Is There a Development Gap in Rationality?*

Alexander W. Cappelen, Shachar Kariv,

Erik Ø. Sørensen, and Bertil Tungodden†

January 23, 2015

Abstract

We compare the rationality of choice under risk – utility maximization, stochastic dominance, and expected-utility maximization – of students from one of the best universities in the US and one of the best universities in Africa. The US subjects came nearer to consistency with utility maximization and the dominance principle, but there are no differences between the two samples in consistency with expected-utility maximization. A canonical cognitive ability (IQ) test indicates a much larger development gap relative to our tests of economic rationality. The results are robust to the inclusion of controls for non-cognitive abilities and personality traits.

JEL Classification Numbers: D01, D03, D81, F61, F63, O12.

Keywords: Rationality, revealed preference, stochastic dominance, expected utility, cognitive skills, non-cognitive skills, personality traits, Big Five, development, experiment.

*We thank Siv-Elisabeth Skjelbred for excellent research assistance. We also thank Raymond Fisman, Pamela Jakiela and Edward Miguel for many thoughtful comments. The paper has also benefited from suggestions by the participants of seminars at several universities and the 2012 Symposium on Economic Experiments in Developing Countries (SEEDEC). We acknowledge financial support from the Peder Sather Center for Advanced Study (An International Research and Educational Collaboration between the University of California, Berkeley, and Norway), the Research Council of Norway (Grant No. 202484) and the National Science Foundation (Grant No. SES-0962543). The experiments were administered by the Choice Lab at the Norwegian School of Economics (NHH) and the Experimental Social Science Laboratory (Xlab) at the University of California, Berkeley.

†Cappelen: Norwegian School of Economics (email: alexander.cappelen@nhh.no); Kariv: University of California, Berkeley (email: kariv@berkeley.edu); Sørensen: Norwegian School of Economics (email: erik.sorensen@nhh.no); Tungodden: Norwegian School of Economics (email: bertil.tungodden@nhh.no).

This paper is dedicated to Cecilie Rasmussen, the first research coordinator of the Choice Lab at Norwegian School of Economics (NHH), who passed away far too young on a research trip to Berkeley in the spring of 2012.

1 Introduction

In this paper we provide experimental evidence on the question "is there a development gap in rationality?" We draw the subjects for the experiment from the student body at the University of California, Berkeley, ranked as the world's top public university and one of the most prestigious (public or private) ones, and the University of Dar es Salaam, the oldest and biggest public university in Tanzania, and one of the best ranked universities in Africa.[1] These subject pools carry an *intrinsic* interest. Although the students at UC Berkeley and the University of Dar es Salaam come from different backgrounds and face different economic prospects, they are united by being among the most able in their societies. We thus procure experimental subjects at the high end of the "ability spectrum" when assessed for economic rationality.

The subject pools are also worth studying for *extrinsic* reasons. The students of UC Berkeley and the University of Dar es Salaam will presumably assume positions of power in various sectors of their economy when they graduate into the world.[2] This is especially true for the Tanzania students in our sample, many of whom will assume positions of substantial power in national economic and political affairs.[3] As Acemoglu and Robinson (2012) argue "...poor countries are poor because those who have power make choices that create poverty. They get it wrong not by mistake or ignorance but on purpose (p 68)." The argument stipulates that those who have power must be rational. This highlights the importance of rigorously testing the rationality of the future leadership in developing countries.

Because uncertainty is endemic in a wide variety of circumstances, we provide a thorough experimental test of the touchstones of rationality in decision making under risk – utility maximization, stochastic dominance, and expected-utility maximization. Inconsistencies with revealed preference conditions and violations of first-order

[1]The Webometrics Ranking of World Universities (www.webometrics.info) ranks all universities worldwide based on the impact of the university on the web. When we conducted the experiments in 2012-13, it ranked UC Berkeley as the fourth best university worldwide. The University of Dar es Salaam is ranked 1,419, but it is ranked first in Tanzania and 11^{th} in Africa.

[2]There is credible evidence that quality education has a strong relationship with economic growth. Education affects economic growth through different mechanisms. The neoclassical argument is that education increases the human capital of the labor force, which in turn increases labor productivity (cf., Mankiw et al., 1992). Science, technology, engineering and mathematics education promotes the growth of the innovation economy (cf., Lucas, 1988). For further discussion, see Hanushek and Woessmann (2007).

[3]The President of Tanzania, Jakaya Kikwete, the Prime Minster, Mizengo Pinda, and the Chief Justice, Mohamed Chande Othman, all graduated from the University of Dar es Salaam.

stochastic dominance (FOSD) are regarded irrational, regardless of risk attitudes, because they leave “money on the table.” Expected Utility Theory (EUT) serves as a normative guide for choice under risk (how people ought to choose). There is no need to assume EUT to investigate rational behavior under uncertainty (in the sense of a complete, transitive preference ordering), but choices can be rational and yet fail to be reconciled with any utility function that is normatively appealing given the decision problem at hand.

The experiment. We presented subjects with a sequence of standard consumer decision problems: selection of a bundle of commodities from a standard budget set. These decision problems were presented using the graphical interface pioneered by Choi et al. (2007a). The approach has a number of advantages over earlier approaches. First, the choice of a bundle subject to a budget constraint provides more information about preferences than a typical discrete choice. Second, because the interface is extremely user-friendly, it is possible to present each subject with *many* choices in the course of a single experimental session, yielding a much larger data set.

This makes it possible to analyze behavior at the level of the individual subject, without the need to pool data or assume that subjects are homogenous, and allows for the thorough testing of observed behavior. The power of the tests depends on two factors. The first is that the range of choice sets is generated so that budget lines cross frequently. The second is that the number of decisions made by each subject is large. This is a crucial point, because in most experimental studies, the number of individual decisions is too small to provide a powerful test.

In the experiment, there are two equally probable states of nature, $s = 1,2$ and an Arrow security for each state. An Arrow security for state s promises a token (the experimental currency) in state s and nothing in the other state. Each decision problem is presented as a choice from a two-dimensional budget line. A choice of the allocation $\mathbf{x} = (x_1, x_2)$ from the budget line denotes an allocation of securities, where x_s denotes the number of units of security s. The budget line is $B(\mathbf{p})$ where $\mathbf{p} = (p_1, p_2)$ is the vector of security prices and p_s denotes the price of security s. The subject can choose any allocation $\mathbf{x}$ that satisfies this constraint.

Measures of rationality. The most basic question to ask about choice data is whether the data are consistent with individual utility maximization – that is, is there a utility function $U(\mathbf{x})$ such that for any price vector $\mathbf{p}$, the chosen allocation $\mathbf{x}$ maximizes $U(\mathbf{x})$ subject to $\mathbf{x} \in B(\mathbf{p})$? Classical revealed preference theory (Afriat, 1967; Varian, 1982, 1983) provides a direct test: choices in a finite collection of budget lines are consistent with maximizing a utility function if and only if they satisfy the Generalized Axiom of Revealed Preference (GARP). Because our subjects make choices in a wide range of budget lines, our data provide a stringent test of GARP.

Violations of monotonicity with respect to FOSD are regarded as errors – that is, a failure to recognize that some allocations yield payoff distributions with unambiguously lower returns (Hadar and Russel, 1969). Since the two states are equally probable, any decision to allocate *fewer* tokens $x_s < x_{s'}$ to the *cheaper* security $p_s < p_{s'}$ violates FOSD but need not involve a violation of GARP. The dominance principle is compelling and generally accepted in decision theory.[4]

If preferences satisfy the Savage (1954) axioms, then the preference ordering over allocations $\mathbf{x}$ can be represented by a function of the form $U(\mathbf{x}) = u(x_1) + u(x_2)$ where $u(\cdot)$ is the utility function over money (the two states are equally probable). Such an individual will choose an allocation $\mathbf{x}^*$ to maximize the expected value of utility subject to the budget constraint $\mathbf{x} \in B(\mathbf{p})$. EUT lies at the very heart of economics. It is meant to serve as a normative guide for choice and also as a descriptive model of how individuals choose. At the same time, much of the experimental and empirical evidence of "anomalies" in choice behavior suggests that EUT may not be the right model of choice under uncertainty. To test EUT, we estimate at the individual level preferences generated by the disappointment aversion model of Gul (1991), which embeds EUT as a parsimonious and tractable special case.

Experimental results. Since GARP offers an exact test (either the data satisfy GARP or they do not), we use Afriat's (1972) Critical Cost Efficiency Index (CCEI) to quantify the extent of violation. The CCEI measures the amount by which each budget constraint must be relaxed in order to remove all violations of GARP. This measure is bounded between zero and one. The closer it is to one, the smaller the perturbation of budget lines required to remove all violations and thus the closer the data are to satisfying GARP. Figure 1 below summarizes the cumulative distributions of CCEI scores in the two samples. Alternative measures of GARP violations yield similar results.

[Figure 1 here]

The distributions reported in Figure 1 show that the US subjects display greater levels of consistency than the Tanzanian subjects: mean CCEI scores across all subjects are 0.950 in the US and 0.856 and 0.869 with low and high stakes in Tanzania, respectively.[5] The magnitudes imply that the Tanzanian subjects on average waste as much as 9.4 (low stakes) and 8.1 (high stakes) percentage points more of their earnings by making inconsistent choices relative to the US subjects. Nonetheless, many of our subjects in both pools are close enough to passing GARP so that we may not want to reject that their choices are consistent with utility maximization.[6] This is, to our knowledge, the first quantifiable and economically interpretable measure of the development gap in rationality.

[4] All preference relations usually considered satisfy FOSD. In fact, as noted by Quiggin (1990) and Wakker (1993), theories of choice under uncertainty were amended to avoid violations of dominance.

[5] Budget lines intersect at least one axis at or above the 50 token level and intersect both axes at or below the 100 token level. Each experimental token was worth 0.5 USD in the US and 100 TZS (equivalent to 0.06 USD) in the low-stakes treatment in Tanzania. In the high-stakes treatment in Tanzania, each experimental token was worth 1000 TZS.

[6] Varian (1990, 1991) suggested a threshold of 0.95 for the CCEI, but this is purely subjective. A more scientific approach, proposed by Bronars (1987), calibrates the CCEI using a hypothetical subject whose choices are uniformly distributed on the budget line (more below).

The rest of our results are summarized as follows. We use expected payoff calculations to assess how closely individual behaviors comply with FOSD. The choices made by subjects in our experiment show very low rates of FOSD violations, but the US subjects violate FOSD less. The difference is very small but statistically significant. We next test if individual behaviors comply with EUT. The parameter estimates vary dramatically across subjects in both samples, implying that individual behavior is very heterogeneous. Nevertheless, most of our subjects are well-approximated by preferences consistent with EUT. Using a constant relative risk aversion (CRRA) specification, assuming the power form, for 65.1 percent of the US subjects and 70.1 percent of the Tanzanian subjects, we cannot reject the null hypothesis of preferences consistent with EUT using a five percent significance level. The corresponding numbers according to a constant absolute risk aversion (CARA) specification, assuming the exponential utility function, increase to 81.7 and 89.3, respectively. Overall, the results on FOSD violations and consistency with EUT indicate an even smaller or no development gap in economic rationality.

Economic rationality, cognitive ability and non-cognitive abilities. An alternative approach in behavioral economics is to 'proxy' economic rationality with a test of cognitive ability (IQ), which indeed seems to capture aspects of the ability to make rational economic decisions (Dohmen et al., 2010). Consistency with GARP, and the other rationality postulates above, offers a theoretically disciplined metric for the rationality of economic decisions, and there is no comparable, theoretically disciplined means for using, interpreting, and evaluating an IQ test. Nevertheless, it is useful to investigate the relationship between IQ and the measures of economic rationality above. If these measures and IQ are very well correlated, then analysts interested in measuring economic rationality might replace our tests with one of the many IQ tests and the conceptual distinctions between the measures would have little practical import.

To compare economic rationality and cognitive ability, our subjects also completed a standard (non-incentivized) Wechsler Adult Intelligence Scale test (WAIS-IV). The IQ test generated substantial variation in both samples and the scores are essentially uncorrelated with the CCEI from the experiment (0.063 and 0.110 in the US and Tanzania, respectively). Figure 2 presents the cumulative distributions of IQ scores (the fraction of questions answered correctly) in both samples. The difference between the distributions of IQ scores depicted in Figure 2 is much larger than the difference between the distributions of CCEI scores depicted in Figure 1 above. In fact, the distributions of IQ scores hardly overlap – the 90th percentile score in Tanzania *equals* the 10th percentile score in the US. This provides a clear illustration of the extent to which using IQ as a proxy for economic rationality will inflate the development gap.

[Figure 2 here]

To further assess the development gap in economic rationality, we control for non-cognitive skills (Heckman and Kautz, 2014) measured by the Big Five personality traits. These influential measures from psychology are derived from factor analysis of wide-ranging personality surveys. The Big Five factors are conscientiousness, openness, extraversion, agreeableness, and neuroticism. Barrick and Mount (1991) conclude that conscientiousness is the best predictor of economic outcomes (see, Block (2010) for a recent description and assessment of the Big Five). Most importantly, adding the controls for the Big Five has very little effect on the estimated coefficients, which suggests that the development gap in economic rationality is unlikely to be driven by unmeasured correlations between our economic rationality measures and personality traits.

The rest of the paper is organized as follows. Section 2 describes the experimental design and procedures. Section 3 summarizes the experimental results on economic rationality, and Section 4 compares them to results of the cognitive ability test. Section 5 contains some concluding remarks. The paper also uses data and technical online appendices for the interested reader. The experimental instructions are provided in Appendix I, and individual-level estimates are provided in Appendix II.[7]

2 The experiment

2.1 Subject pools

We conducted the experiments at UC Berkeley and the University of Dar es Salaam in Tanzania. The Experimental Social Science Laboratory (Xlab) at UC Berkeley draws its subjects from a large and diverse group of students and administrative staff; but all participants in our experiment were undergraduate students. Table 1 provides sociodemographic information on the subject pools. Overall, these data show that UC Berkeley students come from diverse socioeconomic backgrounds. As expected, the parents of the US subjects are much more educated. The US pool has a much higher proportion of females and younger subjects. All our results below are robust to the inclusion of controls for gender and age.

[Table 1 here]

2.2 Experimental procedures

The experimental procedures described below are identical to those used by Choi et al. (2007a). Full experimental instructions, including screen shots of the computer program dialog windows, are available in Appendix I. The experimental instructions were in English (the official languages of Tanzania are Swahili and English, and English is the language of instruction at the University of Dar es Salaam). Each experimental session consisted of 50 independent decision problems. In each decision problem,

[7] Appendix #: http://emlab.berkeley.edu/~kariv/CKST_I_A#.pdf.

a subject was asked to allocate tokens between two accounts, labeled x and y. The x account corresponds to the x-axis and the y account corresponds to the y-axis in a two-dimensional graph.

An example of a budget line in our experiment is the line AC drawn in Figure 3 below. Let x_s denote the demand for the security that pays off in state s and let p_s denote its price. The point B, which lies on the 45 degree line, corresponds to the safe allocation with a certain payoff $x_1 = x_2$. This allocation is consistent with infinite risk aversion. By contrast, point C represents an allocation in which all tokens are allocated to the cheaper security $x_1 = 0$ and $x_2 = 1/p_2$. This allocation is consistent with risk neutrality. Also note that the midpoint of the budget line corresponds to the allocation with equal expenditures, $p_1x_1 = p_2x_2$. This allocation is consistent with maximizing the logarithmic von Neumann-Morgenstern utility function.

[Figure 3 here]

Each choice involved choosing a point on a budget line of possible token allocations. Each decision problem started by having the computer select a budget line randomly from the set of lines that intersect at least one axis at or above the 50 token level and intersect both axes at or below the 100 token level. The budget lines selected for each subject in their decision problems were independent of each other and of the budget lines selected for other subjects in their decision problems. To choose an allocation, subjects used the mouse or the arrows on the keyboard to move the pointer on the computer screen to the desired allocation.[8] Choices were restricted to allocations on the budget constraint.[9]

The payoff at each decision round was determined by the number of tokens in the x account and the number of tokens in the y account. At the end of the round, the computer randomly selected one of the accounts, x or y. The two accounts were equally likely to be chosen. Each subject received the number of tokens allocated to the account that was chosen. During the course of the experiment, subjects were not provided with any information about the account that had been selected in each round. No subject reported any difficulty in understanding the procedures or using the computer program.

---

[8] It is of course possible that presenting choice problems graphically biases choice behavior in some particular way – and that would be a useful topic for research – but there is currently no evidence that this is the case. Ahn et al. (2013) extended the work in Choi et al. (2007a) on risk to settings with ambiguity. Choi et al. (2014) investigated the correlation between individual behavior and demographic and economic characteristics in the CentERpanel (a representative sample of more than 2,000 Dutch households). Among others, Fisman et al. (2007) employed a similar platform to study distributional preferences and produce very different behaviors. Since all experimental designs share the same graphical interface, we are building on expertise we have acquired in previous work.

[9] Choi et al. (2007a) also restricted choices to allocations on the budget line so that subjects could not dispose of payoffs. In Fisman et al. (2007) choices were not restricted to allocations on the budget constraint, but very few subjects violated budget balancedness by choosing strictly interior allocations. Restricting choices to allocations on the budget constraint makes the computer program easier to use.

At the end of the experiment, the computer selected one decision round for each subject, where each round had an equal probability of being chosen, and the subjects were paid the amount they had earned in the selected round. Payoffs were calculated in terms of tokens and then converted into the local currency. Each token was worth 0.5 USD in the US and 100 TZS (equivalent to 0.06 USD) in the low-stakes treatment in Tanzania, which are roughly comparable in purchasing power parity (PPP) terms. In addition, we also conducted a high-stakes treatment in Tanzania where each token was worth 1000 TZS, comparable to its worth in the US. Earnings were paid in private at the end of the experimental session.

2.3 IQ test and Big Five questionnaire

Before the experiment our subjects completed a non-incentivized IQ test and a Big Five personality traits questionnaire. For the IQ test, we used the matrix reasoning part of the Wechsler Adult Intelligence Scale (WAIS-IV) test, which is the most frequently administered IQ test. Subjects had 13 minutes to answer 26 multiple choice questions consisting of finding the natural next element following a sequence of five elements or finding the correct element to place in a 2×2 matrix with one missing element. Subjects saw one correct example of each type of questions before starting the test. For the personality traits questionnaire, we used the Big Five Inventory of John et al. (1991). The Big Five factors – conscientiousness, openness, extraversion, agreeableness, and neuroticism – are commonly used in psychology to measure human personality. Subjects are asked to evaluate the accuracy of statements related to these factors as descriptions of themselves on a five point scale. The personality scores are calculated using the procedure of John et al. (2008).

Table 2 below summarizes the results of the IQ-test and the Big Five personality traits questionnaire in the two subject pools. We see qualitatively similar patterns both in levels and in correlations of the Big Five measures. Note, however, that the variances of the measures are slightly lower in Tanzania, and that in Tanzania "openness" is negatively correlated with "agreeableness" and positively with "conscientiousness" – both these correlations are close to zero among the US subjects. The correlations between IQ and the Big Five measures are not large in either sample. The proportions of correct answers on the IQ test are quite different in the two samples: In Tanzania, the level is much lower, and the standard deviation is higher (more below).

[Table 2 here]

3 Experimental results

3.1 Utility maximization

The most basic question to ask about choice data is whether it is consistent with individual utility maximization (Samuelson, 1947). If budget sets are linear (as in our experiment), classical revealed preference theory provides a direct test. Afriat's (1967) theorem tells us that if a *finite* data set generated by an individual's choices satisfies the Generalized Axiom of Revealed Preference (GARP), then the data can be rationalized by a well-behaved (that is, piecewise linear, continuous, increasing, and concave) utility function.[10] GARP requires that if x^i is indirectly revealed preferred to x^j, then x^j is not *strictly* directly revealed preferred to x^i ($p^j \cdot x^i \geq p^j \cdot x^j$) where p^i denotes the i-th observation of the price vector and x^i denotes the associated allocation.

Although testing conformity with GARP is conceptually straightforward, there is an obvious difficulty: GARP provides an exact test of utility maximization – either the data satisfy GARP or they do not – but individual choices frequently involve at least some errors: subjects may compute incorrectly, or execute intended choices incorrectly, or err in other less obvious ways. To account for the possibility of errors, we assess how closely individual choice behavior complies with GARP by using Afriat (1972) Critical Cost Efficiency Index (CCEI), which measures the fraction by which each budget constraint must be shifted in order to remove all violations of GARP.

By definition, the CCEI is between zero and one: indices closer to one mean that the data are closer to perfect consistency with GARP and hence to perfect consistency with utility maximization. To clarify, Figure 4 below illustrates one such adjustment involving two portfolios, x^1 and x^2. It is clear that x^1 is revealed preferred to x^2 because $p^1 \cdot x^1 > p^1 \cdot x^2$, yet x^1 is cheaper than x^2 at the prices at which x^2 is purchased, $p^2 \cdot x^1 < p^2 \cdot x^2$. The "least costly" shift of the budget constraint that removes the violation is through x^2, since $A/B > C/D$. The CCEI can thus be interpreted as saying that the individual is 'wasting' as much as $1 - A/B$ of their income by making an inconsistent choice.[11]

[Figure 4 here]

To calibrate the CCEI, we follow Bronars (1987), which builds on Becker (1962), and compare the behavior of our actual subjects to the behavior of simulated subjects who randomize uniformly on each budget line. Mean CCEI score for a random sample of 25,000 simulated subjects is only 0.600. Furthermore, a large majority of actual subjects have CCEI scores above 0.90, while less than 0.5 percent of simulated subjects have CCEI scores that high. Our experiment is thus sufficiently powerful to exclude the possibility that consistency is the accidental result of random behavior. Therefore, the consistency of our subjects' behavior under these conditions is not accidental.[12]

[10]This statement of the theorem follows Varian (1982, 1983), who replaced the condition Afriat (1967) called *cyclical consistency* with GARP. The papers by Vermeulen (2012), Afriat (2012), Diewert (2012) and Varian (2012) published in a special issue of the *Economic Journal* on the Foundations of Revealed Preference provide an excellent overview and a discussion of some recent developments in this literature.

[11]We refer the interested reader to Choi et al. (2007a,b) for further details on the testing for consistency with GARP. The computer program and details of the algorithm are available from the authors upon request.

[12]The power of the experiment depends on two factors. The first is that the range of choice sets is

Complementing the graphical presentation in Figure 1, Table 3A below reports summary statistics and percentile values for the CCEI scores. Of the 126 US subjects, 108 (85.7 percent) have CCEI scores above 0.9, and of those, 83 subjects (65.9 percent) have values above 0.95. The corresponding numbers for the 216 Tanzanian subjects, are 113 (52.3 percent) and 88 (40.7 percent), respectively. We interpret these numbers as confirmation that many subjects exhibit behavior that appears to be *almost* optimizing in the sense that their choices nearly satisfy GARP, so that the violations are minor enough to be ignored for the purposes of recovering preferences or constructing appropriate utility functions.

However, the percentile distributions reported in Table 3A also show that the distribution of CCEI scores is skewed to the right for the US subjects, indicating greater overall consistency with utility maximization. Mean CCEI scores across all subjects are 0.950 in the US and 0.856 and 0.869 with low and high stakes in Tanzania, respectively. The mean CCEI scores imply that by making inconsistent choices the Tanzanian subjects 'waste' as much as 9.4 (low stakes) or 8.1 (high stakes) percentage points more of their earnings relative to the US subjects. For comparison, the CCEI scores in similar (unpublished) experiments at Yale Law School, the University of California, Los Angeles (UCLA), and the University of Bergen, Norway averaged 0.982, 0.932 and 0.936, respectively. Table 3B reports summary statistics and percentile values for the CCEI for these subject pools.[13]

[Table 3 here]

We next turn to individual-level regression analyses that examine the consistency scores more systematically. We define indicators for both the Tanzania sample and for the high-stakes experimental treatment (in Tanzania). We also include age, gender and the Big Five personality traits as controls. The dependent variable is the subject's CCEI score. Table 4 below reports the results of ordinary least squares (OLS) estimation. The results show that the differences in consistency with utility maximization between the US and Tanzania subjects are not driven by the personality traits or demographic profiles as the inclusion of these controls leaves the point estimate of the

generated so that budget lines cross frequently. The second is that the number of decisions made by each subject is large. Choi et al. (2007b) generate a random sample of simulated subjects who maximize a utility function with error (the likelihood of error is assumed to be a decreasing function of the utility cost of an error). The analysis demonstrates that if utility maximization is not in fact the correct model, then our experiment is sufficiently powerful to detect it.

[13]The experiment of Choi et al. (2014) consisted of 25, rather than 50, decision problems so the CCEI scores are not directly comparable. Choi et al. (2014) combine the actual data from the experiment and the mirror-image of these data obtained by reversing the prices and the associated allocation for each observation. Choi et al. (2014) report that, in the CentERpanel sample (representative of the Dutch-speaking population in the Netherlands), the CCEI scores for the combined dataset involving 50 decisions, like in our experiment, averaged only 0.733. They also find that, in the combined dataset, subjects with primary or pre-vocational secondary education on average waste as much as 6.8 percentage points more of their earnings by making inconsistent choices relative to subjects with vocational college or university education.

coefficient on Tanzania little changed. There are also no significant differences in the CCEI scores between the low- and high-stakes treatments in Tanzania. We generate virtually identical parameter values using a Tobit specification, which allows for censoring of the CCEI at one. We also employ quantile regressions that are less sensitive to extreme values. The quantile regressions for the 25th, 50th and 75th percentiles similarly detect significant differences between the two subject pools. These results are omitted to economize on space.

[Table 4 here]

3.2 First-order stochastic dominance

Next we ask whether choices are consistent with the dominance principle – that is, the requirement that, regardless of risk attitudes, an allocation should be preferred to another if it yields unambiguously higher monetary payoff. A simple violation of FOSD is illustrated in Figure 5 below. The budget line is defined by the straight line *AE* and the axes measure the value of a possible allocation in each of the two states. The point *B*, which lies on the 45 degree line, corresponds to the safe allocation with a certain outcome. The individual chooses allocation x (position along *AB*), but could have chosen any allocation x' (position along *CD*) such that $F_{x'} \leq F_x$, where $F_{x'}$ and F_x are the resulting payoff distributions. If this individual only cares about the distribution of monetary payoffs, then he will be willing to pay a positive price for a lottery yielding $F_{x'} - F_x$, which has only nonpositive payoffs (that is, for a lottery in which each account has an equal probability of being chosen). Notice that any decision to allocate fewer tokens to the cheaper security (that is, corresponding to a position along *AB*) violates FOSD, but need not involve a violation of GARP, whereas any decision to allocate more tokens to the cheaper security (that is, corresponding to a position along *BE*) never violates FOSD.

[Figure 5 here]

We use expected payoff calculations to assess how closely individual choice behavior complies with FOSD. Suppose that we observe an individual choosing allocation x at prices p where $F_{x'} \leq F_x$ for some x' such that $p \cdot x' = 1$. The extent to which allocation x violates FOSD can be measured by its expected return as a fraction of the *maximal* expected return that could be achieved by choosing an allocation x'. The construction of this violation index is also illustrated in Figure 5 above. The point *D* corresponds to the allocation x' with the highest expected return, yielding the largest upward probabilistic shift (referring to Figure 5, the outcome "α points" is shifted up to "γ points" and the outcome "β points" remains unchanged). This suggests the following approach. For each observation (p^i, x^i), if no feasible allocation dominates the chosen allocation, then it has the highest possible value of one. Otherwise, it has a value of less than one; specifically $(\alpha + \beta)/(\gamma + \beta)$ (the two states are equally likely). Since a single number is desired for each subject, we average this violation index across all decision problems. Table 3C above reports summary statistics and percentile values for the FOSD scores.[14]

Over all subjects, the FOSD scores average 0.992 in the US and 0.975 and 0.978 with low- and high-stakes treatments in Tanzania, respectively. Out of the 126 US subjects, 123 subjects (97.6 percent) have FOSD scores above 0.95, and of the 216 Tanzanian subjects, 185 subjects (85.8 percent) have scores that high. Overall, the choices made by subjects in our experiment show very low rates of FOSD violations. Nevertheless, there is also some heterogeneity in the FOSD scores within and across subject pools. Table 5 reports the results of OLS specifications. The dependent variable is the subject's FOSD score. We again include age, gender and the Big Five personality traits as controls. The results show that the small difference in our FOSD violations scores between the US and Tanzania subjects is statistically significant. We generate virtually identical parameter values using Tobit specifications. We also note that there is considerable heterogeneity in the CCEI and FOSD, and that their values are positively correlated ($\rho = 0.485$ and $\rho = 0.793$ in the US and Tanzania, respectively). Finally, there are no significant differences in FOSD violations between the low- and high-stakes treatments in Tanzania.

[Table 5 here]

3.3 Expected-utility maximization

We now turn to the next level of analysis involving estimation of individual-level, parametric utility functions generated by the model proposed by Gul (1991):

$$U(x_1,x_2) = \min\{\alpha u(x_1) + u(x_2), u(x_1) + \alpha u(x_2)\},$$

where $\alpha \geq 1$ is a parameter measuring disappointment aversion and $u(\cdot)$ is the utility of consumption in each state (the two states are equally likely). In this interpretation, the safe allocation $x_1 = x_2$ is taken to be the reference point. If $\alpha > 1$ the indifference curves of $U(x_1,x_2)$ have a *kink* at the 45 degree line where $x_1 = x_2$, and if $\alpha = 1$ we have the standard EUT representation.[15] Figure 6 below illustrates a "kinked" indifference curve. For EUT, in contrast, the indifference curves are smooth everywhere.

[14]Choi et al. (2014) provide a unified measure of the extent of GARP violations *and* violations of FOSD by combining the actual data from the experiment and the mirror-image of these data obtained by reversing the prices and the associated allocation for each observation (that is, assuming that if (x_1,x_2) is chosen subject to the budget constraint $p_1x_1 + p_2x_2 = 1$, then (x_2,x_1) would have been chosen subject to the mirror-image budget constraint $p_2x_1 + p_1x_2 = 1$). Choi et al. (2014) compute the CCEI for this combined data set, and compare that number to the CCEI for the actual data. This measures the extent of GARP violations and violations of stochastic dominance (for a given subject). We favor a "low-tech" approach, which provides an independent measure of FOSD and is conceptually straightforward.

[15]These preferences can also be generated by a rank-dependent utility function (Quiggin, 1993). As proposed by Quiggin (1982), the tendency to equate the demand for the pair of securities can also be explained by pessimism (overweighting the probabilities of low payoffs and underweighting the

Maximizing $U(x_1,x_2)$ subject to a budget constraint yields a non-linear demand curve. If the security prices are very different, then the optimum is the boundary portfolio, $(x_1,0)$ or $(0,x_2)$, with the larger expected payoff. If the security prices are very similar, then the optimum is the safe portfolio $x_1 = x_2$. In these cases, the optimal choice is insensitive to small price changes. For log-price ratios that are neither extreme nor close to zero, the optimum is an intermediate portfolio and the choice is sensitive to small changes in the risk-return tradeoff.[16]

[Figure 6 here]

We estimate individual-level utility functions directly from the data using CARA

$$u(x) = -e^{-\gamma x}$$

and CRRA

$$u(x) = x^{1-\rho}/((1-\rho)$$

specifications.[17] In estimating the parameters, we impose a number of restrictions.

- First, Afriat's (1967) theorem tells us that when a rationalizing utility function exists, it may be chosen to be well behaved (piecewise linear, continuous, increasing, and concave). We therefore restrict the parameters so that preferences are always risk $\gamma,\rho \geq 0$ and disappointment averse $\alpha \geq 1$.
- Second, because of computational difficulties when the parameters are large, we also impose the restriction $\alpha \leq 10$ in both specifications, and that $\gamma \leq 1$ and $\rho \leq 5$. This involves minimal loss of fit, since the predicted choices with such a high level of risk aversion and/or disappointment aversion are virtually identical.

probabilities of high payoffs). The disappointment aversion model of Gul (1991) is identical to rank-dependent utility in the present design involving two states of nature $s = 1,2$. With more than two states, in Gul's (1991) model the indifference curves have a kink only at the safe allocation $x_s = x_{s'}$ for all s and s', whereas in the rank-dependent utility model the indifference curves have a kink at all allocations where $x_s = x_{s'}$ for some s and s'.

[16] We note that GARP implies rationality in the sense of a complete, transitive preference ordering, but it does not imply the Savage axioms. Using revealed preference methods to test whether the data are consistent with a utility function with some special structure, particularly EUT, is beyond the scope of this paper. Diewert (2012) provides a combinatorial condition that is necessary and sufficient for extending Afriat's (1967) Theorem to choice under uncertainty. Unfortunately, to the best of our knowledge, Diewert's (2012) condition is not a simple adjustment to the usual tests, which are all computationally intensive for large datasets like our own.

[17] The power function is not well defined for the boundary allocations. We incorporate the boundary observations $(1/p_1,0)$ or $(0,1/p_2)$ into our estimation by replacing the zero component by a small consumption level such that the relative demand x_1/x_2 is either $1/\omega$ or ω, respectively. The minimum ratio is $\omega = 0.001$. The selected level did not substantially affect the estimated coefficients for any subject. The exponential function accommodates boundary allocations even when initial income is zero.

The estimation is carried out using both non-linear least squares (NLLS) and maximum likelihood (ML) methods. To economize on space, we only report the NLLS estimation results. See Choi et al. (2007a) for precise details on the estimation technique. Choi et al. (2007a) also show that the estimation does a good job of fitting the data and allowing us to classify different types of behavior.

To economize on space, the individual-level estimates are relegated to Appendix II. Appendix II also lists, by subject, the number of violations of GARP, and also reports the values of the CCEI and FOSD scores. Subjects are ranked according to (descending) CCEI scores. Table 6 provides a population-level summary of the individual-level estimation results by reporting summary statistics and percentile values for the full sample. The distributions are similar for the subsample of subjects with CCEI scores above 0.80, as reported in Table 7. Although in both the CARA and CRRA specifications there is considerable heterogeneity in both parameters, the *majority* of the subjects in both samples are well approximated by preferences consistent with EUT $\hat{\alpha} = 1$, a much higher proportion than found in other experimental studies.[18] There is also considerable heterogeneity in subjects' risk preferences in both the disappointment-averse and the disappointment-neutral subsamples, but they are within the range of recent estimates of risk aversion (more below).

[Table 6 here]
[Table 7 here]

Since some subjects are better described by one or other of the models we estimate, the two classes of parametric utility functions do not yield exactly the same results. Using the CARA specification (top panels), we cannot reject the hypothesis that $\hat{\alpha} = 1$ for a total of 103 US subjects (81.7 percent) at the 95 percent significance level. The remainder appear to have significant degrees of disappointment aversion $\hat{\alpha} > 1$. The corresponding numbers for the Tanzanian subjects with low and high stakes are 95 (89.6 percent) and 98 (89.1 percent), respectively. Using the CRRA specification (bottom panels), we cannot reject the hypothesis that $\hat{\alpha} = 1$ for a total of 82 US subjects (65.1 percent) at the 95 percent significance level. The corresponding numbers for the Tanzanian subjects with low and high stakes are 80 (75.5 percent) and 73 (66.4 percent), respectively. The behavior of these subjects is consistent with EUT. The results are similar for the subsample of subjects with CCEI scores above 0.80.

Overall, our individual-level analysis shows that preferences vary widely across subjects and that there is considerable heterogeneity in both parameters in both the

[18] We will not attempt to review the large and growing literature on choice under risk. Camerer (1995) provides a comprehensive discussion, though now somewhat dated, of the experimental and theoretical work, and Starmer (2000) provides a more recent review that focuses on evaluating non-EUT theories. For the most part, these experimental investigations use several pair-wise choices to test EUT and its various generalizations. Generally speaking, previous experimental work especially in development countries has, on the one hand, collected only a few decisions from each subject and, on the other, presented subjects with an extreme binary choice, designed to discover violations of specific theories.

CARA and the CRRA specifications. Finally, we turn to regression analyses. We define indicators for both the Tanzania sample and the high-stakes experimental treatment (in Tanzania), and we include age, gender and the Big Five personality traits as controls. The dependent variable is an indicator of consistency with EUT $\hat{\alpha} = 1$. Tables 8 and 9 report the results for the CARA and CRRA specifications, respectively. We find no significant differences in the fraction of subjects consistent with EUT between the two samples.

[Table 8 here]
[Table 9 here]

4 Economic rationality versus cognitive ability

Our experimental results stand in sharp contrast to a very large development gap in rationality if measured by a test of cognitive ability (IQ), which is often used as a proxy for how rational individuals are in making economic decisions (see, for example, Dohmen et al. (2010)).[19] Figure 1 above summarizes the cumulative distributions of CCEI scores and Figure 2 summarizes the cumulative distributions of IQ scores in both samples. Table 10 summarizes the 25^{th}, 50^{th} and 75^{th} percentile values of the CCEI, FOSD and IQ scores for the full sample and reports the fractions of Tanzanian and US subjects whose scores are below these values. Most notably, 82.7 percent of the Tanzanian subjects have IQ scores below the joint median, compare to only 7.9 percent of the US subjects. The corresponding numbers for the CCEI scores are 59.3 and 34.1 percent and for the FOSD scores 63.0 and 27.8 percent, respectively. Finally, in Table 11, we repeat the OLS estimation reported in Tables 4 and 5 using the subjects' IQ scores instead of the CCEI or FOSD scores as the dependent variable.

[Table 10 here]
[Table 11 here]

We note that the measures of economic rationality tested in this paper – utility maximization, stochastic dominance, expected-utility maximization – offer a theoretically disciplined metric for quality of economic decisions. These measures have

[19] In a non-strategic setting, Burks et al. (2009) report a correlation of approximately 0.22 between IQ and switching more than once in Holt and Laury's (2002) multiple price list experiments. In a strategic setting, Gill and Prowse (2014) investigate the correlation between cognitive ability and non-cognitive skills and learning to play equilibrium in a repeated p-beauty contest experiment. They estimate a structural model based on level-k reasoning (Stahl and Wilson, 1995; Nagel, 1995; Duffy and Nagel, 1997), and find a systematic relationship between subjects' cognitive ability and their level-k types. We, by contrast, test the assumption of rational choice on individual behavior. Equilibrium is a more restrictive concept. It assumes that each player chooses a strategy that maximizes his payoff taking as given the strategies of his opponents. In a level-k model, it is assumed that level-k subjects believe opponents behave as level-$(k-1)$ and best respond to those beliefs. The role of beliefs in games complicates the analysis of the rationality of strategic behavior. Crawford et al. (2013) provide a comprehensive survey of the applications of level-k models.

well-established economic interpretations and classical theory tells us whether we have enough data to make them statistically useful. There is no comparable, theoretically disciplined means for using, interpreting, and evaluating an IQ test. Another advantage of the economic measures of rationality over IQ tests is that the former are easily portable to a variety of choice problems. We can thus make domain-specific predictions and study a comparable measure of decision-making quality across domains. For example, the same experimental platform and analysis of individual choice problems can be used to study consumption over time. In addition, our experimental task does not involve right and wrong answers, and does not demand any outside particular knowledge or expertise. Virtually all IQ tests have right and wrong answers, and thus draw on outside knowledge and depend on preferences for obtaining certain skills. For example, Raven's matrix tests, spatial relations tests, and number series tests, all have right and wrong answers and all involve skills developed by learning mathematics. Stanovich (2009) provides a critique of IQ tests as measures of decision-making ability.

5 Concluding remarks

We tested the touchstones of rationality in a choice under risk experiment. Our subjects are students from UC Berkeley, one of the best universities in the United States, and the University of Dar es Salaam, one of the best universities in Africa. The Tanzanian and the US subjects differ substantially in sociodemographic and economic backgrounds and face very different economic prospects. Nevertheless, they represent the same 'slice' of the most able in their societies. Since innovation and the resulting economic growth require a skilled workforce, the students of UC Berkeley and the University of Dar es Salaam are among the ones who will drive economic growth once they graduate into the world.

Our results are summarized as follows. We find that the US subjects tend more toward utility maximization than the Tanzanian subjects. We use the CCEI (and other goodness-of-fit indices) to measure the extent of GARP violation. The magnitudes imply that the Tanzanian subjects on average waste as much as 8.7 percentage points more of their earnings by making inconsistent choices than the US subjects. This provides a quantifiable and economically interpretable measure of the development gap in rationality. Tests of monotonicity with respect to FOSD, expected-utility maximization, and small-stakes risk neutrality indicate a very small or no development gap in rationality.

References

Acemoglu, Daron and James Robinson (2012). *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, Crown Business.

Afriat, S. N. (1967). “The construction of a utility function from expenditure data,” *International Economic Review*, 8(1): 67–77.

Afriat, S. N. (1972). “Efficiency estimation of production functions,” *International Economic Review*, 13(3): 568–598.

Afriat, S. N. (2012). “Afriat's theorem and the index number problem,” *Economic Journal*, 122(560): 295–304.

Ahn, David, Syngjoo Choi, Douglas Gale, and Shachar Kariv (2013). Forthcoming, *Quantitative Economics*.

Barrick, Murray R. and Michael K. Mount (1991). “The Big Five personality dimensions and job performance: A meta-analysis,” *Personnel Psychology*, 44(1): 1–26.

Becker, Gary S. (1962). “Irrational behavior and economic theory,” *Journal of Political Economy*, 70(1): 1–13.

Block, Jack (2010). “The five-factor framing of personality and beyond: Some ruminations,” *Psychological Inquiry*, 221(1): 2–25.

Bronars, Stephen G. (1987). “The power of nonparametric tests of preference maximisation,” *Econometrica*, 55(3): 693–698.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini (2009). “Cognitive skills affect economic preferences, strategic behavior, and job attachment,” *Proceedings of the National Academy of Sciences*, 106(19): 7745–7750.

Camerer, Colin (1995). “Individual decision making,” in John Henry Kagel and Alvin E. Roth (eds.), “Handbook of Experimental Economics,” Princeton University Press.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv (2007a). “Consistency and heterogeneity of individual behavior under uncertainty,” *American Economic Review*, 97(5): 1921–1938.

Choi, Syngjoo, Raymond Fisman, Douglas M. Gale, and Shachar Kariv (2007b). “Revealing preferences graphically: An old method gets a new tool kit,” *American Economic Review*, 97(2): 153–158.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman (2014). “Who is (more) rational?” *American Economic Review*, 104(6): 1518–1550.

Crawford, Vincent P., Miguel A. Costa-Gomes, and Nagore Iriberri (2013). “Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications,” *Journal of Economic Literature*, 51(1): 5–62.

Diewert, W. Erwin (2012). “Afriat’s theorem and some extensions to choice under uncertainty,” *Economic Journal*, 122(560): 305–331.

Dohmen, Thomas, Armin Falk, and David Huffman (2010). “Are risk aversion and impatience related to cognitive ability?” *American Economic Review*, 100(3): 1238–1260.

Duffy, John and Rosemarie Nagel (1997). “On the robustness of behaviour in experimental ‘beauty contest’ games,” *Economic Journal*, 107(445): 1684–1700.

Fisman, Raymond J., Shachar Kariv, and Daniel Markovits (2007). “Individual preferences for giving,” *American Economic Review*, 97(5): 1858–1876.

Gill, David and Victoria Prowse (2014). “Cognitive ability, character skills and learning to play equilibrium: A level-k analysis,” Department of Economics Discussion Paper Series no 712, University of Oxford.

Gul, Faruk (1991). “A theory of disappointment aversion,” *Econometrica*, 59(3): 667–686.

Hadar, Josef and William R. Russel (1969). “Rules for ordering uncertain prospects,” *American Economic Review*, 59(1): 25–34.

Hanushek, Eric A. and Ludger Woessmann (2007). “The role of school improvement in economic development,” NBER Working Papers 12832, National Bureau of Economic Research.

Heckman, James J. and Tim Kautz (2014). “Achievement tests and the role of character in American life,” in James J. Heckman, John Eric Humphries, and Tim Kautz (eds.), “The Myth of Achievement Tests: The GED and the Role of Character in American Life,” University of Chicago Press.

Holt, Charles A. and Susan K. Laury (2002). “Risk aversion and incentive effects,” *American Economic Review*, 92(5): 1644–1655.

John, O. P., E. M. Donahue, and R. L. Kentle (1991). “The Big Five Inventory—versions 4a and 54,” Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, Oliver P., Laura P. Naumann, and Christopher J. Soto (2008). “Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues,” in Oliver P. John, Richard W. Robins, and Lawrence A. Pervin (eds.), “Handbook of Personality: Theory and Research,” chapter 4, New York: Guildford Press, 3 edition, pp. 114–158.

Lucas, Robert Jr. (1988). “On the mechanics of economic development,” *Journal of Monetary Economics*, 22(1): 3–42.

Mankiw, N. Gregory, David Romer, and David N. Weil (1992). “A contribution to the empirics of economic growth,” *Quarterly Journal of Economics*, 107(2): 407–437.

Nagel, Rosemarie (1995). “Unraveling in guessing games: An experimental study,” *American Economic Review*, 85(5): 1313–1326.

Quiggin, John (1982). “A theory of anticipated utility,” *Journal of Economic Behavior & Organization*, 3(4): 323–343.

Quiggin, John (1990). “Stochastic dominance in regret theory,” *Review of Economic Studies*, 57(3): 503–511.

Quiggin, John (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model*, Kluwer Academic Publishers.

Samuelson, Paul Anthony (1947). *Foundations of Economic Analysis*, Harvard University Press.

Savage, Leonard J. (1954). *The Foundations of Statistics*, New York: Wiley.

Stahl, Dale O. and Paul W. Wilson (1995). “On players’ models of other players: Theory and experimental evidence,” *Games and Economic Behavior*, 10(1): 218–254.

Stanovich, Keith E. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought*, Yale University Press.

Starmer, Chris (2000). “Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk,” *Journal of Economic Literature*, 38(2): 332–382.

Varian, Hal R. (1982). “The nonparametric approach to demand analysis,” *Econometrica*, 107(4): 945–972.

Varian, Hal R. (1983). “Non-parametric tests of consumer behavior,” *Review of Economic Studies*, 50(1): 99–110.

Varian, Hal R. (1990). “Goodness-of-fit in optimizing models,” *Journal of Econometrics*, 46(1-2): 125–140.

Varian, Hal R. (1991). “Goodness of fit for revealed preference tests,” Technical Report 13, Michigan – Center for Research on Economic & Social Theory.

Varian, Hal R. (2012). “Revealed preference and its applications,” *Economic Journal*, 122(560): 332–338.

Vermeulen, Frederic (2012). “Foundations of revealed preference: Introduction,” *Economic Journal*, 122(560): 287–294.

Wakker, Peter (1993). “Savage’s axioms usually imply violation of strict stochastic dominance,” *Review of Economic Studies*, 60(2): 487–493.

Figure 1. The cumulative distributions of CCEI scores

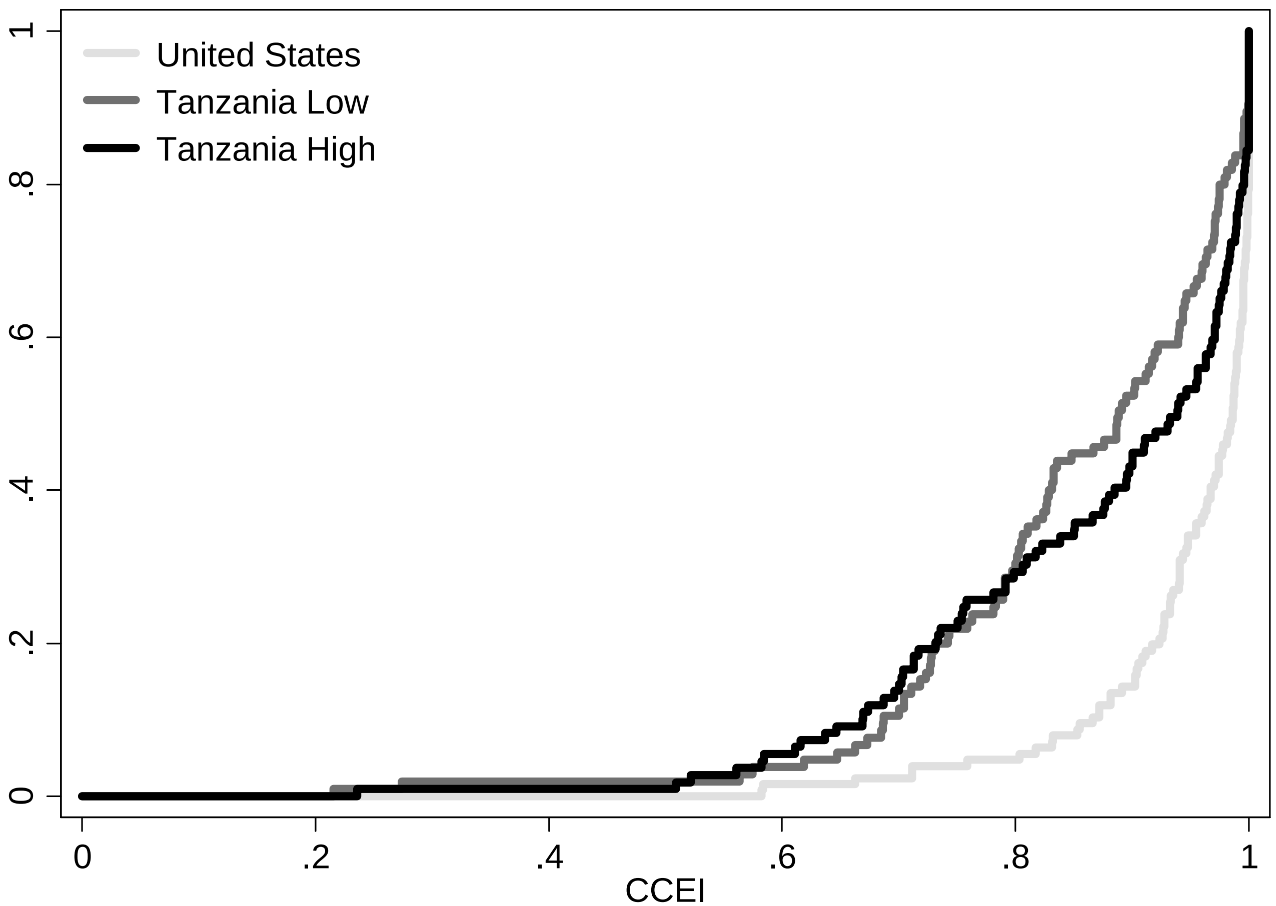


Figure 2. The cumulative distributions of IQ scores

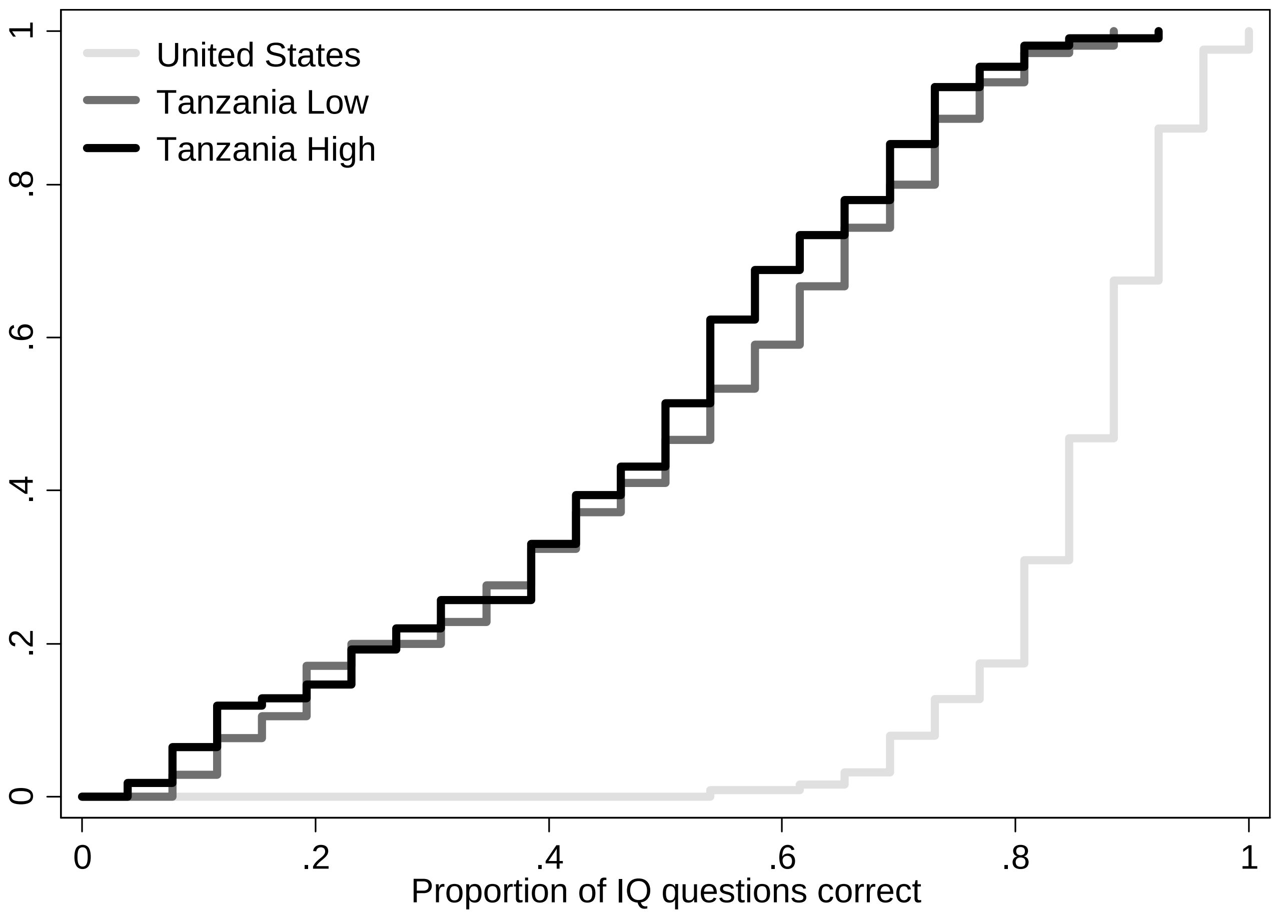
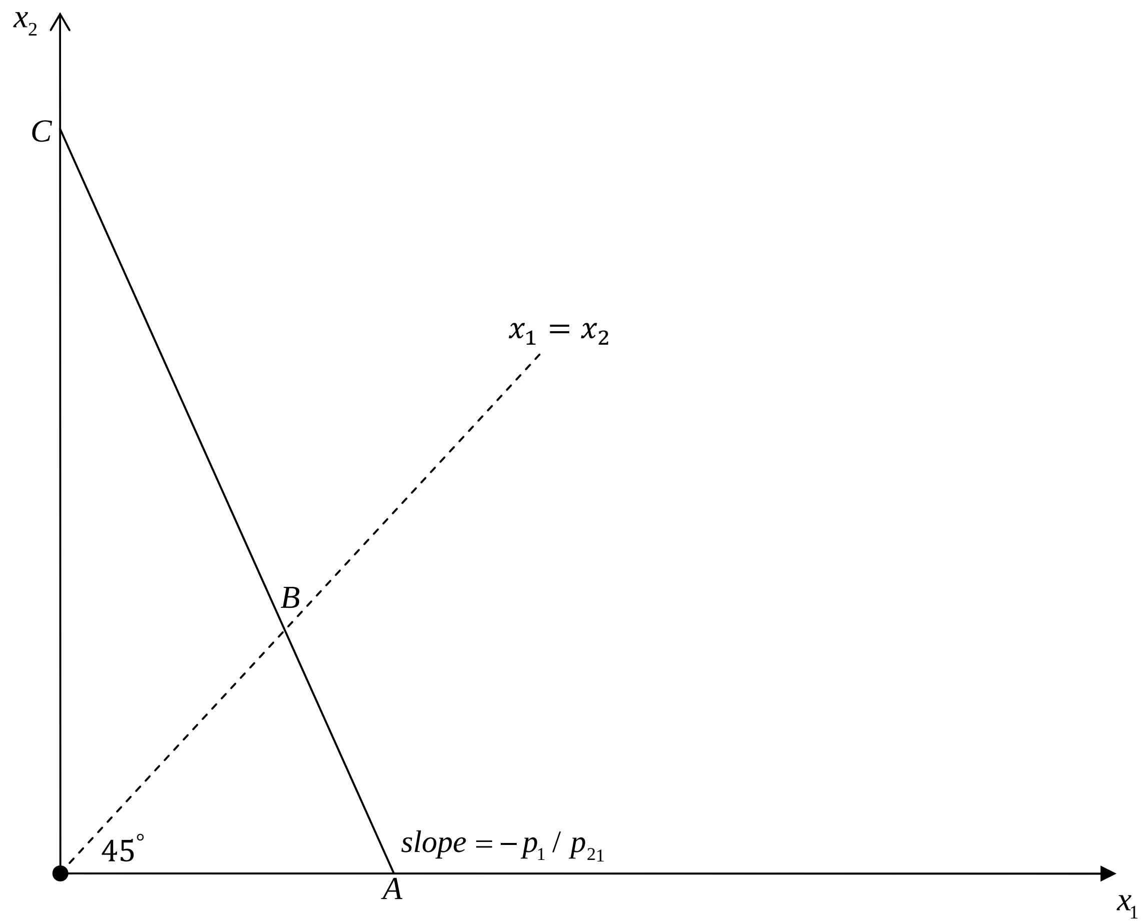
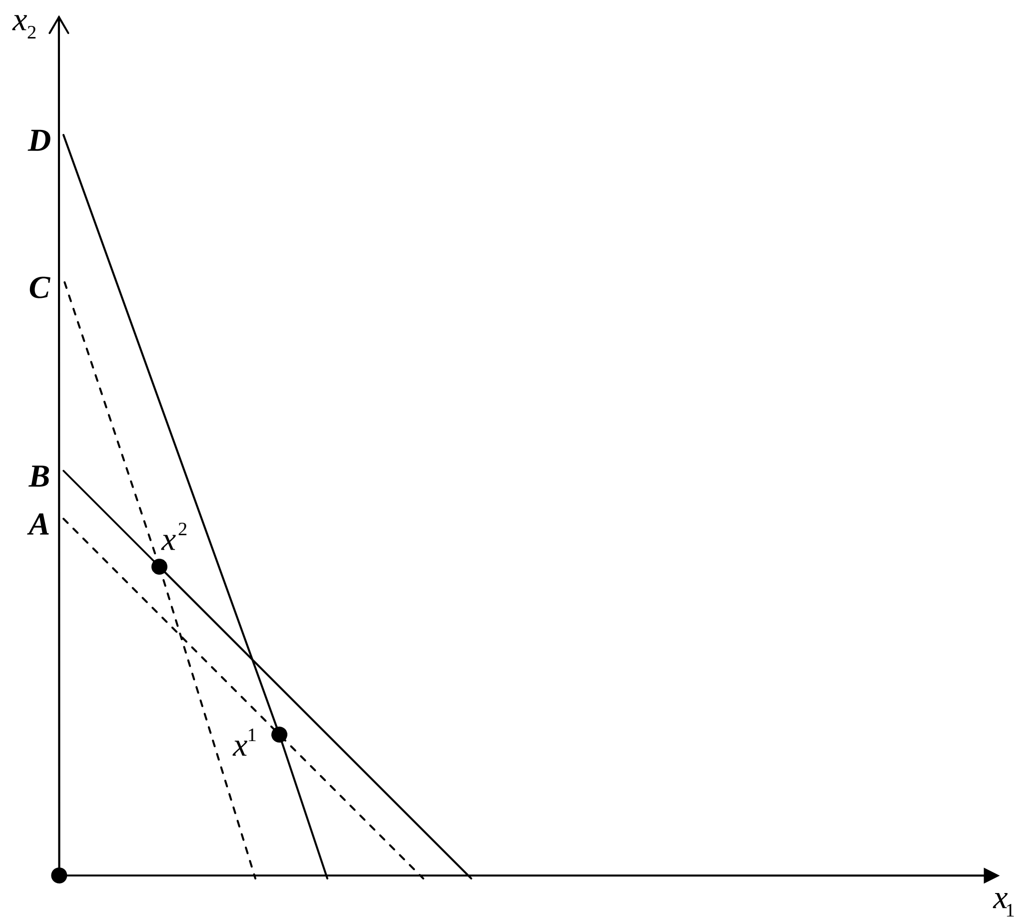


Figure 3: An example of a budget line



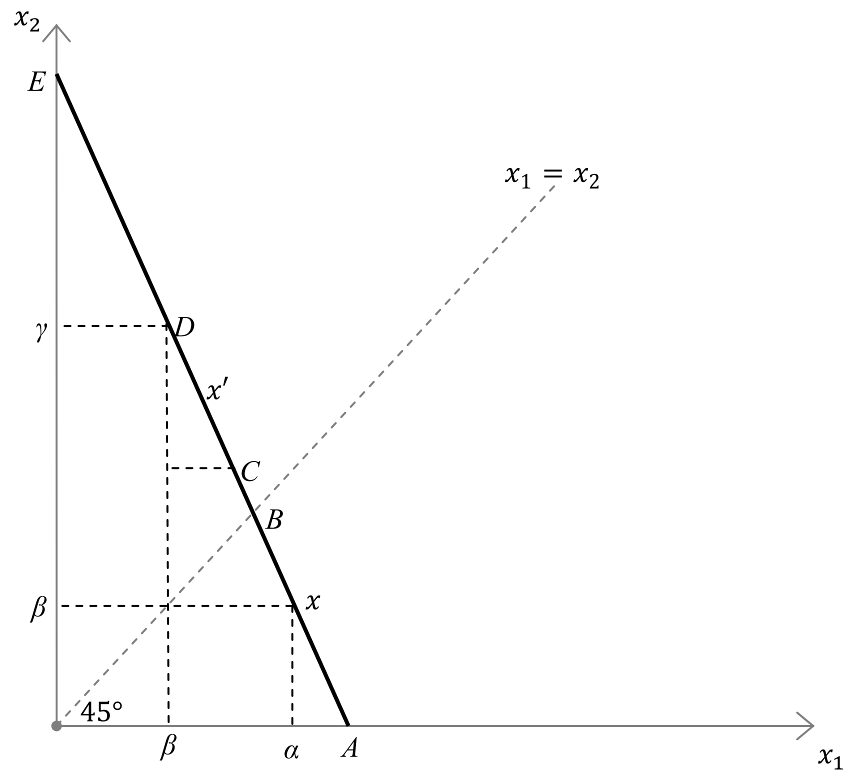
An example of a budget line in the experiment. The point *B*, which lies on the 45 degree line, corresponds to the safe allocation with a certain payoff. This allocation is consistent with infinite risk aversion. By contrast, point *C* represents an allocation in which all tokens are allocated in the cheaper security. This allocation is consistent with risk neutrality. Any decision to allocate *more* tokens to the more expensive security (position along *AB*) is a violation of FOSD (more below).

Figure 4. The construction of the CCEI for a simple violation of GARP



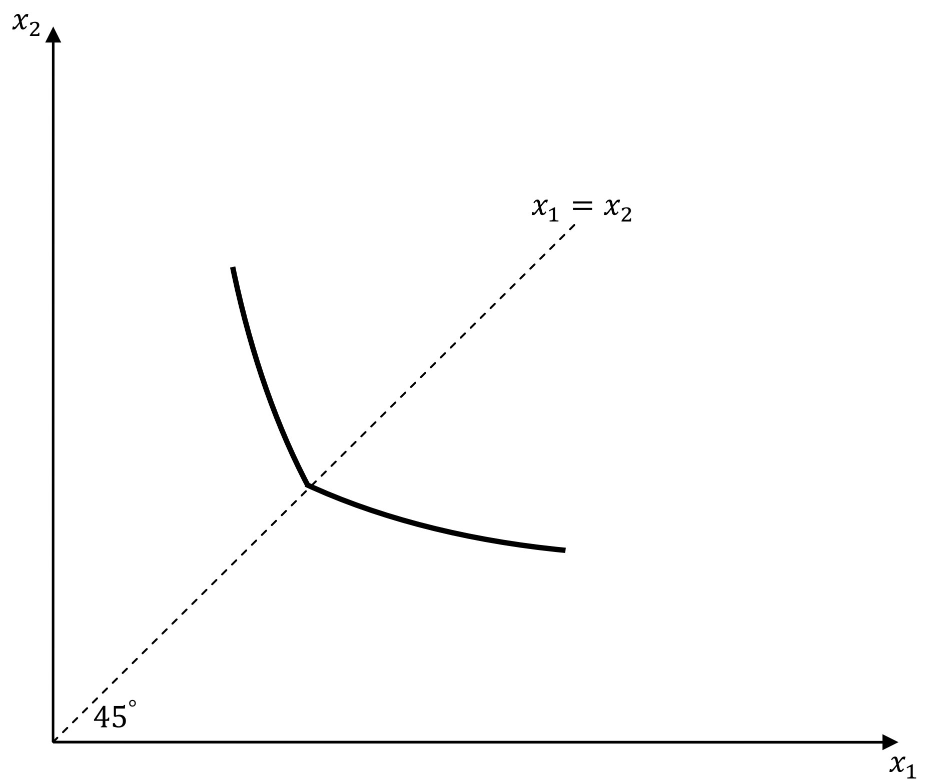
Here we have a violation of the Weak Axiom of Revealed Preference (WARP) since x^1 is directly revealed preferred to x^2 and x^2 is directly revealed preferred to x^1. A perturbation $A/B > C/D$ of the budget line through allocation x^1 removes the violation. The CCEI can thus be interpreted as saying that the individual is `wasting' as much as $1 - A/B$ of the income by making an inconsistent choice.

Figure 5. A violation of first-order stochastic dominance



The individual can choose any allocation x' (position along CD) but prefers allocation x (position along AB) such that $F_{x'} \leq F_x$ where $F_{x'}$ and F_x are the resulting payoff distributions. Violations of first-order stochastic dominance may reasonably be regarded as errors, regardless of risk attitudes---that is, as a failure to recognize that some allocations yield payoff distributions with unambiguously lower returns.

Figure 6. An illustration of an indifference curve for a disappointment averse individual (Gul, 1991)



The indifference curves have a kink at the 45 degree line. The nature of the kink is determined by the individual's disappointment aversion (α).The shape of the indifference curve on either side of the 45 degree line is determined by the individual's attitude toward risk (ρ in the CRRA specification and γ in the CARA specification). Note that an ambiguity averse individual chooses safe allocations satisfying $x_1 = x_2$ when the security prices, are sufficiently similar. For EUT (α = 1), in contrast, the indifference curves are smooth everywhere.

Table 1. Sociodemographic information

	US	Tanzania		
		Low	High	p-value
Age	20.611	23.154	23.361	0.45
	(0.38)	(0.18)	(0.21)	
Fraction Female	0.706	0.311	0.355	0.50
	(0.041)	(0.045)	(0.046)	
Fraction with own income from work	0.317	0.085	0.046	0.24
	(0.042)	(0.027)	(0.020)	
Fraction with two parents with secondary school	0.849	0.387	0.330	0.36
	(0.032)	(0.048)	(0.045)	
Fraction with at least one parent with college/university	0.825	0.311	0.303	0.86
	(0.034)	(0.045)	(0.044)	
# of obs.	126	106	110	

Self reported. Standard errors in parentheses. The *p*-values are for tests of the hypothesis that the means in the two treatment groups in Tanzania are the same, and are from *t*-tests for mean age and Pearson Chi-square tests for the other variables.

Table 2. IQ-test and the Big Five personality traits questionnaire

A. United States

	Mean	Sd	Correlations					
			A	C	E	N	O	IQ
Agreeableness (A)	0.60	0.49	1					
Conscientousness (C)	0.40	0.55	0.200	1				
Extroversion (E)	0.15	0.64	0.120	0.060	1			
Neuroticism (N)	-0.12	0.61	-0.370	-0.230	-0.300	1		
Openness (O)	0.43	0.50	-0.020	-0.060	0.360	-0.090	1	
IQ (proportion correct)	0.86	0.09	0.000	-0.030	-0.160	-0.050	-0.100	1

B. Tanzania

	Mean	Sd	Correlations					
			A	C	E	N	O	IQ
Agreeableness (A)	0.77	0.28	1					
Conscientousness (C)	0.84	0.31	0.460	1				
Extroversion (E)	0.29	0.32	0.150	0.150	1			
Neuroticism (N)	-0.33	0.35	-0.430	-0.340	-0.150	1		
Openness (O)	0.40	0.30	-0.380	0.320	0.450	-0.230	1	
IQ (proportion correct)	0.49	0.22	-0.030	0.030	0.050	-0.040	0.040	1

The personality instrument used is the Big Five Inventory of John et al (1991), and measures were calculated using the “ipsatizing” procedure of John et al (2008). The IQ score is the matrix reasoning part of the WAIS-IV.

Table 3. Summary statistics

A. CCEI scores

		US	Tanzania Low	Tanzania High
Mean		0.950	0.856	0.868
Sd		0.079	0.143	0.150
Percentiles	1	0.584	0.274	0.509
	5	0.803	0.648	0.585
	10	0.866	0.687	0.658
	25	0.933	0.783	0.755
	50	0.986	0.890	0.935
	75	0.999	0.971	0.990
	90	1.000	1.000	1.000
	95	1.000	1.000	1.000
	99	1.000	1.000	1.000
# of obs.		126	106	110

B. A comparison of CCEI scores

		Yale Law School	UCLA	U. of Bergen
Mean		0.982	0.932	0.936
Sd		0.036	0.117	0.135
Percentiles	1	0.858	0.348	0.304
	5	0.889	0.705	0.670
	10	0.925	0.809	0.791
	25	0.990	0.924	0.942
	50	1.000	0.981	0.987
	75	1.000	0.999	1.000
	90	1.000	1.000	1.000
	95	1.000	1.000	1.000
	99	1.000	1.000	1.000
# of obs.		49	121	135

C. FOSD violations

		US	Tanzania Low	Tanzania High
Mean		0.992	0.975	0.978
Sd		0.013	0.029	0.026
Percentiles	1	0.933	0.866	0.903
	5	0.969	0.919	0.920
	10	0.982	0.947	0.942
	25	0.990	0.968	0.966
	50	0.997	0.983	0.989
	75	1.000	0.994	0.996
	90	1.000	0.999	0.999
	95	1.000	1.000	1.000
	99	1.000	1.000	1.000
# of obs.		126	106	110

Table 4. The development gap in CCEI

	(1)	(2)	(3)
Tanzania	-0.087***	-0.095***	-0.098***
	(0.014)	(0.017)	(0.020)
High-stakes		0.015	0.016
		(0.017)	(0.017)
Age			-0.003
			(0.002)
Gender			0.010
			(0.015)
Big Five			
Extraversion			0.008
			(0.016)
Agreeableness			0.021
			(0.021)
Conscientiousness			0.014
			(0.018)
Neuroticism			-0.012
			(0.017)
Openness			-0.007
			(0.019)
Constant	0.950***	0.950***	0.967***
	(0.011)	0.011	(0.058)

Table 5. The development gap in FOSD

	(1)	(2)	(3)
Tanzania	-0.016***	-0.018***	-0.019***
	(0.003)	(0.003)	(0.004)
High-stakes		0.004	0.004
		(0.003)	(0.003)
Age			0.000
			(0.000)
Gender			0.003
			(0.003)
Big Five			
Extraversion			-0.001
			(0.003)
Agreeableness			0.003
			(0.004)
Conscientiousness			0.005
			(0.003)
Neuroticism			-0.001
			(0.003)
Openness			0.001
			(0.004)
Constant	0.992***	0.992***	0.985***
	(0.002)	(0.002)	(0.011)

Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.

Table 6. Estimation results (all subjects)

A. CRRA

		US		Tanzania Low		Tanzania High	
		α	ρ	α	ρ	α	ρ
Mean		1.966	1.070	2.437	1.408	2.560	1.840
Sd		1.910	1.317	2.755	1.515	2.987	1.529
Percentiles	1	1.000	0.000	1.000	0.001	1.000	0.000
	5	1.000	0.058	1.000	0.158	1.000	0.291
	10	1.000	0.108	1.000	0.221	1.000	0.545
	25	1.000	0.270	1.000	0.527	1.000	0.806
	50	1.196	0.624	1.179	0.828	1.086	1.265
	75	1.983	1.254	2.118	1.250	2.065	2.350
	90	3.701	3.151	9.358	5.000	9.506	5.000
	95	6.849	5.000	10.000	5.000	10.000	5.000
	99	10.000	5.000	10.000	5.000	10.000	5.000
% EUT		65.1		75.5		66.4	
# of obs.		126		106		110	

B. CARA

		US		Tanzania Low		Tanzania High	
		α	γ	α	γ	α	γ
Mean		1.688	0.063	1.726	0.116	1.759	0.149
Sd		1.847	0.176	2.006	0.259	2.150	0.288
Percentiles	1	1.000	0.000	1.000	0.000	1.000	0.000
	5	1.000	0.002	1.000	0.004	1.000	0.010
	10	1.000	0.004	1.000	0.007	1.000	0.017
	25	1.000	0.009	1.000	0.016	1.000	0.024
	50	1.000	0.018	1.000	0.025	1.000	0.036
	75	1.298	0.039	1.177	0.052	1.245	0.086
	90	2.858	0.091	2.266	0.374	2.939	0.625
	95	6.683	0.194	5.574	1.000	9.042	1.000
	99	10.000	1.000	10.000	1.000	10.000	1.000
% EUT		81.7		89.6		89.1	
% of obs.		126		106		110	

Table 7. Estimation results (subjects with CCEI ≥ 0.8)

A. CRRA

		US		Tanzania Low		Tanzania High	
		α	ρ	α	ρ	α	ρ
Mean		2.008	1.067	2.771	1.588	2.918	2.034
Sd		1.948	1.344	3.052	1.687	3.319	1.652
Percentiles	1	1.000	0.000	1.000	0.000	1.000	0.000
	5	1.000	0.050	1.000	0.158	1.000	0.337
	10	1.000	0.106	1.000	0.351	1.000	0.594
	25	1.000	0.255	1.000	0.555	1.000	0.863
	50	1.201	0.622	1.192	0.857	1.196	1.286
	75	1.992	1.162	2.512	1.913	2.172	2.936
	90	3.758	3.161	9.444	5.000	10.000	5.000
	95	6.855	5.000	10.000	5.000	10.000	5.000
	99	10.000	5.000	10.000	5.000	10.000	5.000
% EUT		64.2		68.9		61.0	
# of obs.		120		74		77	

B. CARA

		US		Tanzania Low		Tanzania High	
		α	γ	α	γ	α	γ
Mean		1.719	0.064	1.936	0.154	1.903	0.191
Sd		1.887	0.180	2.329	0.302	2.346	0.330
Percentiles	1	1.000	0.000	1.000	0.000	1.000	0.000
	5	1.000	0.002	1.000	0.004	1.000	0.014
	10	1.000	0.004	1.000	0.008	1.000	0.019
	25	1.000	0.008	1.000	0.016	1.000	0.026
	50	1.000	0.018	1.000	0.027	1.000	0.041
	75	1.311	0.035	1.380	0.076	1.262	0.101
	90	2.925	0.097	4.666	0.813	4.115	1.000
	95	6.990	0.213	10.000	1.000	10.000	1.000
	99	10.000	1.000	10.000	1.000	10.000	1.000
% EUT		80.8		87.8		87.0	
% of obs.		120		74		77	

Table 8. CARA regression results

Dependent variable: consistency with EUT

Tanzania	0.057	0.065	0.075	0.066	0.212***	0.062	0.151*	0.075	0.060	0.058	0.091*	0.071	0.055	0.053	0.162*
	(0.038)	(0.042)	(0.049)	(0.047)	(0.082)	(0.052)	(0.085)	(0.050)	(0.053)	(0.046)	(0.054)	(0.049)	(0.052)	(0.051)	(0.087)
High-stakes		-0.015	-0.016	-0.018	-0.012	-0.016	-0.014	-0.016	-0.013	-0.025	-0.019	-0.014	-0.015	-0.024	-0.012
		(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.041)	(0.041)	(0.041)	(0.040)	(0.040)	(0.041)	(0.041)
Age			-0.002	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	-0.004	-0.004	-0.002	-0.002	-0.004	-0.005
			(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female			0.015	0.013	0.012	0.013	0.003	0.015	0.015	0.014	0.019	0.012	0.012	0.012	0.012
			(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.037)	(0.038)	(0.039)	(0.038)	(0.037)
Agreeableness				0.052	0.126**									0.002	0.046
				(0.046)	(0.063)									(0.047)	(0.057)
Agreeableness×Tanzania					-0.208**										-0.129
					(0.088)										(0.105)
Conscientiousness						0.028	0.077							-0.008	0.015
						(0.051)	(0.067)							(0.053)	(0.068)
Conscientiousness×Tanzania							-0.137								-0.058
							(0.103)								(0.119)
Extraversion								0.002	-0.015					0.011	-0.035
								(0.035)	(0.047)					(0.041)	(0.054)
Extraversion×Tanzania									0.058						0.121
									(0.065)						(0.080)
Neuroticism										-0.127***	-0.172***			-0.142***	-0.176***
										(0.038)	(0.051)			(0.043)	(0.058)
Neuroticism×Tanzania											0.126*				0.080
											(0.072)				(0.083)
Openness												-0.107**	-0.123*	-0.137***	-0.126*
												(0.046)	(0.068)	(0.050)	(0.069)
Openness×Tanzania													0.041		0.021
													(0.081)		(0.096)
Constant	0.849***	0.849***	0.881***	0.853***	0.817***	0.868***	0.854***	0.880***	0.881***	0.896***	0.902***	0.936***	0.942***	0.970***	0.942***
	(0.032)	(0.032)	(0.129)	(0.133)	(0.138)	(0.133)	(0.133)	(0.129)	(0.129)	(0.122)	(0.120)	(0.135)	(0.139)	(0.133)	(0.140)
# of obs.	340	340	338	338	338	338	338	338	338	338	338	338	338	338	338

Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.

Table 9. CRRA regression results

Dependent variable: consistency with EUT

Tanzania	-0.001	0.032	0.007	0.010	0.179*	-0.023	0.020	0.019	0.026	-0.005	0.002	-0.001	-0.002	-0.030	0.070
	(0.050)	(0.058)	(0.065)	(0.065)	(0.107)	(0.070)	(0.108)	(0.065)	(0.068)	(0.065)	(0.071)	(0.064)	(0.080)	(0.070)	(0.115)
High-stakes		-0.065	-0.063	-0.063	-0.056	-0.064	-0.063	-0.068	-0.069	-0.070	-0.069	-0.060	-0.060	-0.071	-0.067
		(0.061)	(0.061)	(0.061)	(0.062)	(0.061)	(0.061)	(0.061)	(0.061)	(0.062)	(0.062)	(0.061)	(0.061)	(0.061)	(0.062)
Age			0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.001	0.000	0.001	0.001	-0.000	-0.000
			(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Female			-0.045	-0.044	-0.045	-0.049	-0.054	-0.048	-0.048	-0.045	-0.044	-0.051	-0.051	-0.054	-0.051
			(0.053)	(0.053)	(0.052)	(0.053)	(0.054)	(0.053)	(0.053)	(0.053)	(0.053)	(0.052)	(0.052)	(0.052)	(0.053)
Agreeableness				-0.014	0.072									-0.063	0.011
				(0.061)	(0.078)									(0.062)	(0.071)
Agreeableness×Tanzania					-0.242*										-0.242
					(0.127)										(0.147)
Conscientiousness						0.066	0.090							0.060	0.047
						(0.060)	(0.076)							(0.059)	(0.073)
Conscientiousness×Tanzania							-0.067								0.076
							(0.126)								(0.136)
Extraversion								-0.084*	-0.076					-0.046	-0.056
								(0.047)	(0.056)					(0.058)	(0.071)
Extraversion×Tanzania									-0.026						0.022
									(0.104)						(0.123)
Neuroticism										-0.093*	-0.103			-0.136**	-0.124
										(0.054)	(0.070)			(0.061)	(0.078)
Neuroticism×Tanzania											0.028				-0.038
											(0.111)				(0.127)
Openness												-0.220***	-0.221***	-0.219***	-0.206**
												(0.063)	(0.079)	(0.071)	(0.088)
Openness×Tanzania													0.003		0.003
													(0.132)		(0.154)
Constant	0.730***	0.730***	0.726***	0.734***	0.693***	0.698***	0.690***	0.732***	0.731***	0.738***	0.739***	0.839***	0.839***	0.866***	0.826***
	(0.040)	(0.040)	(0.156)	(0.161)	(0.165)	(0.160)	(0.160)	(0.157)	(0.158)	(0.154)	(0.154)	(0.162)	(0.162)	(0.164)	(0.170)
# of obs.	340	340	338	338	338	338	338	338	338	338	338	338	338	338	338

Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.

Table 10. The overlap of the CCEI and IQ scores distributions

		Fraction of subjects			
	(Joint) Centile		US	Tanzania	
				Low	High
CCEI	25	0.827	0.063	0.390	0.330
	50	0.950	0.341	0.657	0.532
	75	0.995	0.635	0.838	0.798
FOSD	25	0.977	0.087	0.362	0.330
	50	0.991	0.278	0.714	0.550
	75	0.998	0.556	0.895	0.835
IQ	25	0.462	0.000	0.410	0.431
	50	0.692	0.079	0.800	0.853
	75	0.846	0.468	0.981	0.991
# of obs.		342	126	106	110

The fractions of Tanzanian and US subjects whose scores are below the (joint) centile.

Table 11. The development gap in IQ

	(1)	(2)	(3)
Tanzania	-0.367***	-0.355***	-0.326***
	(0.020)	(0.024)	(0.029)
High-stakes		-0.024	-0.027
		(0.025)	(0.025)
Age			-0.003
			(0.003)
Gender			0.041
			(0.021)
Big Five			
Extraversion			-0.010
			(0.024)
Agreeableness			-0.032
			(0.030)
Conscientiousness			0.001
			(0.025)
Neuroticism			-0.028
			0.024
Openness			-0.003
			0.028
Constant	0.856***	0.856***	0.903***
	(0.016)	(0.016)	(0.075)

Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.