NHH

INSTITUTT FOR SAMFUNNSØKONOMI

DEPARTMENT OF ECONOMICS

SAM 8 2014

ISSN: 0804-6824

March 2014

Discussion paper

# Is There a Development Gap in Rationality?

BY

**Alexander W. Cappelen, Shachar Kariv, Erik Ø. Sørensen,**
AND **Bertil Tungodden**



Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

# Is There a Development Gap in Rationality?*

Alexander W. Cappelen, Shachar Kariv,

Erik Ø. Sørensen, and Bertil Tungodden†

January 28, 2014

**Abstract**

We report an experimental test of the four touchstones of rationality in choice under risk – utility maximization, stochastic dominance, expected-utility maximization and small-stakes risk neutrality – with students from one of the best universities in the United States and one of the best universities in Africa, the University of Dar es Salaam. Although the US and the Tanzanian subjects come from different backgrounds and face different economic prospects, they are united by being among the most able in their societies. Importantly, many of whom will exercise an outsized influence over economic and political affairs. We find very small or no significant differences between the two samples in the degree of rationality according to a number of standard economic measures. An alternative approach is to take cognitive ability (IQ) as a proxy for economic rationality. We show that a canonical IQ test indicates a much larger development gap in rationality relative to our economic tests.

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

**Keywords**: Development, rationality, revealed preference, stochastic dominance, expected utility, risk aversion, cognitive ability, experiment.

*We thank Siv-Elisabeth Skjelbred for excellent research assistance. We also thank Ray Fisman, Pam Jakiela and Ted 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 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 new 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 many backgrounds, 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. If our subjects do not fit the model of Homoeconomicus, then others in their society are therefore even less likely to be rational utility maximizers.

These subject pools are also worth studying for *extrinsic* reasons. Because the students of UC Berkeley and the University of Dar es Salaam are among the most able in their societies, they will presumably assume positions of power in different sectors of their economy when they graduate into the world and can thus spur growth and development. This is especially true for the Tanzania students in our sample, of whom many will assume positions of substantial power in national economic and political affairs – the current President of Tanzania, Jakaya Kikwete, the Prime Minster, Mizengo Pinda, and his predecessor, Edward Lowassa, all graduated from the University of Dar es Salaam. Acemoglu and Robinson (2012) argue that "Poor countries are poor not because those who have power make choices that create poverty. They get it wrong not by mistake or ignorance but on purpose." The argument stipulates that those who have power must be rational. This highlights the importance of rigorously testing the

[1] Webometrics (`www.webometrics.info`) ranks all universities worldwide based on the impact of the university on the web. It ranks 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$^{\text{th}}$ in Africa.

rationality of the future leadership in developing countries.[2]

This is especially true for the Tanzania students in our sample, of whom many will assume positions of substantial power in national economic and political affairs. Acemoglu and Robinson (2012) argue that "Poor countries are poor not because those who have power make choices that create poverty. They get it wrong not by mistake or ignorance but on purpose." This theory stipulates that those who have power must be rational. This highlights the importance of developing a rigorous test of the rationality of students in the upper echelons of institutions in their society.[3]

We focus on choice under risk, because uncertainty is endemic in a wide variety of economic circumstances and enters every realm of economic decision-making. We provide a thorough experimental test of the four touchstones of rationality in decision making under risk: utility maximization, stochastic dominance, expected-utility maximization and small-stakes risk neutrality. Inconsistencies with revealed preference conditions and violations of first-order 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), so violations of expected utility and risk aversion over modest stakes lack normatively appealing properties of rationality. Clearly, choices cannot be considered rational if there is no utility function that those choices maximize. However, choices can be rational yet fail to be reconciled with any utility function that is normatively appealing given the decision problem at hand.

**The experiment.** In the experiment, we present subjects with a standard economic decision problem that can be interpreted either as a portfolio choice problem (the allocation of income between two risky assets) or a consumer decision problem (the selection of a bundle of contingent commodities from a standard budget line). Put pre-

[2]There is credible evidence that quality education has a strong causal relationship to 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]There is credible evidence that quality education has a strong causal relationship to 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).

cisely, 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. Let $\mathbf{x} = (x_1, x_2)$ denote an allocation of securities, where $x_s$ denotes the number of units of security $s$. Without essential loss of generality, assume the individual's payout is normalized to 1. The budget constraint is then $B(\mathbf{p})$, where $\mathbf{p} = (p_1, p_2)$ is the vector of security prices and $p_s$ denotes the price of security $s$. The individual can choose any allocation $\mathbf{x}$ that satisfies this constraint.

To characterize an individual's decision-making under uncertainty, it is necessary to generate many observations per subject over a wide range of budget lines. A graphical interface was developed for this purpose by Choi et al. (2007b), where subjects see a geometrical representation of the budget line on a computer screen and choose the portfolio through a simple 'point-and-click.' This intuitive and user-friendly interface allows for the quick and efficient elicitation of many decisions per subject under a wide range of budget lines. In particular, the changes in endowments and relative prices are such that budget lines cross frequently. The broad range of budget lines faced by each subject allows us to thoroughly test the rationality of decision-making under risk.

**Measures of rationality.** We study the four rationality postulates underlying the theory of choice under risk:

- **Utility maximization.** The most basic question to ask about choice data is whether the data 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.

- **First-order stochastic dominance.** 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). The dominance principle is compelling and generally accepted in decision theory. Notice that any decision to allocate *fewer* tokens to the *cheaper* security violates FOSD but need not involve a violation of GARP. Choices can be consistent with GARP and yet fail to be reconciled with any utility function that is appropriate to the decision problem at hand.

- **Expected-utility maximization.** If preferences satisfy the Savage (1954) axioms, then the preference ordering over allocations $\mathbf{x}$ can be represented by a function of the form $U(\mathbf{x}) = \frac{1}{2}u(x_1) + \frac{1}{2}u(x_2)$ where $u(\cdot)$ is the utility function over money. Such an individual will choose an allocation $\mathbf{x}^*$ to maximize the expected value of utility subject to the budget constraint $\mathbf{x} \in B(\mathbf{p})$. To test EUT, we estimate at the individual level preferences generated by the disappointment aversion model of Gul (1991). The model embeds EUT as a parsimonious and tractable special case.

- **Small-stakes risk neutrality.** Arrow (1971) showed that expected-utility maximizers are arbitrarily close to risk neutral when stakes are arbitrarily small. Rabin (2000) presented an empirical claim and a (calibration) theorem which together show that the risk-neutrality implication of EUT is not restricted to arbitrarily small stakes. In our experiment, the test of risk neutrality is straightforward, since it requires always allocating all tokens to the cheaper security so that the preference ordering over allocations $\mathbf{x}$ can *only* be represented by a function of the form $U(\mathbf{x}) = \frac{1}{2}x_1 + \frac{1}{2}x_2$.

**Summary of results.** We assess how the data comply with GARP by calculating Afriat's (1972) Critical Cost Efficiency Index (CCEI). This measures the amount by which each budget constraint must be relaxed in order to remove all violations of GARP. The CCEI 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. In our experiment, mean CCEIs across all subjects are 0.950 in the US and 0.856 and 0.869 with low and high stakes in Tanzania, respectively.[4,5] More importantly, we find that the US subjects display greater levels of consistency than the Tanzanian subjects. The magnitudes imply that the Tanzanian

[4]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.

[5]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).

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, we argue that most of our subjects are close enough to passing GARP so that we may not want to reject that their choices are consistent with utility maximization. This is, to our knowledge, the first quantifiable and economically interpretable measure of the development gap in rationality. Figure 1 below summarizes the cumulative distributions of CCEI scores.

*[Figure 1 here]*

The rest of our results are summarized as follows. We use expected payoff calculations to assess how closely individual choice behaviors comply with FOSD. Overall, 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. As a practical note, the consistency with FOSD also suggests that subjects did not have any difficulties in understanding the procedures or using the computer program. 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. Finally, perhaps as expected, the behavior of only very few subjects in both samples is consistent with risk neutrality (over small stakes). Overall, the results on FOSD violations, consistency with EUT and small-stakes risk neutrality indicate no development gap in economic rationality.

**Economic rationality versus cognitive ability.** An alternative approach in behavioral economics is to proxy economic rationality with one of the many tests of cognitive ability (IQ), which indeed seem to capture aspects of the ability to make rational economic decisions (Dohmen et al., 2010). Clearly, 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 correlation 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, in some circumstances, the conceptual distinctions between the measures would have little practical import. To this end, 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 graphical illustration of the extent to which using IQ as a proxy for economic rationality will inflate the development gap. In sharp contrast, although individual behaviors are complex and heterogeneous, we find a high degree of rationality in the sense that most subjects' choices come close to satisfying rationality according to a number of standard economic measures.

*[Figure 2 here]*

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.[6]

# 2 The experiment

We conducted laboratory experiments at UC Berkeley and the University of Dar es Salaam in Tanzania. 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

[6]The paper also uses several data and technical online appendices for the interested reader (http://emlab.berkeley.edu/~kariv/CKST_I_A#.pdf).

English is the language of instruction at the University of Dar es Salaam). The experiment began by the subjects completing a non-incentivized WAIS-IV test, which is the most frequently administered IQ test, and a standard Big Five personality traits questionnaire.[7]

Appendix II provides information on the subject populations. It also presents the results of their IQ test and Big Five personality test. For most attributes, there is no significant difference in the sample composition between the two subject pools.[8] The two exceptions are that the US pool has a much higher fraction of females (0.706 versus 0.336, $p < 0.001$) and younger subjects (20.6 versus 23.3, $p < 0.001$). All our results are robust to the inclusion of controls for gender and age.

The experimental procedures described below are identical to those used by Choi et al. (2007a). Each experimental session consisted of 50 independent decision problems. In each decision problem, 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. 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 his 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.

[7]We use the matrix progression part of the WAIS-IV and 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. Barrick and Mount (1991) conclude that conscientiousness is the best predictor of economic outcomes. Subjects were asked to evaluate the accuracy of statements as descriptions of themselves on a five point scale. The personality scores are calculated using the procedure of John et al. (2008).

[8]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. Overall, the UC Berkeley students come from diverse socioeconomic backgrounds. Notably, 37 percent of UC Berkeley undergraduates received Pell Grants in 2011-12 (awards of up to $5,550 per academic year for undergraduate students with family incomes generally less than $45,000 a year).

Choices were restricted to allocations on the budget constraint.[9,10]

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. 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 that 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.

An example of a budget line defined in this way 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 point that lies in the middle 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

[9] It is of course possible that presenting choice problems graphically biases choice behavior in some particular way – and that is a useful topic for research – but there is 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. Among others, Choi et al. (2007b) employ a similar platform to study distributional preferences and produce very different behaviors. Building on our experimental methods, Choi et al. (2011) investigate the correlation between consistency with GARP and demographic and economic characteristics in the CentERpanel (a representative sample of more than 2,000 Dutch households). Since all experimental designs share the same graphical interface, we are building on the data sets and expertise we have acquired in previous work.

[10] 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.

function.

*[Figure 3 here]*

# 3 Experimental results

In this section, we first discuss the revealed preference tests used to determine whether individual choices in our experiment are consistent with utility maximization. Beyond consistency, it is natural to ask whether choices are also consistent with FOSD. Finally, we also ask whether choices can be reconciled with a utility function with some normatively appealing properties, namely EUT and small-stakes risk neutrality.

## 3.1 Utility maximization

The most basic measure of economic rationality is whether the individual choices are consistent with individual utility maximization (Samuelson, 1947). If budget sets are linear (as in our experiment), classical revealed preference theory provides a direct test of consistency: choices in a finite collection of budget lines are consistent with maximizing a well-behaved (that is, piecewise linear, continuous, increasing, and concave) utility function if and only if they satisfy GARP. Conversely, Afriat's (1967) theorem tells us that if a *finite* data set generated by an individual's choices satisfies GARP, then the data can be rationalized by a well-behaved utility function.[11] Hence, in order to decide whether the choices are consistent with utility-maximizing behavior we only need to check whether they satisfy GARP.

Let $\left\{(p^i,x^i)\right\}_{i=1}^{50}$ be the data generated by an individual's choices in our experiment, where $p^i$ denotes the $i$-th observation of the price vector and $x^i$ denotes the associated allocation. In this context GARP requires that if $x^i$ is indirectly revealed preferred to $x^j$, then $x^j$ is not *strictly* directly revealed preferred to $x^i$. The broad range of budget lines faced by each participant in the present experiment provides a rigorous test of GARP, where the relative prices are such that budget lines cross frequently. This means that our data lead to high power tests of the revealed preference conditions.

[11] 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.

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. We therefore need to measure how closely individual choice behavior complies with consistency, where we rely on the Critical Cost Efficiency Index (CCEI) proposed by Afriat (1972). The CCEI measures the fraction by which each budget constraint must be shifted in order to remove all violations of GARP.

By definition, the CCEI 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. 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 the income by making an inconsistent choice. The CCEI score thus overstate the extent of inefficiency.

*[Figure 4 here]*

Although the CCEI provides a summary statistic of the overall consistency of the data with GARP, it does not give any information about which of the observations are causing the most severe violations. A single large violation may lead to a small value of the index while a large number of small violations may result in a much larger efficiency index. Alternative measures have been suggested by Houtman and Maks (1985) and Varian (1990). In practice, all these measures yield similar conclusions. The various indices are all computationally intensive for even moderately large data sets.[12]

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

[12]We refer the interested reader to Choi et al. (2007b,a) 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.

subjects have CCEIs above 0.90, while less than 0.5 percent of simulated subjects have CCEIs 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.[13]

Complementing the graphical presentation in Figure 1, Table 1A below reports summary statistics and percentile values for the CCEI.[14] Of the 126 US subjects, 108 (85.7 percent) have CCEI scores above 0.9, and of those, 83 subjects (65.9 percent) had 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 1 also show that the distribution of CCEI scores is skewed to the left 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 1B reports summary statistics and percentile values for the CCEI for these subject pools.[15]

[13]The power of the experiment 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. Choi et al. (2007b) generate a random sample of simulated subjects who implement the power utility function $u(x) = x^{1-\rho}/(1-\rho)$, commonly employed in the empirical analysis of choice under uncertainty, with error. The likelihood of error is assumed to be a decreasing function of the utility cost of an error. More precisely, they assume an idiosyncratic preference shock that has a logistic distribution where the logistic precision parameter reflects sensitivity to differences in utility. The analysis demonstrates that if utility maximization is not in fact the correct model, then our experiment is sufficiently powerful to detect it.

[14]We need to allow for small mistakes resulting from the slight imprecision of subjects' handling of the mouse. All the GARP results presented below allow for a narrow confidence interval of one token (for any $i$ and $j \neq i$, if $d(x^i, x^j) \leq 1$ then $x^i$ and $x^j$ are treated as the same allocation).

[15]The experiment of Choi et al. (2011) consisted of 25, rather than 50, decision problems so the CCEI scores are not directly comparable. Choi et al. (2011) combine the actual data from the experiment and the mirror-image of these data obtained by reversing the prices and the associated allocation for

*[Table 1 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 the high-stakes experimental treatment. We also include age, gender and the Big Five personality traits as controls. The dependent variable is the subject's CCEI score. Table 2 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 demographic profiles or by dimensions that are used to describe human personality in Psychology. 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 dependent variable at zero and 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 2 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. The dominance principle is compelling and generally accepted in decision theory. All typical 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. 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

each observation. Choi et al. (2011) 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.

to an 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 had 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 "$\alpha$ points" is shifted up to "$\gamma$ points" and the outcome "$\beta$ points" in 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 1C above reports summary statistics and percentile values.[16]

Over all subjects, the FOSD scores averaged 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

[16]Choi et al. (2011) 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. (2011) 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.

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 thus show very low rates of FOSD violations. Nevertheless, there is also some heterogeneity in the FOSD scores within and across subject pools. Table 3 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 3 here]*

### 3.3 Expected-utility maximization

The empirical validity of the Savage axioms, on which EUT is based, has been tested extensively by experimentalists. We now turn to the next level of analysis involving estimation of individual-level, parametric utility functions. Choi et al. (2007a) find that, for some subjects, the choice data are well explained by a preference ordering in which the indifference curves have a kink at the 45 degree line, which corresponds to an allocation with a certain payoff. For EUT, in contrast, the indifference curves are smooth everywhere. Figure 6 below illustrates a "kinked" indifference curve. One interpretation of this preference ordering is given by the disappointment aversion 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.[17] This formulation thus embeds EUT as a parsimonious and tractable special case. 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.[18]

*[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.[19] 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.

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, in-

[17]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 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 $s'$.

[18]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.

[19]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 $\omega$, 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.

creasing, 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 $\gamma \leq 1$ and $\rho \leq 5$. This involves minimal loss in fit, since the predicted choices with such a high level of risk aversion and/or disappointment aversion are virtually identical.

- Finally, subjects with low CCEI scores are not sufficiently consistent to be considered utility-generated. We therefore present results for all subjects and for those with CCEI scores above 0.8 in parallel throughout the remainder of the paper.

To economize on space, the individual-level estimates are relegated to Appendix III. Appendix III 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 4 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 5. 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, a much higher proportion than found in other experimental studies.[20,21] There is also considerable heterogeneity in subjects' risk preferences in both the disappointment-averse and disappointment-neutral subsamples but they are within the range of recent estimates of risk aversion (more below).

*[Table 4 here]*

*[Table 5 here]*

[20]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.

[21]Ahn et al. (2013) report an experiment where subjects make choices over three-dimensional budget sets corresponding to three states of nature, where one state is risky with known probability and two states are ambiguous with unknown probabilities. Ahn et al. (2013) did not reject the null hypothesis of preferences consistent with Subjective Expected Utility (Savage, 1954) for 64.3 percent of their subjects.

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 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 $\hat{\alpha} = 0$ 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 there is considerable heterogeneity in both parameters in both the CARA and CRRA specifications. Nevertheless, there are no significant differences in the fraction of subjects consistent with EUT between the two samples. We omit the regression analysis for the sake of brevity.

## 3.4 Small-stakes risk neutrality

Except for the high-stakes treatment in Tanzania, not allocating all tokens to the cheaper security is hard to reconcile with a standard (Savage-based) theory of choice under risk regarding wealth. Any such small-stakes risk aversion is often called loss aversion. Arrow (1971) showed that expected-utility maximizers are arbitrarily close to risk neutral when stakes are arbitrarily small, and Rabin (2000) showed that the risk-neutrality implication of EUT is not restricted to arbitrarily small stakes. Because the experimental decision problem is symmetric (the two states are equally likely) and budget lines are drawn from a symmetric distribution, we can summarize the risk attitudes of our subjects by reporting the average fraction of tokens allocated to the cheaper security. Always allocating all tokens to the cheaper security corresponds to risk neutrality (and equally dividing all tokens corresponds to infinite risk aversion).

Like the revealed preference tests, the advantage of this measure is that it is nonparametric. It summarizes attitudes toward risk in a single number, making no assumptions about the parametric form of the underlying utility function. Further, because we observe many choices over a wide range of budget lines, it describes preferences with considerable precision. Because the purpose of the risk elicitation is to test for consistency with small-stakes risk neutrality, we favor this "low-tech" nonparametric approach, which provides a good fit and is conceptually straightforward, over the structural estimation above. Table 1D above reports summary statistics and percentile values for the full sample. The distributions are similar for the subsample of subjects with CCEI scores above 0.80. Of all our subjects, only one US subject and one Tanzanian subject (in the low-stakes treatment) allocated all their tokens to $x_1$ if $p_1 < p_2$ and to $x_2$ if $p_1 > p_2$ implying risk-neutral preferences. In addition, the hypothesis that preferences are consistent with risk neutrality is rejected for all other subjects with the CRRA ($\hat{\alpha} = 1$ and $\hat{\rho} = 0$) and CARA ($\hat{\alpha} = 1$ and $\hat{\gamma} = 0$) specifications.

# 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)).[22] Figure 1 above summarizes the cumulative distributions of CCEI scores and Figure 2 summarizes the cumulative distributions of IQ scores in both samples. Table 6 summarizes the $25^{\text{th}}$, $50^{\text{th}}$ and $75^{\text{th}}$ 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 are 63.0 and 27.8 percent, respectively. Finally, in Table 7 we repeat the OLS estimation reported in Tables 2 and 3 using the subjects' IQ scores instead of the CCEI or FOSD scores as the dependent variable.

*[Table 6 here]*
*[Table 7 here]*

We note that the four measures of economic rationality tested in this paper – utility maximization, stochastic dominance, expected-utility maximization and small-stakes

[22] 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.

risk neutrality – offer a theoretically disciplined metric for quality of economic decisions. These measures have 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. Overall, we view our methods as complementing those based on standard experimental methods, psychological factors, and cognitive tasks, and future work will investigate the relationship between our economic approach, performance on these other tasks, and important behaviors in the real world.

# 5 Conclusion

We tested the four 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.

Our study has important practical implications in political economy. This is especially true for the Tanzanian students in our sample who will likely exercise an outsized influence and assume positions of substantial power in national and indeed global economic and political affairs. Their high level of economic rationality supports the theory of Acemoglu and Robinson (2012) that policymakers in poor nations "get it wrong not by mistake or ignorance but on purpose." Finally, our results also suggest a number of potential directions for future work. Policymakers make choice (over lotteries) with outcomes that have both personal and social consequences. Students of elite universities thus graduate into a world that allows them opportunities to implement their social preferences. Our experimental technique allows for the collection of richer data about preferences than has heretofore been possible, and can provide a positive account of preferences for both personal and social consumption in rich choice environments (cf. Fisman et al. (2007)).

# 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*.

Arrow, Kenneth J. (1971). *Essays in the Theory of Risk-Bearing*, Chicago: Markham Publishing.

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.

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*.

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 (2011). "Who is (more) rational?" NBER Working Paper No. 16791.

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.

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

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.

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

Houtman, Martijn and J. A. H. Maks (1985). "Determining all maximal data subsets consistent with revealed preference," *Kwantitatieve Methoden*, 6(19): 89–104.

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.

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.

Rabin, Matthew (2000). "Risk aversion and expected-utility theory: A calibration theorem," *Econometrica*, 68(5): 1281–1292.

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

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

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

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

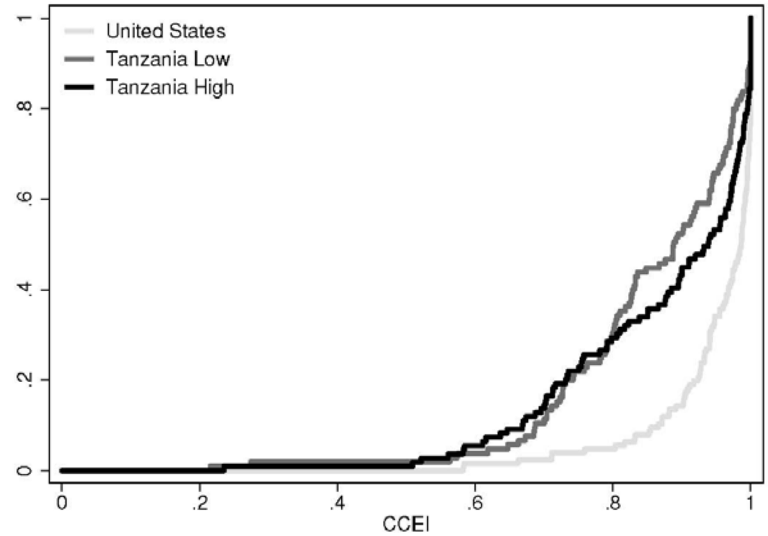



Figure 2. The cumulative distributions of IQ scores

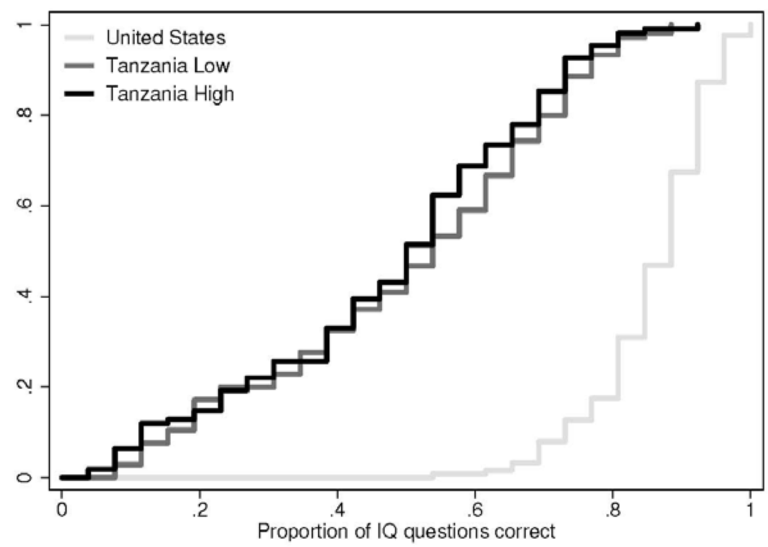
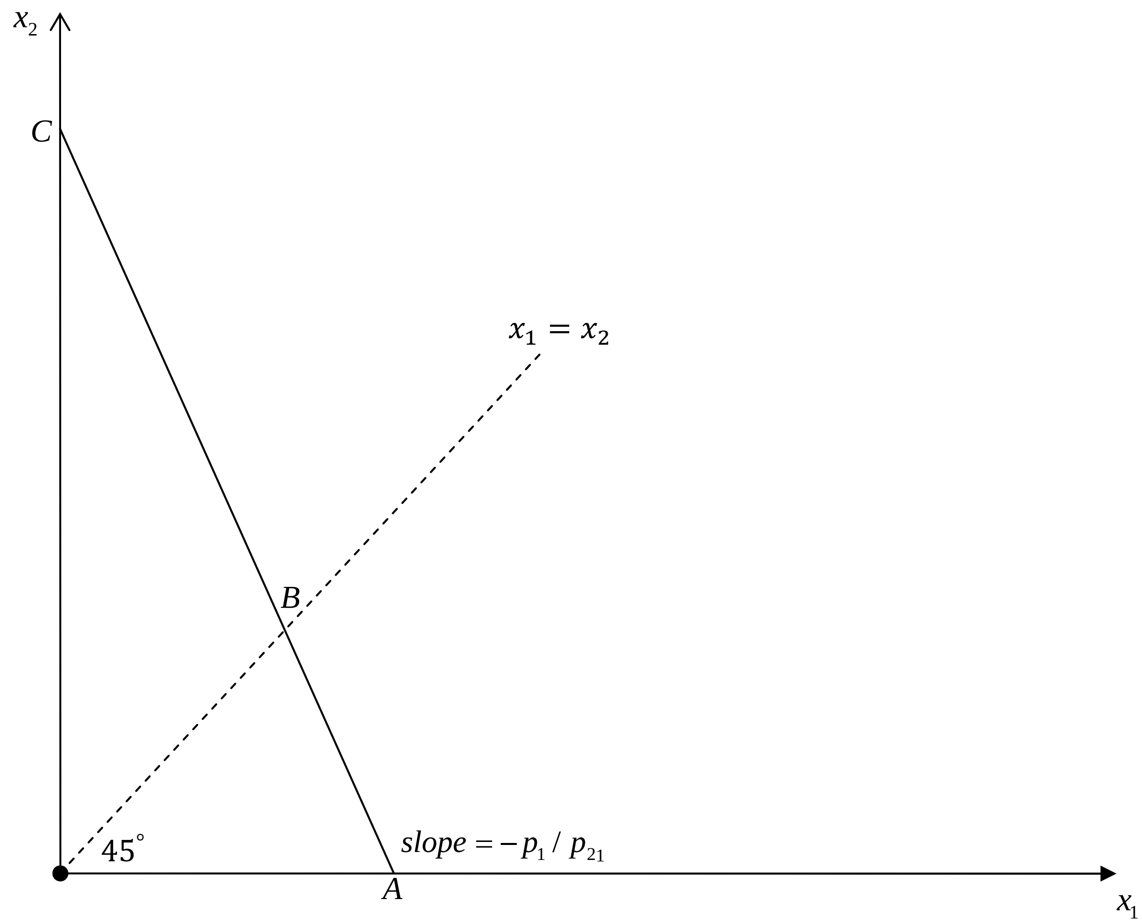
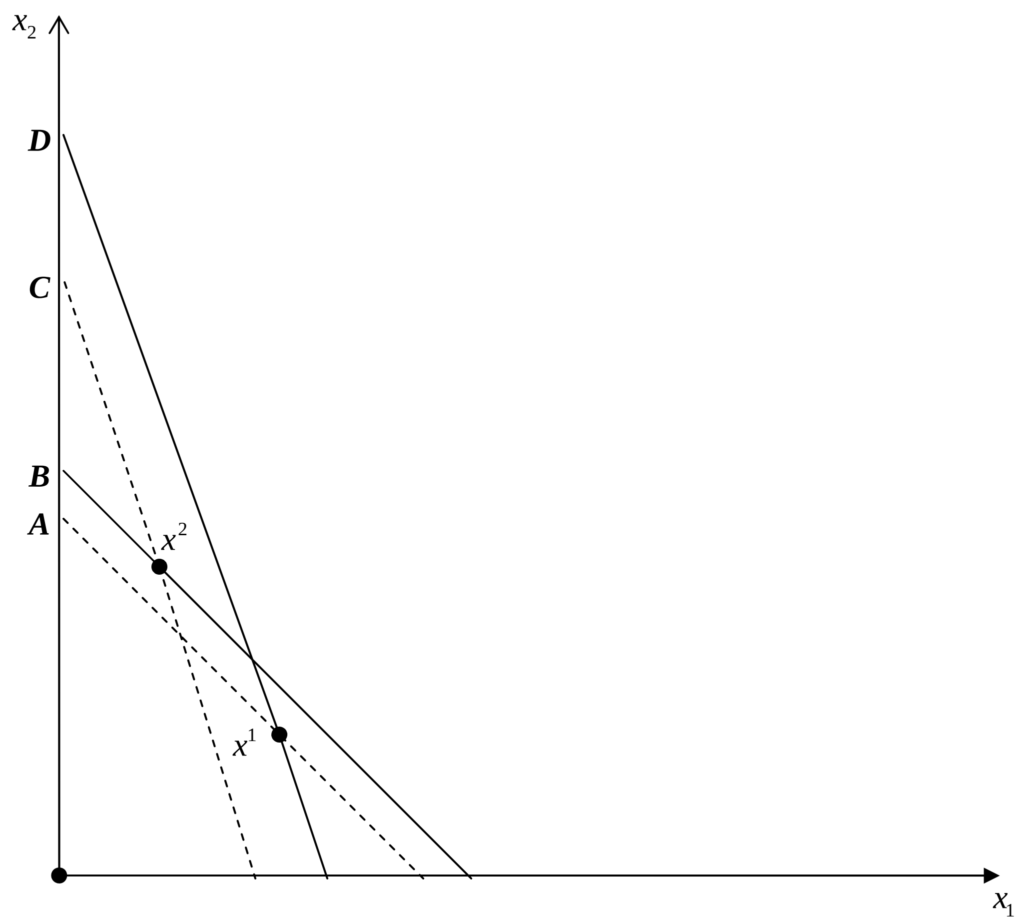


Figure 3: An example of a budget line



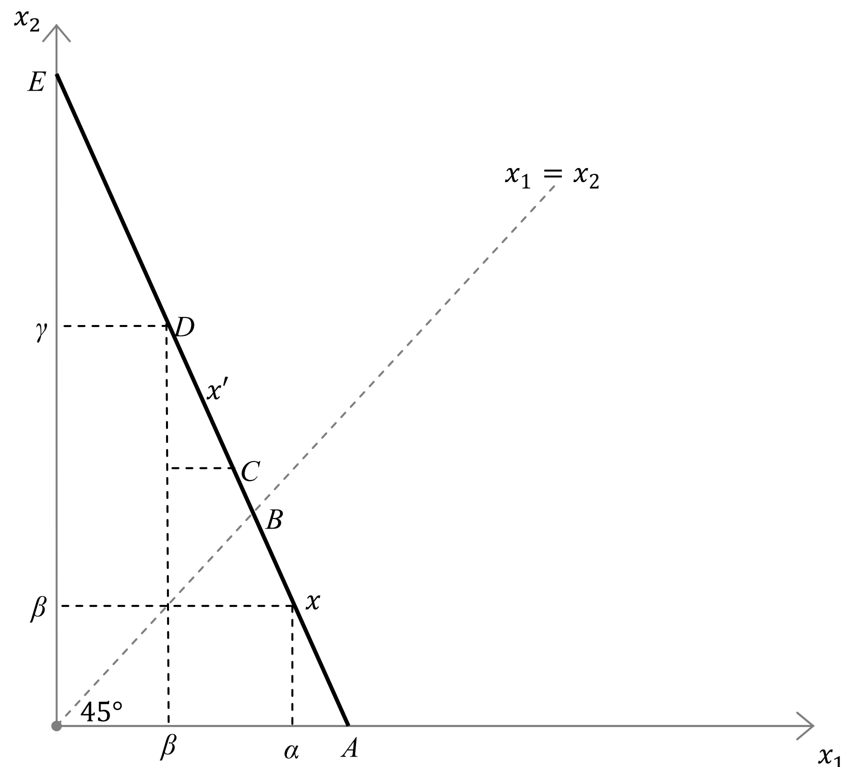

An example of a budget line in the experiment. The point *B*, which lies on the 45 degree line, corresponds to the safe allocation with a certain payoff. This allocation is consistent with infinite risk aversion. By contrast, point *C* represents an allocation in which all tokens are allocated in the 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



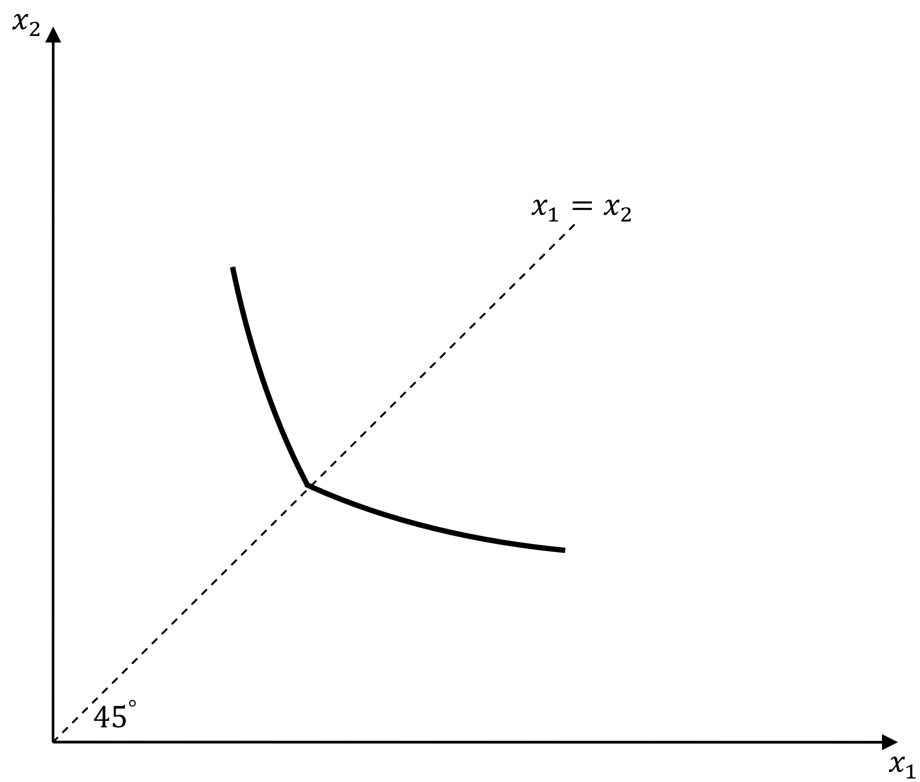

Here we have a violation of the Weak Axiom of Revealed Preference (WARP) since $x^1$ is directly revealed preferred to $x^2$ and $x^2$ is directly revealed preferred to $x^1$. A perturbation $A/B > C/D$ of the budget line through allocation $x^1$ removes the violation. The CCEI can thus be interpreted 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 ($\alpha$).The shape of the indifference curve on either side of the 45 degree line is determined by the individual's attitude toward risk ($\rho$ in the CRRA specification and $\gamma$ 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 ($\alpha$ = 1), in contrast, the indifference curves are smooth everywhere.

Table 1. 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 |

D. Risk aversion

| | | US | Tanzania Low | Tanzania High |
|---|---|---|---|---|
| Mean | | 0.643 | 0.617 | 0.592 |
| Sd | | 0.100 | 0.098 | 0.080 |
| Percentiles | 1 | 0.500 | 0.486 | 0.498 |
| | 5 | 0.511 | 0.500 | 0.500 |
| | 10 | 0.536 | 0.503 | 0.502 |
| | 25 | 0.569 | 0.548 | 0.535 |
| | 50 | 0.625 | 0.610 | 0.576 |
| | 75 | 0.713 | 0.665 | 0.630 |
| | 90 | 0.771 | 0.749 | 0.677 |
| | 95 | 0.830 | 0.782 | 0.703 |
| | 99 | 0.975 | 0.918 | 0.926 |
| # of obs. | | 126 | 106 | 110 |

Table 2. 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 3. 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 4. Estimation results (all subjects)

A. CRRA

| | | US | | Tanzania | | | |
|---|---|---|---|---|---|---|---|
| | | | | Low | | High | |
| | | $\alpha$ | $\rho$ | $\alpha$ | $\rho$ | $\alpha$ | $\rho$ |
| 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 | | High | |
| | | $\alpha$ | $\gamma$ | $\alpha$ | $\gamma$ | $\alpha$ | $\gamma$ |
| 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 5. Estimation results (subjects with CCEI ≥ 0.8)

A. CRRA

| | | US | | Tanzania | | | |
|---|---|---|---|---|---|---|---|
| | | | | Low | | High | |
| | | $\alpha$ | $\rho$ | $\alpha$ | $\rho$ | $\alpha$ | $\rho$ |
| 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 | | High | |
| | | $\alpha$ | $\gamma$ | $\alpha$ | $\gamma$ | $\alpha$ | $\gamma$ |
| 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 6. The overlap of the CCEI and IQ scores distributions

| | | (Joint) Centile | US | Tanzania Low | Tanzania 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 |

Fraction of subjects: US, Tanzania (Low, High).

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

Table 7. 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.

## Issued in the series Discussion Papers 2013

**2013**

**01/13** January, **Lukáš Laffërs**, "Identification in Models with Discrete Variables".

**02/13** January, **Ingvild Almås**, Anders Kjelsrud and Rohini Somanathan, "A Behaviour-based Approach to the Estimation of Poverty in India".

**03/13** February, **Ragnhild Balsvik** and Line Tøndel Skaldebø, "Guided through the `Red tape'? Information sharing and foreign direct investment".

**04/13** February, **Sissel Jensen**, Ola Kvaløy, Trond E. Olsen, and **Lars Sørgard**, "Crime and punishment: When tougher antitrust enforcement leads to higher overcharge".

**05/13** February, **Alexander W. Cappelen**, **Trond Halvorsen**, **Erik Ø. Sørensen**, and **Bertil Tungodden**, "Face-saving or fair-minded: What motivates moral behavior?"

**06/13** March, **Jan Tore Klovland** and **Lars Fredrik Øksendal**, "The decentralised central bank: regional bank rate autonomy in Norway, 1850-1892".

**07/13** March, **Kurt Richard Brekke**, Dag Morten Dalen, and Tor Helge Holmås, "Diffusion of Pharmaceuticals: Cross-Country Evidence of Anti-TNF drugs".

**08/13** April, **Kurt R. Brekke**, Luigi Siciliani, and Odd Rune Straume, "Hospital Mergers:A Spatial Competition Approach".

**09/13** April, **Liam Brunt** and Edmund Cannon, "The truth, the whole truth, and nothing but the truth: the English Corn Returns as a data source in economic history, 1770-1914".

**10/13** April, **Alexander W. Cappelen**, **Bjørn-Atle Reme**, **Erik Ø. Sørensen**, and **Bertil Tungodden,** "Leadership and incentives".

**11/13** April, Erling Barth, **Alexander W. Cappelen**, and Tone Ognedal, "Fair Tax Evasion".

**12/13** June, **Liam Brunt** and Edmund Cannon, "Integration in the English wheat market 1770-1820".

**13/13** June, **Fred Schroyen** and **Nicolas Treich**, "The Power of Money: Wealth Effects in Contests".

**14/13** August, **Tunç Durmaz** and **Fred Schroyen**, “Evaluating Carbon Capture and Storage in a Climate Model with Directed Technical Change”.

**15/13** September, **Agnar Sandmo**, “The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought”.

**16/13** October, **Kai Liu**, “Health Insurance Coverage for Low-income Households: Consumption Smoothing and Investment”.

**17/13** December, **Øivind A. Nilsen**, **Lars Sørgard**, and **Simen A. Ulsaker**, “Upstream Merger in a Successive Oligopoly: Who Pays the Price?”

**18/13** December, Erling Steigum and **Øystein Thøgersen**, “A crisis not wasted – Institutional and structural reforms behind Norway’s strong macroeconomic performance”.

**2014**

**01/14** January**, Kurt R. Brekke**, Tor Helge Holmås, and Odd Rune Straume, "Price Regulation and Parallel Imports of Pharmaceuticals".

**02/14** January, **Alexander W. Cappelen, Bjørn-Atle Reme, Erik Ø. Sørensen,** and **Bertil Tungodden, "**Leadership and incentives".

**03/14** January, **Ingvild Almås, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø**. **Sørensen**, and **Bertil Tungodden**, "Willingness to Compete: Family Matters".

**04/14** February, **Kurt R. Brekke**, Luigi Siciliani, and Odd Runde Straume, "Horizontal Mergers and Product Quality".

**05/14** March, **Jan Tore Klovland**, "Challenges for the construction of historical price indices: The case of Norway, 1777-1920".

**06/14** March, Johanna Möllerström, **Bjørn-Atle Reme**, and **Erik Ø. Sørensen**, "Luck, Choice and Responsibility".

**07/14** March, Andreea Cosnita-Langlais and **Lars Sørgard**, "Enforcement vs Deterrence in Merger Control: Can Remedies Lead to Lower Welfare?"

**08/14** March, **Alexander W. Cappelen, Shachar Kariv, Erik Ø. Sørensen**, and **Bertil Tungodden**, «Is There a Development Gap in Rationality?"

**Norges Handelshøyskole**

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no