The Development Gap in Economic Rationality of Future Elites

Alexander W. Cappelen, Shachar Kariv
Erik Ø. Sørensen, and Bertil Tungodden†

October 19, 2020

Abstract

We test the touchstones of economic rationality—utility maximization, stochastic dominance, and expected-utility maximization—of elite students in the U.S. and in Africa. The choices of most students in both samples are generally rationalizable, but the U.S. students score substantially higher on consistency. Nevertheless, the development gap in economic rationality between these future elites is much smaller than the difference in performance on a canonical cognitive ability test, often used as a proxy for economic decision-making ability in studies of economic development and growth. We argue for the importance of including consistency with economic rationality in studies of decision-making ability.

JEL Classification Numbers: C91, D81, F61.

Keywords: political economy, development, experiment, rationality, revealed preference, stochastic dominance, expected utility, cognitive ability, noncognitive ability, IQ, personality traits, Big Five, human capital.

*We thank Siv-Elisabeth Skjelbred for excellent research assistance. We also thank Raymond Fisman, Pamela Jakiela, Edward Miguel and Matthew Polisson for many thoughtful comments and suggestions. 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 (alexander.cappelen@nhh.no); Kariv: University of California, Berkeley (kariv@berkeley.edu); Sørensen: Norwegian School of Economics (erik.sorensen@nhh.no); Tungodden: Norwegian School of Economics (bertil.tungodden@nhh.no).
The authors dedicate this paper to Cecilie Rasmussen, the founding research coordinator of the Choice Lab at NHH Norwegian School of Economics, who passed away far too young on a research trip to Berkeley in spring 2012.

1 Introduction

We use economic laboratory experiments to assess and measure the development gap in economic rationality. We draw the subjects for the experiment from the student body at the University of California (UC), Berkeley, ranked as the world’s top public university and one of the most prestigious (public or private) universities in the US, and the University of Dar es Salaam, the oldest and largest public university in Tanzania, and one of the top-ranked universities in the whole of Africa.¹

These subject pools carry an *intrinsic* interest. Although the students at UC Berkeley and University of Dar es Salaam come from different backgrounds and face different economic prospects, they are united in being among the most able in their respective 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 University of Dar es Salaam will have a disproportionate impact in various sectors of their economy when they graduate. This is especially true for the Tanzanian students in our sample, many of whom will assume positions of substantial power in national economic and political affairs.² Thus, the decision-making ability of these of future elites can have a large and highly disproportionate impact on the future of the country as a whole.

The idea that decision-makers vary in their decision-making ability—and therefore make choices of different decision-making quality—has intuitive appeal. However, definitive judgement about which choices exhibit low decision-making quality—and which decision-makers have inferior decision-making ability—requires high-quality data at the individual level, where all the relevant trade-offs should be clear,

¹The Webometrics Ranking of World Universities (www.webometrics.info) ranks all universities worldwide based on their web impact. In this, UC Berkeley is ranked fourth best worldwide, and while the University of Dar es Salaam is ranked only 1,930, it is ranked first in Tanzania and 33rd in Africa.

²The President of Tanzania, John Magufuli, his predecessor, Jakaya Kikwete, the Prime Minster, Kassim Majaliwa, and his predecessor, Mizengo Pinda, all graduated from the University of Dar es Salaam. Other alumni are Gertrude Mongella (first president of the Pan-African Parliament), Asha-Rose Migiro (former Deputy Secretary-General of the United Nations), Yoweri Museveni (President of Uganda), and Willy Mutunga (retired Chief Justice of Kenya), among others.
such that the classification of some decisions as “mistakes” is uncontroversial. Perhaps as a result, researchers have largely not incorporated explicit measures of decision-making ability in economic development research, but instead used measures of education and cognitive and noncognitive skills to consider how human capital affects important economic outcomes.\(^3\)

In this paper, we measure decision-making ability using the consistency of choices with economic rationality, taking the view that if there is no utility function that the choices maximize, then those choices cannot be of high decision-making quality. We are motivated by the result in Choi et al. (2014) showing that consistency with utility maximization is strongly related to household wealth differentials in a large and diverse sample.\(^4\) Adopting this standard for decision-making ability, we present subjects with an economic choice experiment in which we can measure the economic rationality of their choices with a high degree of precision. We further examine strengthening our test of economic rationality by demanding compliance with first-order stochastic dominance and expected utility, building on the work of Polisson et al. (2020). By implementing this approach in relevant subject pools in a developed (US) and a developing (Tanzania) country, we provide preliminary evidence about whether there is a development gap in economic rationality.\(^5\)

**The experiment.** Because uncertainty is endemic in a wide variety of circumstances, we provide an experimental test of the touchstones of rationality in decision-making under risk—namely, utility maximization, stochastic dominance, and expected-utility maximization. Inconsistencies with the generalized axiom of revealed preference (GARP) and violations of first-order stochastic dominance (FOSD) are considered irrational because they leave “money on the table.” While there is no need to assume expected utility (EU) to investigate rational behavior under uncertainty (in

\(^3\)Some studies use levels of basic financial knowledge as a measure of decision-making ability. Lusardi and Mitchell (2014) provide a review of the growing body of work on financial literacy. This literature considers financial knowledge as a form of human capital and demonstrates the impact of financial knowledge on economic decision-making.

\(^4\)Choi et al. (2014) show that a one-standard deviation increase in the rationality score is associated with 15-19% more wealth, conditional on socioeconomic variables including current income, education, family structure, risk tolerance, and the results of standard tests of cognitive and noncognitive ability. As Choi et al. (2014) note, predicting wealth differentials provides a particularly strong test of the measure of decision-making ability. The test is strong because it does not just examine the power of choices in the laboratory to predict related choices in a similar natural decision environment. Instead, wealth accumulation is determined by countless individual decisions, successively made over time in many different environments, and involves a host of different trade-offs concerning risk, time, and personal and social consumption. This substantially increases our likelihood of rejecting a relationship (external validation).

\(^5\)Kim et al. (2018) use the experimental technique to measure economic rationality in a sample of secondary-school students in Malawi. They show that an education intervention enhanced economic rationality – as measured by consistency with utility maximization – in the long run.
the sense of a complete, transitive, and monotonic preference ordering), EU does serve as a normative guide for choice under risk (how people ought to choose).\footnote{Choices can be consistent with GARP and yet fail to be reconciled with any utility function that is normatively appealing given the decision problem at hand. As noted by Quiggin (1990) and Wakker (1993), prominent models that have been proposed as alternatives to EU were amended to avoid violations of FOSD.}

In the experiment, we present subjects with a sequence of standard consumer decision problems: that is, selection of a bundle of commodities from a standard budget set. More specifically, 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 $x_s$ is 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 (prices normalized by income so that $p_1 x_1 + p_2 x_2 = 1$). We present these decision problems using the graphical interface introduced by Choi et al. (2007b).\footnote{We are building on expertise acquired in previous work. Ahn et al. (2014) extended the earlier experimental work of Choi et al. (2007a, 2014) from settings with risk (known probabilities) to those with ambiguity (unknown probabilities). Others have also analyzed the data sets in Choi et al. (2007a) and Choi et al. (2014), including Halevy et al. (2018) and Polisson et al. (2020). Fisman et al. (2007, 2015b,a, 2017) and Li et al. (2017) employ a similar experimental methodology to analyze social preferences with different samples.} 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 large individual-level data set.

**Economic rationality.** The most basic question to ask about choice data is whether they are consistent with individual utility maximization. If budget sets are linear (as in our experiment), classical revealed preference theory (Afriat, 1967; Varian, 1982, 1983) provides a direct test in that choices in a finite collection of budget sets are consistent with maximizing a well-behaved utility function if and only if they satisfy GARP. We assess how nearly the data comply with GARP by calculating Afriat’s (1972) critical cost efficiency index (CCEI). The CCEI, denoted by $e^*$, is bounded by zero and one. The closer it is to one, the smaller the perturbation of budget sets required to remove all violations and thus the closer the data are to being rationalizable. We can interpret the CCEI as saying that decision-makers are ‘wasting’ as much as $1 - e^*$ of their income as possible by making inconsistent choices.

Consistency with GARP requires consistent preferences over all possible alter-
natives, but any consistent preference ordering is admissible. In this way, we see consistency as a necessary, but not sufficient, condition for high decision-making ability. Polisson et al. (2020) strengthen GARP to test whether data are FOSD-rationalizable (comply with GARP and FOSD) and EU-rationalizable (comply with GARP, FOSD, and the independence axiom upon which EU is based), and also calculate CCEI-type scores for FOSD-rationalizability and EU-rationalizability, which we denote by $e^{**}$ and $e^{***}$. By definition, $1 \geq e^{*} \geq e^{**} \geq e^{***}$ – choices are perfectly EU-rationalizable if $1 = e^{*} = e^{**} = e^{***}$ and perfectly FOSD-rationalizable but not EU-rationalizable if $1 = e^{*} = e^{**} > e^{***}$. The $e^{*}$, $e^{**}$, and $e^{***}$ scores thus provide summary statistics of the overall consistency with different levels of rationalizability, reflecting the minimum adjustments required to eliminate all rationality violations associated with the data set.

We emphasize that although our experimental design solicits only individual decisions, our tests also apply to all decisions that have a strategic component – in nearly all game theory, each player is assumed to maximize the expected value of a utility function. Hence, EUT is part of the core of strategic rationality, not only individual rationality. We also note that all tests are individual-level and purely nonparametric, making no assumptions about the parametric form of the underlying utility function.

Figure 1 summarizes our individual-level analyses by showing mean $e^{*}$, $e^{**}$, and $e^{***}$ scores and 95% confidence intervals in the two samples. As shown, the US subjects display greater levels of rationality than the Tanzanian subjects: the mean $e^{*}$, $e^{**}$, and $e^{***}$ scores across all subjects are 0.95, 0.90, and 0.89 in the US and 0.86, 0.75, and 0.75 in Tanzania, respectively. Many subjects in both samples thus attain high scores on these measures of economic rationality and some are close to the ideal of perfectly rational behavior. Nonetheless, the scores of the US subjects are substantially higher than those of the Tanzanian subjects. The magnitudes imply that relative to the US subjects, the Tanzanian subjects on average ‘waste’ as much as 8.8, 14.2, and 14.0 percentage points more of their earnings by making choices that are not rationalizable, FOSD-rationalizable, or EU-rationalizable, respectively. This provides, to our best knowledge, the first quantifiable and economically interpretable evidence of a development gap in economic rationality.

It is also worth noting that in both samples, the differences between perfect ra-

---

8At the risk of oversimplifying, the canonical experimental tests of strategic rationality (Ho et al. (1998), and Costa-Gomes et al. (2001), among others) focus on iterated elimination of dominated strategies, and the “level” of rationality is measured by the number of levels of elimination of dominated strategies that subjects actually do. Limited iterated reasoning is similar to FOSD-rationalizability – a failure to recognize that some actions yield unambiguously lower payoffs.
tionalizability and the subjects’ rationalizability \((1 - e^*)\) and between the subjects’ rationalizability and FOSD-rationalizability \((e^* - e^{**})\) are much larger than the difference between the subjects’ FOSD-rationalizability and the EU-rationalizability \((e^{**} - e^{***})\). In fact, for many subjects, choices are not less EU-rationalizable than FOSD-rationalizable \((e^* - e^{**})\). This is consistent with the results in Polisson et al. (2020) which applies these same tests to experimental data previously collected by Choi et al. (2007a) and Halevy et al. (2018) with student samples and Choi et al. (2014) with a broadly population representative sample.

The justification for measuring economic decision-making ability using the compliance of choices with economic rationality is strong. The key in this regard is the distinctive and robust connection of our approach to economic theory. Revealed preference tests offer a theoretically disciplined metric for the quality of economic decisions. The measure also has a well-established economic interpretation and revealed preference theory tells us whether we have sufficient data to make it statistically useful. In this way, our experimental platform and analytical techniques offer a theory-based tool for measuring decision-making ability as an aspect of human capital in economic development research.

**Cognitive ability.** An important question in the present analysis is whether cognitive ability tests also capture decision-making ability. Hanushek and Woessmann (2008) provide an excellent, although now somewhat dated, review of the role of cognitive skills in economic development. Their conclusion is “that the cognitive skills of the population—rather than mere school attainment—are powerfully related to individual earnings, to the distribution of income, and to economic growth.” Hanushek and Woessmann (2012) concentrate directly on the role of cognitive skills and provide further evidence about the cognitive skills–growth relationship. This strand of empirical research shows that differences in economic well-being is driven, at least in part, by differences in aspects of human capital proxied by tests of cognitive ability (that is, the intelligence quotient or IQ).  

Our subjects also completed (part of) a standard (nonincentivized) Wechsler Adult Intelligence Scale (WAIS-IV) test. This IQ test generated substantial vari-

---

9See Murnane et al. (1995), Heckman et al. (2006), and Hanushek and Woessmann (2008) for the relationships between IQ scores and economic and social behaviors and outcomes. Frederick (2005), Burks et al. (2009), Oechssler et al. (2009), Dohmen et al. (2010), and Benjamin et al. (2013), among others, find that variation in so-called preference anomalies—such as impatience over short time horizons and small-stakes risk aversion—is related to variation in cognitive ability. Borghans et al. (2009) find that differences in the traits of cognition, as well as personality traits, account for some of the interpersonal variation in risk aversion (but not in ambiguity aversion). In response to Dohmen et al. (2010), Andersson et al. (2016) show – in an experiment with a large subject pool drawn from the general population – that cognitive ability is related to random decision-making rather than to risk preferences.
ation in both samples and the scores are only weakly correlated with the $e^*$, $e^{**}$
and $e^{***}$ scores from the experiment (0.061, 0.120, and 0.120 in the US and 0.109,
0.165, and 0.159 in Tanzania). Figure 1 shows that the country difference in IQ
scores is much larger than the country difference in economic rationality. In the IQ
test, the US students score 35.4 percentage points higher than the Tanzanian stu-
dents, which is equivalent to 1.45 standard deviations (compared with 0.66, 0.68,
and 0.68 standard deviations in the $e^*$, $e^{**}$, and $e^{***}$ scores). The IQ test thus
indicates a much larger development gap than the revealed preference tests.\footnote{Jones (2015) reviews work that relates differences in IQ between countries to differences in macroeconomic outcomes and discusses what he calls possible “IQ–productivity channels.” Lynn et al. (2002); Lynn and Vanhanen (2006) asserts that IQ is a predictor of GDP. Nonetheless, their work has drawn widespread criticism, including by Ervik (2003), Nechyba (2004), and Palairet (2004).}

To further assess the development gap in economic rationality, we control for
noncognitive skills measured by the Big Five personality traits. These influential
measures from psychology are derived from factor analysis of wide-ranging person-
ality surveys and comprise conscientiousness, openness, extraversion, agreeableness,
and neuroticism. Among these, conscientiousness has the strongest correla-
tion with high-quality economic decision-making. Duckworth et al. (2007) describe
conscientious people as “thorough, careful, reliable, organized, industrious, and
self-controlled,” which are all aspects of decision-making ability.\footnote{See Block (2010) for a description and assessment of the Big Five. Borghans et al. (2008), Almlund et al. (2011), and Becker et al. (2012) discuss the relationship between personality psychology and economics. Laajaj et al. (2019) discuss the interpretation of the Big Five personality traits in low-income populations.} Such noncognitive skills are also important elements of human capital that are related to job mar-
ket outcomes and earnings differences (Heckman et al., 2006). However, we find
that adding controls for the Big Five has very little effect on the estimated coeffi-
cients, which suggests that the development gap in economic rationality is unlikely
driven by unmeasured correlations between our economic rationality measures and
personality traits.

In sum, we observe a development gap in economic rationality when compar-
ing future elites in both a rich country and a poor country, and show that decision-
making ability is not fully captured by existing measures of cognitive and noncog-
nitive skills. Typically, definitive judgment about decision-making quality is made
difficult by the twin problems of identification and measurement (Kariv and Sil-
verman, 2013). We argue that the tool kit used in this paper—comprising the ex-
perimental method and analytical techniques—addresses these problems, and thus
represents a promising approach to measuring an aspect of human capital that may
be of great importance when studying economic development.\footnote{This approach is also important for the examination of paternalistic policies aimed at steering people toward making better decisions (Camerer et al., 2003; Thaler and Sunstein, 2003).}
The remainder of the paper is organized as follows. Section 2 describes the experimental design and procedures, and Section 3 explains the measures of rationality. Section 4 reports the experimental results, and Section 5 provides some concluding remarks and summarizes what the reader should take away from the analysis.

2 The Experiment

2.1 Subject Pools

We conducted the experiments at UC Berkeley (126 subjects) and the University of Dar es Salaam in Tanzania (216 subjects). The US pool has a much higher proportion of females (70.6% versus 33.8%) and younger subjects (average age 20.6 years versus 23.3 years). As expected, the parents of the US subjects are much more educated – 82.5% of the US subjects have at least one parent with a college/university education compared with only 30.6% of the Tanzanian subjects. All our results are robust to the inclusion of controls for gender, age and parental education.

The US subjects in the experiment were recruited from all undergraduate classes at UC Berkeley. While student subjects at major American universities have been extensively studied, our subjects from the University of Dar-es-Salaam are from a less studied population. While the university is in Dar-es-Salaam, we have subjects from 25 out of the 30 Tanzanian regions (per 2012 administrative boundaries). Those born in Dar-es-Salaam are well represented (16.7%), but there are slightly more from Kagera (17.2%) and almost as many from Kilimanjaro (13.5%). Only 6% or less of the subjects come from each of the other regions. According to data by the National Bureau of Statistics (Tanzania), the Dar-es-Salaam region’s GDP per capita is 5,699 USD (PPP in 2018), the highest in Tanzania. The Kilimanjaro and Kagera regions are ranked 6th and 23rd with GDP per capita of 4,029 USD and 1,773 USD, respectively.

The most popular reported field of study among our Tanzanian subjects is one or more of the social sciences (42.8%), with science and engineering close behind (31.6%). Education (9.8%) and humanities (8.4%) are also represented. On average, the Tanzanian subjects are 2.3 years into their study (with 4 as the maximum). The median Tanzanian subject reported yearly expenses of 2 million TZS, corresponding to about 1,200 USD (at 2012 exchange rates). Many report that they have support from their family (80.5%), with a median support of 30,000 TZS (equivalent to 18 USD) per week, and almost all (99.5%) report that they are at least partly funded by government loans (with, on average, 68.6% of the costs of study covered by the student loans).
2.2 Design and Procedures

The experimental procedures described below are identical to those used by Choi et al. (2007a). 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). No subject reported difficulty understanding the procedures or using the computer interface. 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 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 to move the pointer on the computer screen to the desired allocation. Choices were restricted to allocations on the budget constraint.\(^\text{13}\)

The number of tokens in the $x$ and $y$ accounts determined the payoff in each decision round. At the end of the round, the computer randomly selected one of the accounts, $x$ or $y$, with the two accounts 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.

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 US dollars (USD) in the US and 100 Tanzanian shillings (TZS) (equivalent to 0.06 USD) in the low-stakes treatment in Tanzania (106 subjects), which were roughly comparable in purchasing power terms. In addition, we conducted a high-stakes treatment in Tanzania (110 subjects) where each token was worth 1,000 TZS. We pool the data from the low- and high-stakes treatments in Tanzania in the main analysis, but all results hold if we analyze the treatments separately.

\(^\text{13}\)Choi et al. (2007a) also restricted choices to allocations on the budget line, which makes the computer program easier to use. 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.
Our subjects also completed a nonincentivized 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. 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 were 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 in John et al. (2008).

3 Testing for Rationalizability

Let \( \{(p^i, x^i)\}_{i=1}^{50} \) be the data generated by some individual’s choices, where \( p^i \) denotes the \( i \)-th observation of the price vector and \( x^i \) denotes the associated allocation.

3.1 Rationalizability

We first test whether the choices can be utility-generated. The crucial test for this is provided by the GARP. Afriat (1967) shows 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. GARP requires that if \( x^i \) is indirectly revealed as being preferred to \( x^j \), then \( x^j \) is not strictly directly revealed as being preferred to \( x^i \) (\( x^i \) must cost at least as much as \( x^j \) at the prices prevailing when \( x^j \) is chosen, \( p^j \cdot x^i \geq p^j \cdot x^j \)). Our definition follows Varian (1982), which replaced Afriat’s (1967) “cyclical consistency” condition with GARP.15

Given GARP offers an exact test (either the data satisfy GARP or they do not) and choice data almost always contain at least some violations, we assess how nearly the data comply with GARP by calculating Afriat’s (1972) CCEI, denoted by \( e^* \). By definition, \( 0 \leq e^* \leq 1 \) and the closer it is to 1, the smaller the perturbation of

---

14Subjects 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 \( \times \) 2 matrix with one missing element. Subjects were shown one correct example of each type of question before starting the test.

15We refer the interested reader to Choi et al. (2007a) for further details on the testing for consistency with GARP. Choi et al. (2007b) also show that if utility maximization is not in fact the correct model, then our experiment is sufficiently powerful to detect it. See Afriat (2012), Diewert (2012), Varian (2012), Vermeulen (2012), and Chambers and Echenique (2016) for a review of revealed preference theory.
the budget constraints required to remove all violations and thus the closer the data are to perfect consistency with GARP.\footnote{Put precisely, for any number \(0 \leq e \leq 1\), define the direct revealed preference relation \(R^D(e)\) as \(x^i R^D(e)x^j\) if \(ep^i \cdot x^i \geq p^j \cdot x^j\), and define \(R(e)\) to be the transitive closure of \(R^D(e)\). \(e^*\) is the largest value of \(e\) such that the relation \(R(e)\) satisfies GARP. We interpret this as saying that a subject is ‘wasting’ as much as \(1 - e^*\) of his or her earnings by making choices inconsistent with GARP.}

### 3.2 FOSD-rationalizability

Choices can be consistent with GARP and yet fail to be reconciled with any utility function that is normatively appealing given the decision problem at hand. Given the two states are equally likely, allocating fewer tokens to the cheaper security \((x_s < x_s'\text{, when } p_s < p_s')\) is a violation of monotonicity with respect to FOSD. Violations of FOSD 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. As Starmer (2000) points out, “it is widely held that any satisfactory theory—descriptive or normative—should embody monotonicity.”

To test whether choice behavior satisfies GARP and FOSD (for a given subject), we combine the actual data from the experiment and the mirror-image data and compute the CCEI for this combined data set.\footnote{The data generated by an individual’s choices are \(\{(\bar{x}_i, x_i')\}^{50}_{i=1}\), where \(\bar{x} = (\bar{x}_1, \bar{x}_2)\) are the endpoints of the budget line. The mirror-image data are obtained by reversing the prices and the associated allocation for each observation \(\{(\bar{x}_1', x_1', x_2')\}^{50}_{i=1}\).} Clearly, any decision to allocate fewer tokens to the cheaper security (positions along the shorter side of the budget line relative to the 45-degree line) will necessarily generate a simple violation of the weak axiom of revealed preference (WARP) involving its mirror-image decision.

By definition, the CCEI score for the combined data set consisting of 100 observations can be no bigger than the CCEI score for the actual data. Polisson et al. (2020) show that when states are equally likely (as in our experiment), the CCEI score for the combined data set is a measure of consistency with GARP and FOSD, which we call FOSD-rationalizability and denote by \(e^{**}\). By definition, \(e^{**} \leq e^* \leq 1\) and the difference between \(e^{**}\) and \(e^*\) reflects an upper bound on the additional income that the subject is “wasting” by violating FOSD.

### 3.3 EU-rationalizability

Using the generalized restriction of infinite domains (GRID) method applied by Polisson et al. (2020), we can further strengthen the GARP test to also test for consistency with expected-utility maximization, which we call EU-rationalizability.
and denote by $e^{***}$. Any choice data consistent with the familiar non-EU theories, such as weighted expected utility (Chew, 1983), implicit expected utility (Dekel, 1986), and prospect theory (Kahneman and Tversky, 1979) are FOSD-rationalizable, but not EU-rationalizable. To see this, consider the rank-dependent utility function (Quiggin, 1993):

$$U(x) = \omega u(\min \{x_1, x_2\}) + \overline{\omega} u(\max \{x_1, x_2\}),$$

where $0 < \omega \leq \overline{\omega} < 1$ are weights and $u(\cdot)$ is the Bernoulli utility function defined on the amounts of money. This formulation encompasses a number of different non-EU models and reduces to EU when $\omega = \overline{\omega}$. When $\omega > \overline{\omega}$—interpreted as “pessimism”—the indifference curves have “kinks” at safe allocations $x_1 = x_2$ that lie on the 45-degree line. Such allocations will be chosen for a nonnegligible set of price ratios around $-1$, which is inconsistent with EU (as prices are randomly generated, smooth preferences should give rise to allocations satisfying $x_1 = x_2$ with probability zero).

Figure 2 illustrates a simple violation of EU-rationalizability involving three allocations, $x^1 = (a, a)$, $x^2 = (b, b)$, and $x^3 = (c, d)$ (Example 2 in Polisson et al. (2020)). The price vectors are such that $B(p^1)$ is the flattest and $B(p^3)$ is the steepest. Note that $x^1$ and $x^2$ are safe allocations $x_1 = x_2$ consistent with infinite risk aversion, whereas $x^3$ is not. As Choi et al. (2007a) show, these choices are individually consistent with EU, but mutually inconsistent. In fact, they are consistent with pessimism or disappointment aversion (Gul, 1991) where the safe allocation is the reference point. To see why, suppose that $\omega = \overline{\omega}$ so we have the standard EU representation and note that

$$U(a, a) = 2u(a) \geq u(b) + u(c),$$

since $(b, c)$ is cheaper than $(a, a)$ at the prices at which $(a, a)$ is chosen, and

$$U(b, b) = 2u(b) \geq u(a) + u(d),$$

given $(a, d)$ is cheaper than $(b, b)$ at the prices at which $(b, b)$ is chosen. But rearranging yields $u(a) + u(b) \geq u(c) + u(d)$, which contradicts that $(c, d)$ is directly revealed as being preferred to $(a, b)$—$(a, b)$ costs less than $(c, d)$ at the prices prevailing when $(c, d)$ is chosen. It is clear that these choices are consistent with GARP and FOSD so the data are FOSD-rationalizable, but they are not EU-rationalizable. In this example, $e^{***} < e^{**} = e^* = 1$, which is perfectly consistent with a non-EU alternative.

---

18Other papers that provide revealed preference tests of whether choice data are consistent with a utility function with some special structure include Green and Srivastava (1986), Varian (1988), Bayer et al. (2013), Echenique and Saito (2015), Chambers and Echenique (2016), and Chambers et al. (2016). See Polisson et al. (2020) for a discussion.
4 Results

We first test whether individual choices can be rationalized, in the sense that they satisfy the GARP. Consistency with GARP requires consistent preferences, but any consistent preference ordering is admissible. In this way, we see GARP as a necessary—but not sufficient—condition for choices to be considered of high decision-making quality. We then strengthen GARP to test whether choice behavior is also FOSD-rationalizable and EU-rationalizable. To account for the possibility of errors, we assess how closely individual choice behavior complies with rationalizability, FOSD-rationalizability, and EU-rationalizability by calculating the efficiency indices described above $e^*$, $e^{**}$, and $e^{***}$, where $e^{***} \leq e^{**} \leq e^* \leq 1$.

Figure 3 tabulates the proportion of US and Tanzanian subjects whose $e^*$ (Figure 3A), $e^{**}$ (Figure 3B), and $e^{***}$ (Figure 3C) lie above different critical values. This provides a clear graphical illustration of the extent to which subjects did worse than choosing rationally and the extent to which the US subjects did better than the Tanzanian subjects. For comparison purposes, we also tabulate the proportion subjects (Figure 3D) whose IQ score (fraction of questions answered correctly) is above the same critical values. If we choose 0.90 as our critical value, we find that 87.5%, 67.5%, and 65.1% of the US subjects have $e^*$, $e^{**}$, and $e^{***}$ above this threshold, while only 52.3%, 37.0%, and 34.7% of the Tanzanian subjects have scores that high. Hence, the probability that a US subject will pass each of the rationalizability tests—utility maximization, stochastic dominance, and expected-utility maximization—is about one-third higher.\footnote{If we follow the purely subjective threshold of 0.95 suggested by Varian (1991), the corresponding percentages are 65.9%, 48.4%, and 43.7% for the US subjects and 40.7%, 20.4%, and 17.6% for the Tanzanian subjects for $e^*$, $e^{**}$, and $e^{***}$, respectively. Bronars (1987) calibrated the CCEI using a hypothetical subject whose choices are uniformly distributed on the budget line. The mean $e^*$ for a random sample of 25,000 simulated subjects is only 0.60. Choi et al. (2007b) generated a benchmark level of consistency using hypothetical subjects with idiosyncratic preference shock that has a logistic distribution.}

We also note that the vast majority of subjects in both samples are not less EU-rationalizable than FOSD-rationalizable—in fact, the difference between $e^{**}$ and $e^{***}$ is less than 0.01 for 85.7% of the US subjects and 94.0% of the Tanzanian subjects. By contrast, the values of $e^{**}$ in both samples are distinctly lower than those of $e^*$ so FOSD-rationalizability is systematically lower than rationalizability (which is systematically lower than perfect rationalizability). Polisson et al. (2020) also find that EU-rationalizability does not alter the goodness-of-fit relative to the restrictions already imposed by FOSD-rationalizability using the experimental data.
collected by Choi et al. (2007a, 2014) and Halevy et al. (2018).

We next turn to individual-level regression analyses. We define indicators for both the Tanzania sample and for the high-stakes experimental treatment (in Tanzania). We also include age, gender, a measure of parental education, and a nonparametric measure of risk attitudes—the average share of tokens the subject allocated to the cheaper security. We note that there is considerable heterogeneity in risk attitudes in both samples. On average, the US subjects allocated 64.3% of their tokens to the cheapest security, whereas the Tanzanian subjects allocated 61.7% and 59.2% of their tokens to the low- and high-stakes treatments, respectively.

We also note that risk attitudes and the $e^*$, $e^{**}$, and $e^{***}$ scores are only weakly correlated in the US (0.066, 0.117, and 0.124) and effectively uncorrelated in Tanzania (−0.025, −0.026, and −0.022). The risk attitudes and IQ scores are also only weakly correlated in the US (0.210) and effectively uncorrelated in Tanzania (−0.023).

Table 1 reports the results of the ordinary least squares (OLS) estimates. The Tanzanian sample size decreases from 216 to 212 subjects due to missing survey information. In column (1), we present estimates for the baseline specification with the $e^*$ scores as the dependent variable. In columns (2) and (3) we repeat the estimation reported in column (1) using the $e^{**}$ and $e^{***}$ scores. The point estimates of −0.079, −0.143, and −0.140 for the coefficients on Tanzania are highly significant and robust to the size of the stakes in the experiment in Tanzania. These results are consistent with the development gap in rationalizability, FOSD-rationalizability, and EU-rationalizability presented in Figure 1.

In columns (4)–(6), we include the proportion of questions answered correctly in the IQ test to the baseline specification. The point estimates of the coefficient on the IQ are 0.066 for $e^*$, 0.176 for $e^{**}$, and 0.169 for $e^{***}$ and all are statistically significant. Adding IQ to the basic specification reduces the estimated coefficients for Tanzania somewhat. These results suggest that IQ captures some of the development gap in economic rationality measured by $e^*$, $e^{**}$, and $e^{***}$, but cannot fully

---

20 An important advantage of our experimental task is that it delivers measures of rationality and (risk) preferences from a single realm of decision-making. We summarize attitudes toward risk with a single statistic: the average share of tokens allocated to the cheaper security. Clearly, subjects who are less averse to risk will allocate a larger share of tokens to $x_1$ when $p_r < p_s$ (infinite risk aversion is consistent with always allocating the tokens equally and risk neutrality is consistent with always allocating all tokens to the cheaper security). Like the rationalizability tests, an advantage of this measure is that it is nonparametric and thus measures attitudes toward risk without making assumptions about the parametric form of the underlying utility function.

21 Our individual-level measures of risk aversion are similar to those in Choi et al. (2007a), and they are also within the range of estimates from other studies. For a discussion of these studies, see Choi et al. (2007a) and Halevy et al. (2018).
account for the observed country difference. In a final effort to evaluate the development gap in economic rationality, in columns (7)–(9) we add the list of controls for the Big Five personality traits. Adding these controls has almost no effect on the estimates of the coefficient for Tanzania, which shows that the differences in decision-making ability between the US and Tanzania subjects are not driven by personality traits. We also observe that adding the Big Five does not affect the estimated coefficients for IQ, which shows that cognitive and noncognitive skills capture independent aspects of human capital.

Finally, for the sake of comparison, in column (10) we repeat the estimation in the baseline specification using the subjects’ IQ scores instead of the economic rationality scores as the dependent variable. In column (11) we add the list of controls for the Big Five personality traits. The coefficients on Tanzania show that the estimated development gap in IQ is more than twice as large as the estimated development gap in our measures of economic rationality.

[Table 1 here]

5 Conclusion

In this paper, we tested the touchstones of economic rationality in a choice under risk experiment. Our subjects are students from UC Berkeley, one of the best universities in the US, and students at the University of Dar es Salaam, one of the best universities in Africa. The Tanzanian and US subjects differ substantially in their sociodemographic and economic backgrounds and face very different economic prospects. Nevertheless, they represent similar ‘slices’ of the future elites in their respective societies. As graduates of top universities in their country, the students of UC Berkeley and the University of Dar es Salaam will have a disproportionate impact on their economies and societies.

Our results are summarized as follows. We find that the US subjects display a higher degree of economic rationality than the Tanzanian subjects. We use revealed preference tests to measure the extent of violations of utility maximization, stochastic dominance, and expected-utility maximization. The magnitudes imply that the Tanzanian subjects on average waste about nine percentage points more of their earnings than the US subjects by making choices that are inconsistent with utility maximization. This provides a quantifiable and economically interpretable measure of the development gap in economic rationality. Also requiring monotonicity with respect to FOSD and expected-utility maximization widens the development gap in economic rationality to about 14 percentage points. These results are robust to the inclusion of controls for cognitive and noncognitive skills. At the same time,
our results show that the observed development gap in decision-making ability is much smaller than the country difference in performance on a canonical cognitive ability test.

We argue that the justification for measuring decision-making ability by compliance with economic rationality is strong because of the distinctive and robust connection of these measures to economic theory. This connection makes the measure economically quantifiable and interpretable. Moreover, the same economic theory that inspires the measure also tells us when we have enough data to make it statistically useful. Furthermore, because we present subjects with a standard economic decision problem, the revealed preference tests are applicable to, and comparable across, all sorts of economic choice problems. Thus, another important advantage of our methods is that they can be transported—with relative ease—to different decision domains, including intertemporal choice and social choice. The approach thus allows for both domain-specific predictions and a unified measure of decision-making quality across domains.

The experimental technique offers a new opportunity to improve on existing laboratory and field experimental methods in economic development research. The graphical interface has been integrated with web-based survey instruments by Choi et al. (2014) and Fisman et al. (2017) to conduct experiments with large representative samples in developed countries, and the present study and Kim et al. (2018) show that the tool is applicable in a development setting. We did not experience any significant problems in the implementation of the experimental technique in our study, and believe that the simple experimental interface would also make it highly feasible to implement with nonstudent populations in low-income countries.

The aim of our paper has been to illustrate how economic theory and experimental techniques can be used to study decision-making ability in economic development research. We believe that there are several important research avenues to pursue with this tool. First, it would be interesting to consider economic rationality in a development setting more broadly. We focused this paper on the high end of the ability spectrum and on the risk domain, but it is also important to understand the decision-making ability of the poorest of the poor across the key domains in economics (time, risk, and social). These people also make many critical decisions in their daily lives, where leaving money on the table may be highly consequential. Second, it would also be of great importance to examine what kind of intervention can be used to improve decision-making ability (Kim et al., 2018). Decision-making ability is an essential aspect of human capital that should lie at the forefront of studies of economic development.
References


Figure 1: Mean rationalizability violation and IQ scores, by country

Note: Mean $e^*$, $e^{**}$, $e^{***}$ and IQ scores and 95% confidence intervals in the two samples. $e^*$, $e^{**}$, $e^{***}$ measure how close the individual’s choice data are to being rationalizable, FOSD-rationalizable, and EU-rationalizable. By definition, $e^{***} \leq e^{**} \leq e^* \leq 1$. IQ scores refer to the proportion of questions answered correctly.
Figure 2: An example of choices that are FOSD-rationalizable but not EU-rationalizable

Note: Example 2 in Polisson et al. (2020). These choices are consistent with maximizing a rank-dependent utility function but inconsistent with expected-utility maximization. In this example, $e^{***} \leq e^{**} = e^* = 1$. 
Figure 3: The proportion of subjects above different rationalizability levels, by country

Note: The proportions of subjects whose $e^*$, $e^{**}$, and $e^{***}$ are above different critical values (the survival function, $S(e) = 1 - F(e)$, where $F$ is the empirical distribution function of the index $e = e^*, e^{**}, e^{***}$).
### Table 1: The development gap in rationalizability and cognitive ability

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Baseline + IQ</th>
<th>Baseline + IQ and Big-5</th>
<th>Cognitive ability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$e^*$</td>
<td>$e^{**}$</td>
<td>$e^{***}$</td>
<td>IQ</td>
</tr>
<tr>
<td><strong>Tanzania</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(10)</td>
</tr>
<tr>
<td>$e^*$</td>
<td>$-0.079$</td>
<td>$-0.143$</td>
<td>$-0.140$</td>
<td>$-0.316$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$-0.058$</td>
<td>$-0.087$</td>
<td>$-0.087$</td>
<td>$-0.315$</td>
</tr>
<tr>
<td>$e^{***}$</td>
<td>$-0.700$</td>
<td>$-0.106$</td>
<td>$-0.105$</td>
<td>$-0.024$</td>
</tr>
<tr>
<td><strong>High stakes (Tanzania)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.017$</td>
<td>$0.054$</td>
<td>$0.052$</td>
<td>$0.016$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.019$</td>
<td>$0.058$</td>
<td>$0.056$</td>
<td>$0.053$</td>
</tr>
<tr>
<td>$e^{***}$</td>
<td>$0.018$</td>
<td>$0.056$</td>
<td>$0.053$</td>
<td>$0.024$</td>
</tr>
<tr>
<td><strong>Risk aversion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e^*$</td>
<td>$-0.001$</td>
<td>$0.038$</td>
<td>$0.047$</td>
<td>$0.051$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$-0.004$</td>
<td>$0.029$</td>
<td>$0.038$</td>
<td>$0.033$</td>
</tr>
<tr>
<td>$e^{***}$</td>
<td>$-0.007$</td>
<td>$0.028$</td>
<td>$0.039$</td>
<td>$0.016$</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>$-0.002$</td>
<td>$-0.003$</td>
<td>$-0.003$</td>
<td>$-0.002$</td>
</tr>
<tr>
<td>$e^*$</td>
<td>$-0.002$</td>
<td>$-0.003$</td>
<td>$-0.003$</td>
<td>$-0.002$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$-0.002$</td>
<td>$-0.003$</td>
<td>$-0.003$</td>
<td>$-0.002$</td>
</tr>
<tr>
<td><strong>Female</strong></td>
<td>$0.009$</td>
<td>$0.017$</td>
<td>$0.018$</td>
<td>$0.039$</td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.007$</td>
<td>$0.010$</td>
<td>$0.012$</td>
<td>$0.039$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.005$</td>
<td>$0.006$</td>
<td>$0.008$</td>
<td>$0.039$</td>
</tr>
<tr>
<td><strong>Parents have college education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.015$</td>
<td>$0.016$</td>
<td>$0.015$</td>
<td>$0.027$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.014$</td>
<td>$0.011$</td>
<td>$0.010$</td>
<td>$0.027$</td>
</tr>
<tr>
<td><strong>IQ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.066$</td>
<td>$0.176$</td>
<td>$0.169$</td>
<td>$0.067$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.067$</td>
<td>$0.176$</td>
<td>$0.168$</td>
<td>$0.067$</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>$0.974$</td>
<td>$0.917$</td>
<td>$0.902$</td>
<td>$0.816$</td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.920$</td>
<td>$0.773$</td>
<td>$0.765$</td>
<td>$0.844$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.901$</td>
<td>$0.741$</td>
<td>$0.729$</td>
<td>$0.107$</td>
</tr>
<tr>
<td><strong>Big-5 included</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.133$</td>
<td>$0.209$</td>
<td>$0.206$</td>
<td>$0.253$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.133$</td>
<td>$0.209$</td>
<td>$0.206$</td>
<td>$0.253$</td>
</tr>
<tr>
<td><strong>Standard deviation of outcome</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$e^*$</td>
<td>$0.111$</td>
<td>$0.125$</td>
<td>$0.124$</td>
<td>$0.501$</td>
</tr>
<tr>
<td>$e^{**}$</td>
<td>$0.111$</td>
<td>$0.125$</td>
<td>$0.124$</td>
<td>$0.503$</td>
</tr>
</tbody>
</table>

**Note:** Columns (1)–(9): Linear regressions of $e^*$, $e^{**}$ and $e^{***}$ on a country indicator (Tanzania), an indicator for the high-stakes treatment in Tanzania (High), the average share of tokens allocated to the cheaper security (Risk aversion), Age (in years), Gender (dummy for female), whether one or two parents have college education (dummy), IQ score, and controls for the Big Five personality traits. Columns (10) and (11): Linear regressions of IQ scores on a country indicator (Tanzania), an indicator for the high-stakes treatment in Tanzania (High), the average share of tokens allocated to the cheaper security (Risk aversion), Age (in years), Gender (dummy for female), whether one or two parents have college education (dummy), and controls for the Big Five personality traits. Standard errors in parentheses.