formulation "when it comes to charity". Because the term "charity" caused confusion in some countries, for the purposes of the GPS module the formulation was changed to "good cause". The second question is "Are you a person who is generally willing to share with others without expecting anything in return, or are you not willing to do so?" This was used by BDHFS in their validation experiment but was not selected to be included in GPS.

Reciprocity

Quantitative measure. The participants were described the following scenario: "Please think about what you would do in the following situation. You are in a city you are not familiar with, and you realize you lost your way. You ask a stranger for directions. The stranger offers to walk with you and show you the way to your destination. By helping you the stranger misses an hour of work and thus loses 50 shillings in total. However, the stranger says he or she does not want any money from you. When you arrive to your destination, you can buy a gift for the stranger in a shop." Then they were asked whether they buy a "thank-you" gift for the stranger and how much money they would spend on the present. The amount spent on the present is our quantitative measure of reciprocity.

The quantitative question on reciprocity in GPS-Kenya is similar in spirit, but again we made some adjustments in the wording based on the pilot and discussions with the local team. The original GPS question asks "Please think about what you would do in the following situation. You are in an area you are not familiar with, and you realize that you lost your way. You ask a stranger for directions. The stranger offers to take you to your destination. Helping you costs the stranger about KSh 60 in total. However, the stranger says he or she does not want any money from you. You have six presents with you. The cheapest present costs KSh 15, the most expensive one costs KSh 90. Do you give one of the presents to the stranger as a "thank-you" gift?" The participants could choose between giving no presents, or a present which costs KSh. 15/30/45/60/75/90.

The question in the validation of BDHFS is: "Imagine the following situation: you are shopping in an unfamiliar city and realize you lost your way. You ask a stranger for directions. The stranger offers to take you with their car to your destination. The ride takes about 20 minutes and costs the stranger about 20 Euro in total. The stranger does not want money for it. You carry six bottles of wine with you. The cheapest bottle costs 5 Euro, the most expensive one 30 Euro. You decide to give one of the bottles to the stranger as a thank-you gift. Which bottle do you give?".

Qualitative measure. The participants were asked to say how well the following statement describes them: "When someone does me a favor I am willing to return it." They provided answer on a scale from 0 to 10, where 0 means "does not describe me at all", and 10 means "describes me perfectly". The qualitative survey question on reciprocity is identical with the GPS question.

As in the experiments, we elicited further measures of preferences which are not a part of GPS.

Ambiguity aversion

Quantitative measure. In the survey, we included a single (quantitative) measure of ambiguity aversion. The participants were asked to imagine that they were going to play a game and they were described the experimental task we implemented to elicit measure of ambiguity aversion.

Anti-social behavior

Quantitative measure. The participants were asked to imagine a situation mimicking the experimental task in which they and another unknown person unexpectedly received an opportunity to get KSh 3,200 each. Then they were asked to make a choice between two options, one in which both the participant and another person receive KSh 3,200, and one in which the participant receives KSh 3,150 and the other person receives KSh 1,600. The choice of the second option is our quantitative survey measure of anti-social behavior.

Qualitative measures. We designed two qualitative survey items to measure anti-social behavior. We asked participants "How willing or unwilling are you to cause troubles to other people?", and "How willing or unwilling are you to make harm to other people?". In both cases, the participants rated their willingness on a 0-10 scale.

In-group and out-group measures of altruism and anti-social behavior

We included the in-group and out-group variants of all three survey questions on altruism and of all three survey questions on anti-social behavior. As in the experiments, the distinction was made by using the formulation "from your ancestral home area" vs. "from other parts of Kenya other than your ancestral home area"

Payments and procedures

On average, the earnings of the participants from both the experimental and survey sessions were KSh 1,520, i.e. an equivalent of approximately five days earnings. Specifically, for the experimental part, the participant received a show up fee of KSh 250 (KSh 200 for participation and KSh 50 if they arrived on time). After they finished all the experimental tasks, one of the decisions they made was randomly selected to be payoff relevant. On average, the payoff was KSh 820. For the survey part, the participants received a show up fee of KSh 450 (KSh 400 plus KSh 50 if they arrived on time).

The participants visited the lab twice. During one session they made choices in all the experimental tasks, and during the other session they answered the survey questions. The two visits were one week apart and we randomized the order of the experiments vs. survey. This approach aims to limit the spillovers between the experimental and survey measures, for example due to an effort to give consistent answers. To address this concern further, we included a task to measure participants' memory. In the first week, the participants were shown a set of ten letters from the alphabet on a screen for twenty seconds. In the second week, they

were asked to recall these letters. In the analysis, we can test whether the correlation between experimental and survey measures is driven by participants with better memory. In the experimental session, we further randomized at the session level the order of a set of experiments focusing on (i) time preference, risk preference, and ambiguity aversion, and (ii) altruism, reciprocity, and anti-social behavior.