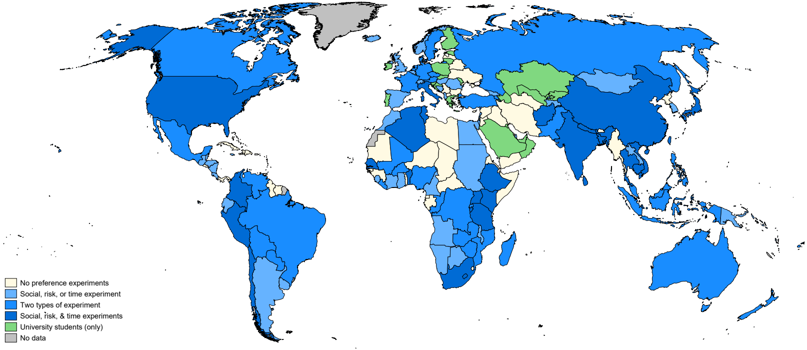
Taking Preferences Seriously, Not Literally

May 2019

Pamela Jakiela
Center for Global Development, BREAD, & IZA

Motivation: Lab-in-the-Field Experiments



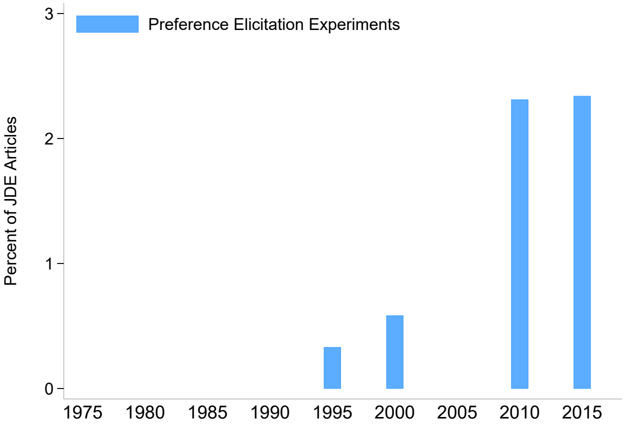
Motivation: Lab-in-the-Field Experiments

Inroads into mainstream development economics?

Search all 2,695 *Journal of Development Economics* abstracts for terms:

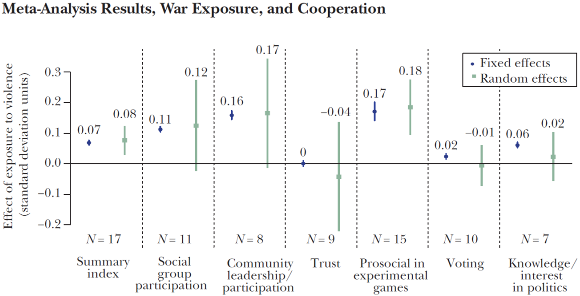
- "lab-in-the-field"
- "laboratory"
- "preference" and "experiment"
- "risk averse" and "experiment"
- "dictator game" or "ultimatum game" or "trust game"
- "hyperbolic discount" and "experiment"
- "loss averse" and "experiment"

Motivation: Lab-in-the-Field Experiments



Data from 2,695 JDE abstracts published between 1974 and 2018.

Conflict and Natural Disasters Change Preferences

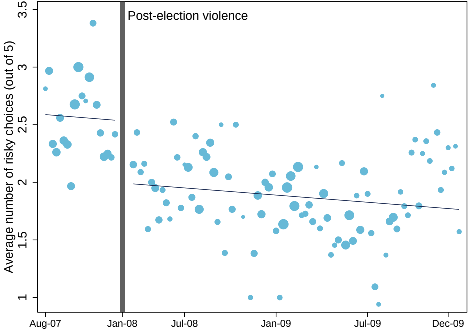


Source: Bauer et al. (2016)

Meta-analysis of 15 experiments estimating (DG, UG, PGG, TG SVO)

Experiments provide clearer evidence that greater conflict exposure increases altruism, cooperation, etc. than more traditional data sources

Conflict and Natural Disasters Change Preferences

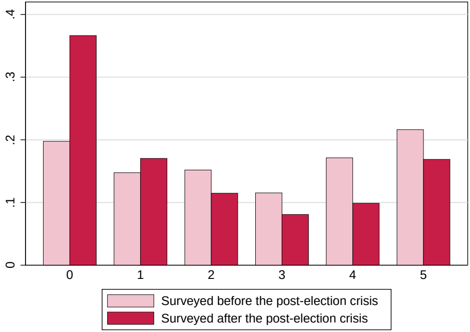


Source: Jakiela & Ozier (2019)

Comparisons of those more vs. less exposed to violence found conflict reduced risk aversion (cf. Voors et al. 2012, Callen et al. 2014)

We found that Kenya's 2008 post-election violence increased risk aversion

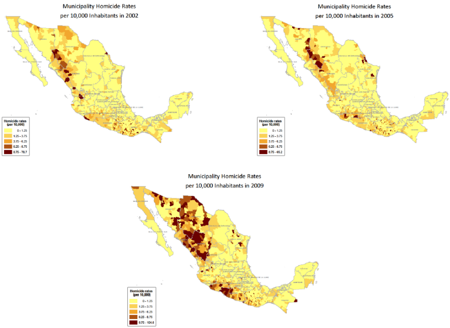
Conflict and Natural Disasters Change Preferences



Source: Jakiela & Ozier (2019)

Living through a conflict doubles likelihood of extreme risk aversion — in a population where very few people had direct experience of victimization

Conflict and Natural Disasters Change Preferences



Source: Brown et al. (2019)

Differing results make sense if any conflict exposure changes preferences, especially if risk aversion predicts migration out of most affected areas

Brown et al. (2019) test this using MxFLS, data from Mexican drug war

Conflict and Natural Disasters Change Preferences

Dependent Variable: Risk Aversion (0 to 100)

MxFLS rounds:	2 only	2 & 3
	(1)	(2)
Homicide rate	-2.203^{***}	1.472^{***}
	(0.722)	(0.465)
Individual fixed effects	No	Yes
Mean of dep. var	17.51	30.82
Observations	11,348	22,696

Source: Brown et al. (2019). Standard errors clustered at the municipality level. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$. All models control for individual characteristics and household characteristics and date of interview fixed effects.

Significant, negative relationship between local-level violence and risk aversion, but sudden increases in violence lead to increases in risk aversion

Conflict and Natural Disasters Change Preferences

Hypothetical elicitation techniques (can) work well in large-scale surveys

- Our hypothetical measure predicts migration, entrepreneurship
- Both Kenya, Mexico studies have $N > 5,000$

All these papers take identification seriously

- Experimental preference elicitation + reduced form identification
- Experiments broaden range of measurable outcomes

Conflict and Natural Disasters Change Preferences

Hypothetical elicitation techniques (can) work well in large-scale surveys

- Our hypothetical measure predicts migration, entrepreneurship
- Both Kenya, Mexico studies have $N > 5,000$

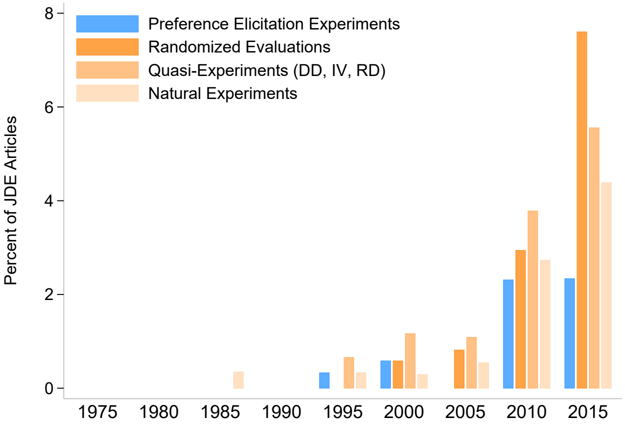
All these papers take identification seriously

- Experimental preference elicitation + reduced form identification
- Experiments broaden range of measurable outcomes

Yet, many of these papers:

- Could have been written with other (non-experimental) outcomes
- Result from fortuitous inclusion of preference measures in surveys

Do Developmentistas Take Preferences Seriously?



Data from 2,695 JDE abstracts published between 1974 and 2018 (Duggan, Jakiela, and Ozier 2037).

Do Developmentistas Take Preferences Seriously?

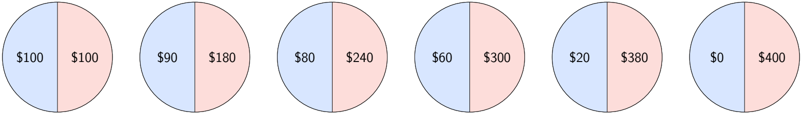
Evidence that **we** don't take preference experiments seriously:

- How many RCTs have been stratified by preferences?
- How many policies have been targeted based on preferences?
- How often are rigorous measures of preferences collected at baseline?

Micro development people now recognize that preference experiments exist, but don't see them as relevant to program evaluation or policy

- Yet, preferences are the central building block of all economic theory
- Need better theoretical links between estimated treatment effects in different contexts and generalize-able policy recommendations

Taking Preferences Literally, Not Seriously



Taking preferences literally:

- Confronting subjects with a single decision problem, then treating preference parameter implied by that single decision as literal truth
  - Human beings are not agents in an Econ 101 problem set!
- Often motivated by desire to keep experiment/instructions simple
- We've fallen off the technological frontier of preference elicitation

Taking Preferences Literally, Not Seriously

People don't have preference parameters inscribed on their foreheads

- We tend to focus on precision, not accuracy

People implement their decisions with error:

- People less familiar with probabilities, expected values, etc. are likely to make more decision errors than university students (Choi et al. 2014, Fisman et al. 2015, Fisman et al. 2017)
- Errors often change the interpretation of behavioral patterns (Hey & Orme 1994, Harless & Camerer 1994, Loomes 2005)

Experimental methods need to account for decision errors

Taking Preferences Literally, Not Seriously

Source: Ensminger (2002)

The wrong lessons from Binswanger (1980), Henrich et al. (2004)?

- Methods should be appropriate for the local context
- Shorter/simpler elicitation approaches aren't always better

Taking Preferences Literally, Not Seriously

Taking language literally: translation

- Some words (eg. random, percent) don't exist in many languages
- Other words (eg. risk in Swahili) have very different connotations

Taking language literally: framing

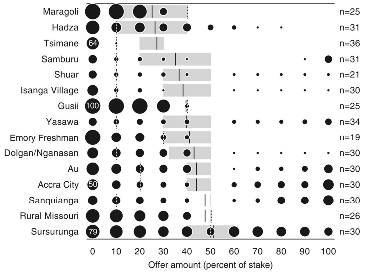
- Framed field experiments may be easier for subjects to understand
- Labels don't change the fundamental nature of the choice

Measurement tools that perform well in university labs may not work in the field, but even simple elicitation tools need adaptation and validation

- Even methods that "work" in the field may work differently

Taking Preferences Literally, Not Seriously

Fig. 1. UG results displayed as the distributions of rejections across possible offers in the UG, which overlay the mean offers and interquartiles.



Source: Henrich et al. (2006)

Many people refuse to accept "too much" in field ultimatum games

- Gifts often come with strings attached, so people don't accept
- We wouldn't know this if Henrich et al. (2006) had followed the common practice of eliciting minimum acceptable offers

How to Take Preferences Seriously

Subjects are trying to tell us something through their choices

- May or may not be their preference parameters
- Progress on preference measurement requires better experimental tools designed to separate signal from noise in LMIC contexts

Remainder of the talk:

- Different approaches to estimating preference parameters in non-student populations with limited education, numeracy, etc.
- Focus on characterizing decision errors
- Identify strategies for building better preference elicitation methods

The Stability of Distributional Preferences

Measuring Noisy Preferences

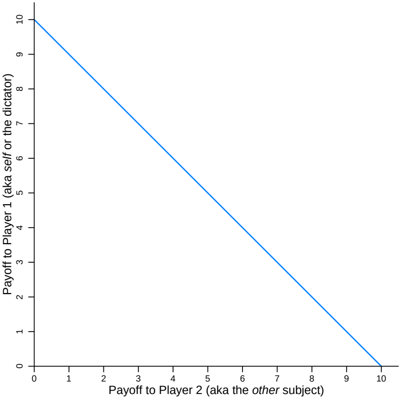
When preferences are noisy, we want to collect multiple observations

- Need to present decision problems in an intuitive way
- Graphical presentations/interfaces allow for this

Choi et al. (2007) propose a simple computer interface for eliciting preferences using decision problems that can be presented as budget lines

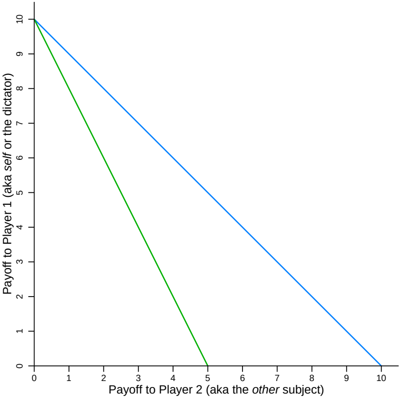
- Can be used to measure social, risk, or time preferences (Choi et al. 2007, Fisman et al. 2007)
- Feasible to collect many decisions from each subject
  - Estimate subject-level preference parameters
  - Test for consistency with economic rationality

A Modified Dictator Game



Standard **dictator game**: Player 1 ("self") receives an endowment of 10, and chooses an amount $x \in [0, 10]$ to allocate to Player 2 ("other")

A Modified Dictator Game

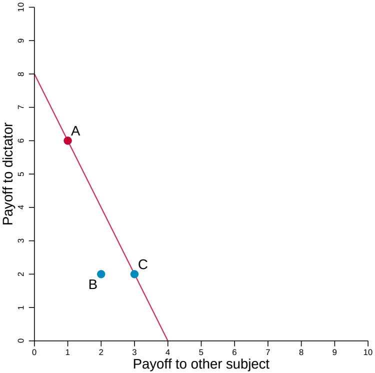


Modified **dictator game**: Player 1 chooses $\pi = (\pi_i, \pi_j)$ given p_i, p_j such that she does not exceed her budget constraint, i.e. $p_i\pi_i + p_j\pi_j \leq m$

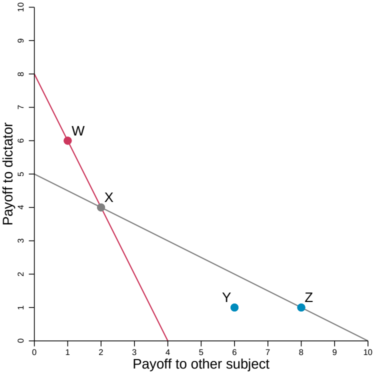
A Modified Dictator Game: Testing Rationality

By choosing an allocation on the budget line, the dictator reveals a preference for that allocation (relative to other feasible distributions)

Directly Revealed Preferred



Indirectly Revealed Preferred



A Modified Dictator Game: Testing Rationality

Rationality $\Rightarrow$ revealing a preference for a bundle is equivalent to demonstrating that it gives you greater utility than the alternatives

A Modified Dictator Game: Testing Rationality

Rationality $\Rightarrow$ revealing a preference for a bundle is equivalent to demonstrating that it gives you greater utility than the alternatives

π is **indirectly revealed preferred** to π' whenever there is some sequence of bundles chosen by i — $\pi^0, \pi^1, \ldots, \pi^{n-1}, \pi^n$ — so that

$$
\begin{aligned}
p_i \pi_i + p_j \pi_j \geq p_i \pi_i^0 + p_j \pi_j^0 \ &\rightarrow \ \pi \text{ is directly revealed preferred to } \pi^0 \\
\text{AND } p_i^0 \pi_i^0 + p_j^0 \pi_j^0 \geq p_i^0 \pi_i^1 + p_j^0 \pi_j^1 \ &\rightarrow \ \pi^0 \text{ is directly revealed preferred to } \pi^1 \\
\ldots \\
\text{AND } p_i^n \pi_i^n + p_j^n \pi_j^n \geq p_i^n \pi_i' + p_j^n \pi_j' \ &\rightarrow \ \pi^n \text{ is directly revealed preferred to } \pi'
\end{aligned}
$$

A Modified Dictator Game: Testing Rationality

Rationality $\Rightarrow$ revealing a preference for a bundle is equivalent to demonstrating that it gives you greater utility than the alternatives

π is **indirectly revealed preferred** to π' whenever there is some sequence of bundles chosen by i — $\pi^0, \pi^1, \ldots, \pi^{n-1}, \pi^n$ — so that

$$\begin{aligned}
p_i\pi_i + p_j\pi_j \geq p_i\pi_i^0 + p_j\pi_j^0 \ &\rightarrow\ \pi \text{ is directly revealed preferred to } \pi^0 \\
\text{AND } p_i^0\pi_i^0 + p_j^0\pi_j^0 \geq p_i^0\pi_i^1 + p_j^0\pi_j^1 \ &\rightarrow\ \pi^0 \text{ is directly revealed preferred to } \pi^1 \\
\ldots \\
\text{AND } p_i^n\pi_i^n + p_j^n\pi_j^n \geq p_i^n\pi_i' + p_j^n\pi_j' \ &\rightarrow\ \pi^n \text{ is directly revealed preferred to } \pi'
\end{aligned}$$

If preferences are rational, this would imply:

$$u(\pi_i, \pi_j) \geq u(\pi_i^0, \pi_j^0) \geq \ldots \geq u(\pi_i^n, \pi_j^n) \geq u(\pi_i', \pi_j')$$

A Modified Dictator Game: Testing Rationality

Distributional preferences satisfy **GARP** when the following is true:

- If (π_i, π_j) is indirectly revealed preferred to (π'_i, π'_j), then (π'_i, π'_j) is **not** directly revealed strictly preferred to (π_i, π_j)
- Intuitively, it can't be the case that both of the following are true:

$$u(\pi_i, \pi_j) \geq u(\pi'_i, \pi'_j)$$
$$u(\pi'_i, \pi'_j) > u(\pi_i, \pi_j)$$

A Modified Dictator Game: Testing Rationality

Distributional preferences satisfy **GARP** when the following is true:

- If (π_i, π_j) is indirectly revealed preferred to (π_i', π_j'), then (π_i', π_j') is **not** directly revealed strictly preferred to (π_i, π_j)
- Intuitively, it can't be the case that both of the following are true:

$$u(\pi_i, \pi_j) \geq u(\pi_i', \pi_j')$$
$$u(\pi_i', \pi_j') > u(\pi_i, \pi_j)$$

Afriat's Theorem: the following conditions are equivalent:

- The data satisfy GARP
- There exists a well-behaved (i.e. concave, monotonic, continuous, non-satiated) utility function that rationalizes the data

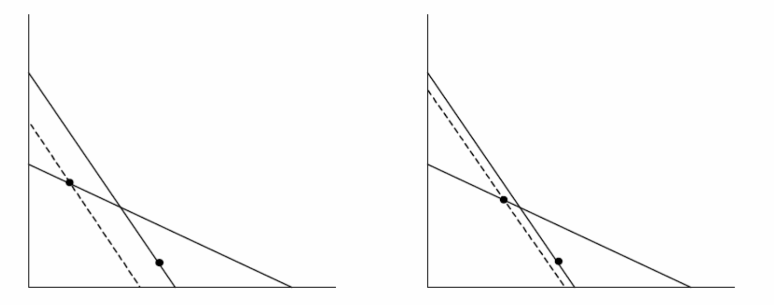
A Modified Dictator Game: Testing Rationality

Afriat (1972) proposes a measure of proximity to satisfying GARP:

A Modified Dictator Game: Testing Rationality

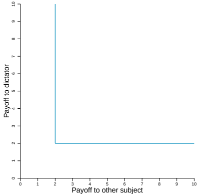
Afriat (1972) proposes a measure of proximity to satisfying GARP:

Critical Cost Efficiency Index (CCEI): the amount by which each budget constraint must be relaxed in order to remove all violations of GARP

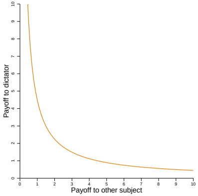


Equality-Efficiency Tradeoffs

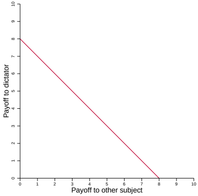
Rawlsian



Cobb-Douglas

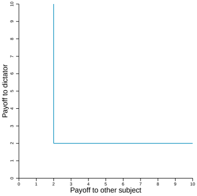


Utilitarian

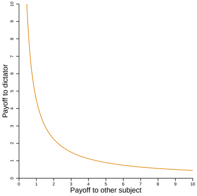


Equality-Efficiency Tradeoffs

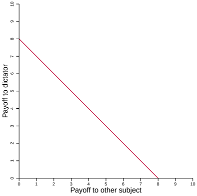
Rawlsian



Cobb-Douglas



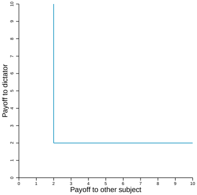
Utilitarian



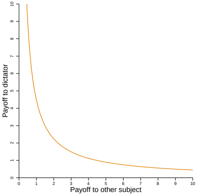
Price changes allow us to characterize equality-efficiency tradeoffs

Equality-Efficiency Tradeoffs

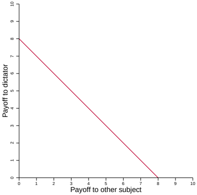
Rawlsian



Cobb-Douglas



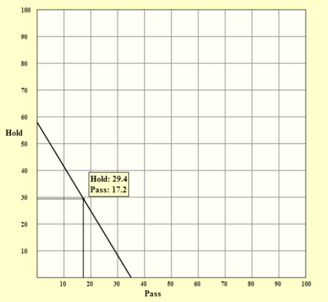
Utilitarian

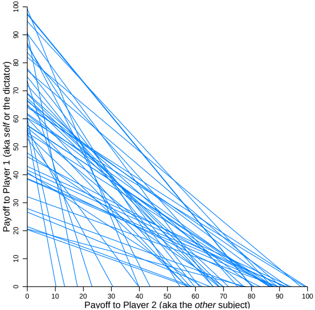


Price changes allow us to characterize equality-efficiency tradeoffs

- Decreasing $p_s \pi_s$ when p_s/p_o increases indicates preferences weighted towards efficiency (in terms of increasing total payoffs)
- Increasing $p_s \pi_s$ when p_s/p_o increases indicates preferences weighted towards equality (in terms of reducing differences in payoffs)

A Modified Dictator Game





Graphical interface allows research to collect many decisions per subjects

- Choose (π_s, π_o) subject to budget constraint $\pi_s + p_o\pi_o = m$

A Modified Dictator Game

We embed interface in the American Life Panel (ALP):

- 687 American adults complete experiment in 2013 and 2016
- Each matched with ALP respondent not sampled for experiment

A Modified Dictator Game

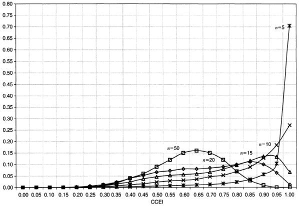
We embed interface in the American Life Panel (ALP):

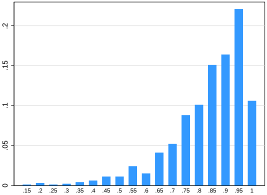
- 687 American adults complete experiment in 2013 and 2016
- Each matched with ALP respondent not sampled for experiment

Implementation:

- Graphical dictator game interface: subject chooses a point on a budget line representing set of feasible payoffs to *self* and *other*
- Confront each subject with a large number of decision problems (50)
- Relative price of redistribution varies across decision rounds
  - Budget lines chosen at random
- One round randomly chosen to determine payoffs

Measuring Rationality



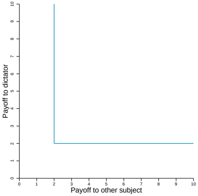


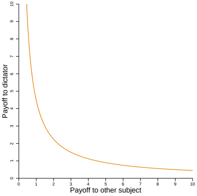
Experimental design also allows us to measure economic rationality

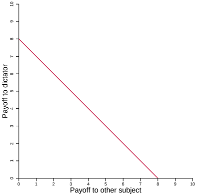
- Almost all subjects violate GARP (more so than students)
- Subjects' choices demonstrate a high degree of consistency

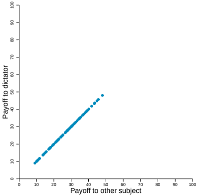
Distributional Preferences: Examples

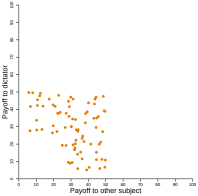
Rawlsian	Cobb-Douglas	Utilitarian
$u(\pi_s, \pi_o) = \min\{\pi_s, \pi_o\}$	$u(\pi_s, \pi_o) = \ln(\pi_s) + \ln(\pi_o)$	$u(\pi_s, \pi_o) = \pi_s + \pi_o$

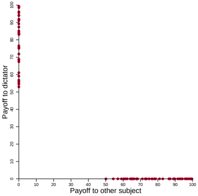












The CES Utility Function

Estimate CES other-regarding utility function at the subject level:

$$u_s(\pi_s, \pi_o) = [\alpha(\pi_s)^\rho + (1-\alpha)(\pi_o)^\rho]^{1/\rho}$$

Generates individual CES parameter estimates for every subject n:

- $\hat{\alpha}_n$: fair-mindedness/selfishness, weight on payoff to *self* vs. *other*
- $\hat{\rho}_n$: curvature of altruistic indifference curves, measures willingness to trade off equality and efficiency (aggregate payoff)

The CES Utility Function

Estimate CES other-regarding utility function at the subject level:

$$u_s(\pi_s, \pi_o) = [\alpha(\pi_s)^\rho + (1-\alpha)(\pi_o)^\rho]^{1/\rho}$$

Generates individual CES parameter estimates for every subject n:

- $\hat{\alpha}_n$: fair-mindedness/selfishness, weight on payoff to *self* vs. *other*
- $\hat{\rho}_n$: curvature of altruistic indifference curves, measures willingness to trade off equality and efficiency (aggregate payoff)

CES utility function spans a range of preference types

- Approaches utilitarian indifference curves as $\rho \to 1$
- Approaches maximin indifference curves as $\rho \to -\infty$

Estimating Individual CES Parameters

CES expenditure function is given by:

$$\frac{\pi_s}{m} = \frac{\left(\frac{\alpha}{1-\alpha}\right)^{1/(1-\rho)}}{(p_o)^{\rho/(\rho-1)} + \left(\frac{\alpha}{1-\alpha}\right)^{1/(1-\rho)}}$$

Individual-level econometric specification for each subject n:

$$\frac{\pi_{s,n,i}}{m_i} = \frac{\left(\frac{\alpha_n}{1-\alpha_n}\right)^{1/(1-\rho_n)}}{(p_{o,n,i})^{\rho_n/(\rho_n-1)} + \left(\frac{\alpha_n}{1-\alpha_n}\right)^{1/(1-\rho_n)}} + \epsilon_{n,i}$$

where $i = 1, ..., 50$ and $\epsilon_{n,i}$ is iid normal with mean zero and variance σ_n^2

Classifying Distributional Preference Types

Fair-mindedness vs. selfishness:

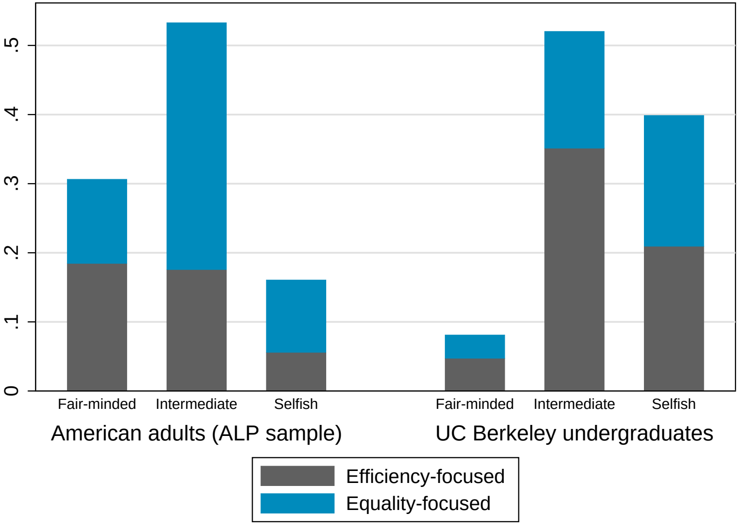
- We classify a subject as **fair-minded** if $0.45 < \hat{\alpha}_n < 0.55$
- We classify a subject as **selfish** if $\hat{\alpha}_n > 0.95$

Equality-efficiency tradeoffs:

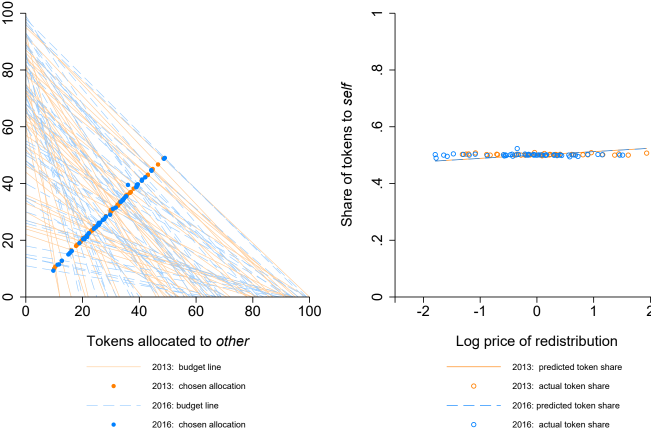
- We classify a subject as **efficiency-focused** if $\hat{\rho}_n > 0$
- We classify a subject as **equality-focused** if $\hat{\rho}_n < 0$

We compare ALP subjects' preference parameters to those of UC Berkeley students who participated in identical DG experiments in 2004 and 2011

Classifying Distributional Preference Types

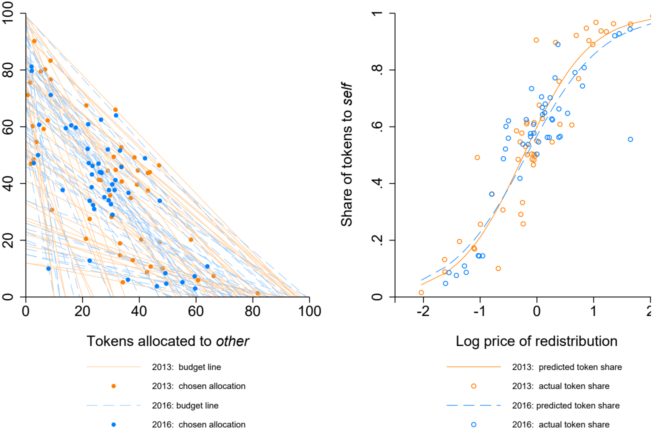


Characterizing Behavior in 2013 and 2016



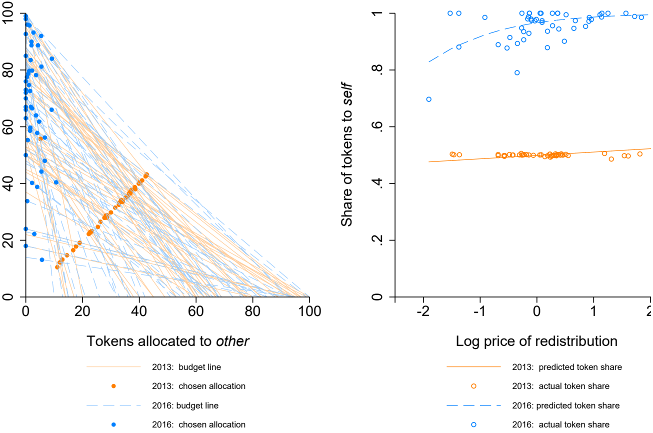
A few subjects implement their choices consistently, without errors

Characterizing Behavior in 2013 and 2016



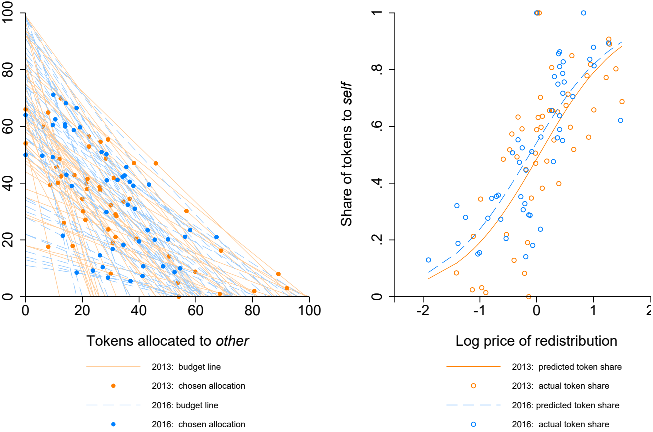
Many subjects are stable and consistent, but make small errors

Characterizing Behavior in 2013 and 2016



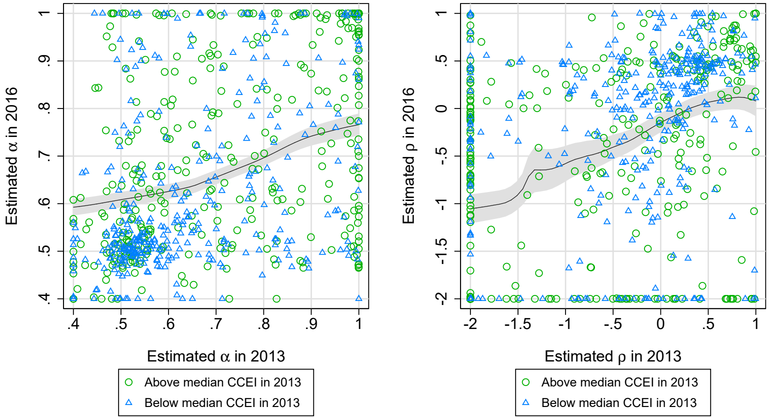
In other cases, preferences are near-rational but not stable over time

Characterizing Behavior in 2013 and 2016



Or not-so-near-rational, but quite stable over time

Preference Stability



Subjects with low CCEI scores display a high degree of preference stability

Measuring Risk Preferences in Household Surveys

Experimental Design

A day in the life of a lab-in-the-field experimentalist:

"Hi Pam!"

"I'm surveying 5,000 [entrepreneurs/ex-combatants/tobacco farmers] in [Bangladesh/Ethiopia/Kenya/Malawi] next month. I'd like to measure their [social/risk/time] preferences, but I can't use [monetary] incentives and any questions I add to the survey can only take 43 seconds."

"What should I do?"

Experimental Design

A multiple price list (Holt & Laury 2002)

- Series of choices between high and low risk lotteries
- Probability of good state increases across decisions

A lottery menu task (Binswanger 1980, Harrison et al. 2009)

- Subjects indicate which of 2, 3, or 6 lotteries they prefer
- Simple probability structure, substantial scope for inconsistency

A portfolio choice task (Gneezy & Potters 1997, Jakiela & Ozier 2016)

- Subject decides how much to invest in a risky asset
- Implicit or explicit choice from a convex choice set

Experimental Design

Modules embedded in a standard household survey

- Administered one-on-one by (16) trained enumerators

Subjects randomly sampled from adult population of 16 communities

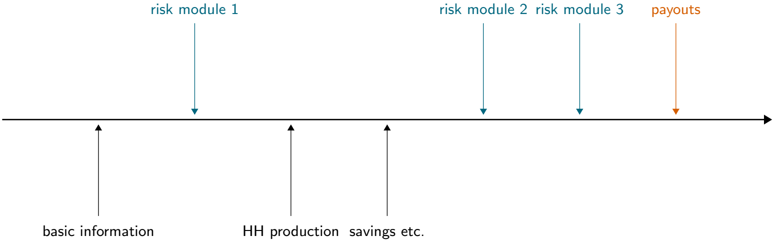
- Catchment areas of rural primary schools in (what are now) Bungoma, Busia, and Kakamega Counties in western Kenya
- Sample drawn after conducting Google-guided census of each village
- Enumerators sought out target survey respondents in random order, but only those present in village on day of visit were interviewed

Sample contains data on 648 respondents; 608 completed risk modules

- More educated, younger respondents more likely to complete survey
- Some incomplete surveys due to data entry errors

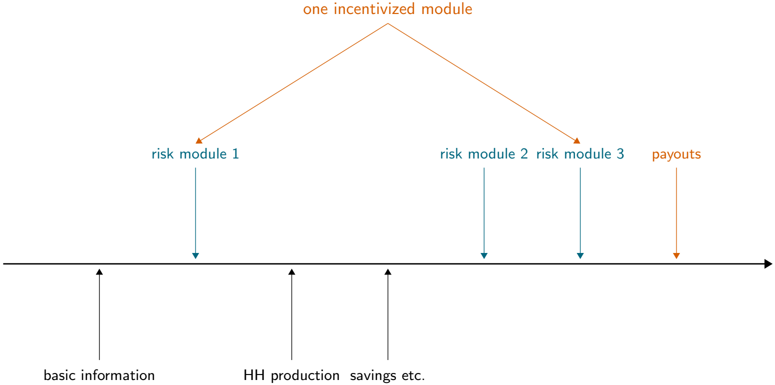
Experimental Design

Three risk preference modules appear at different points in the survey
One module determines monetary payoffs, the other two are hypothetical



Experimental Design

Three risk preference modules appear at different points in the survey
One module determines monetary payoffs, the other two are hypothetical



Experimental Design

Incentivized module:
Now, I am going to ask you to make some choices. The choices are about real money, so I would like you to take the decisions seriously.

Hypothetical module:
Now, I am going to ask you to make some choices. The choices are about money, so although we are not making choices with real money, I would like you to take the decisions seriously, and decide what you would do if real money was at stake.

All modules:
These choices are to find out about what you like the most. There are no right or wrong answers. The information will be used by researchers here in Kenya and in the United States — for us to understand better how people in this area think and behave.

Experimental Design

T1	T2	T3	T4	T5	T6
CUPS	CUPS	MPL	MPL	MENU	MENU
MENU	MENU	CUPS	CUPS	MPL	MPL
MPL	MPL	MENU	MENU	CUPS	CUPS

Vary which module comes first, which module is incentivized

- MPL: Holt-Laury (2002) experiment implemented using visual aids
- MENU: adapted sequence of choices implemented using visual aids
- CUPS: portfolio choice problem implemented using coins and cups

Starting points in survey sequence randomized across enumerator-days

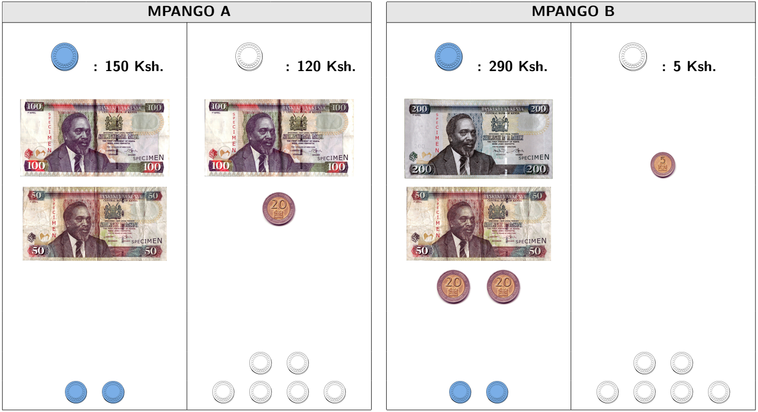
Implementing the MPL Task

Lottery A	Lottery B	EV Diff.	Implied CRRA
1/8 of 150, 7/8 of 120	1/8 of 290, 7/8 of 5	83	$\rho < -1.46$
2/8 of 150, 6/8 of 120	2/8 of 290, 6/8 of 5	51	$-1.46 < \rho < -0.69$
3/8 of 150, 5/8 of 120	3/8 of 290, 5/8 of 5	19	$-0.69 < \rho < -0.22$
4/8 of 150, 4/8 of 120	4/8 of 290, 4/8 of 5	−12	$-0.22 < \rho < 0.13$
5/8 of 150, 3/8 of 120	5/8 of 290, 3/8 of 5	−44	$0.13 < \rho < 0.44$
6/8 of 150, 2/8 of 120	6/8 of 290, 6/8 of 5	−76	$0.44 < \rho < 0.76$
7/8 of 150, 1/8 of 120	7/8 of 290, 1/8 of 5	−108	$0.76 < \rho < 1.16$
8/8 of 150, 0/8 of 120	8/8 of 290, 0/8 of 5	−140	$1.16 < \rho$

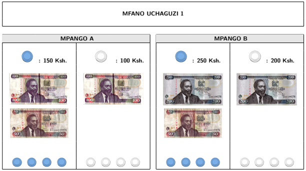
Adapted from Holt & Laury (2002)

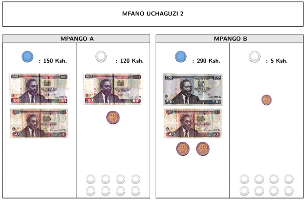
Implementing the MPL Task

Each lottery in the MPL was represented on a laminated card:



Implementing the MPL Task





Complicated probabilities explained intuitively using colored tokens

- 94 percent of subjects understood simple comprehension questions
- 92 percent named the payoffs correctly in example decision problems
- 88 percent chose "correct" lottery in example decision problems

Implementing the MENU Task

MPANGO A	
: 100 Ksh.	: 100 Ksh.

MPANGO D	
: 60 Ksh.	: 300 Ksh.

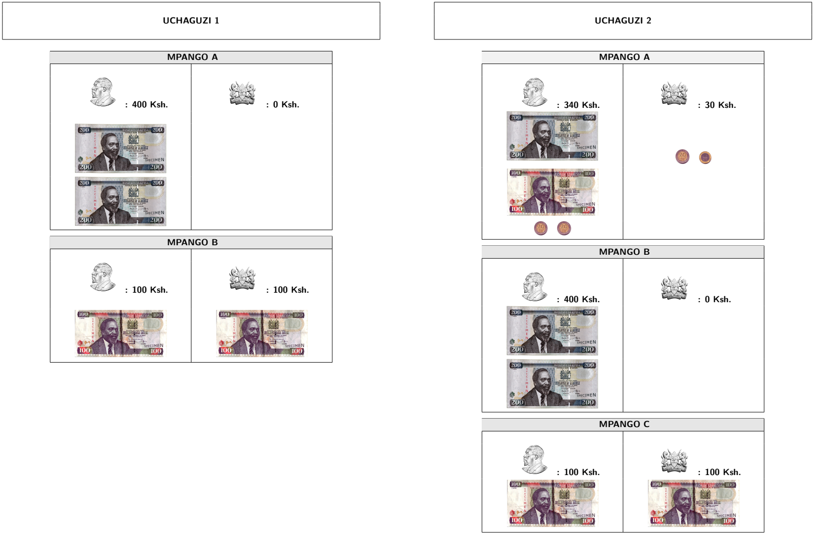
MPANGO B	
: 90 Ksh.	: 180 Ksh.

MPANGO E	
: 40 Ksh.	: 320 Ksh.

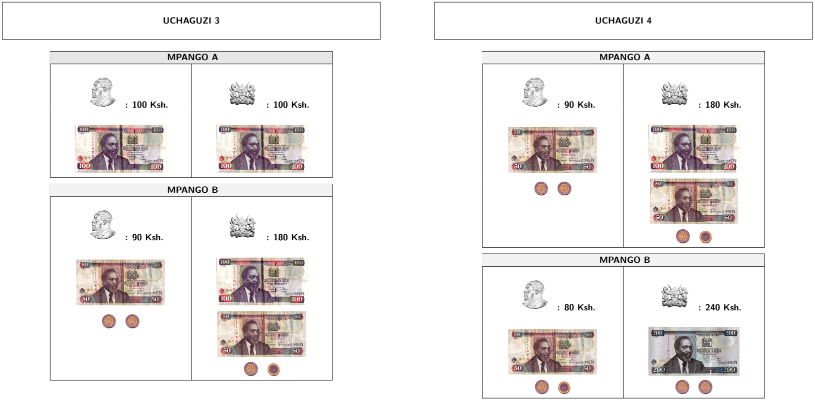
MPANGO C	
: 80 Ksh.	: 240 Ksh.

MPANGO F	
: 20 Ksh.	: 380 Ksh.

Implementing the MENU Task



Implementing the MENU Task



Implementing the MENU Task

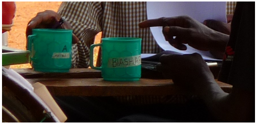
Sequence of 8 decision problems builds in complexity over time

- Practice problems gauge comprehension of the task
- Builds in opportunities for violations of consistency
  - Are binary choices consistent with choice in final decision problem?
  - Are choices consistent with a CRRA utility representation?

CRRA-consistent subjects fall into one of 10 categories

- Inconsistent subjects classified based on frequency of "risky" choice
- Approach validated in Jakiela & Ozier (2019)

Implementing the CUPS Task



Adaptation of standard portfolio choice problem (Jakiela & Ozier 2016)

- Subject divides endowment of 8 coins between two identical cups
  - Savings cup: amount saved is guaranteed (zero risk)
  - Business cup: 50 percent chance investment is multiplied by five
- Illustrated card explains all possible high/low payoffs
- Subject flips a coin to determine whether investment is successful

Implementing the CUPS Task

Investment	High Payoff	Low Payoff	CRRA Coefficients
80	400	0	$\rho < 0.28$
70	360	10	$0.28 < \rho < 0.44$
60	320	20	$0.44 < \rho < 0.55$
50	280	30	$0.55 < \rho < 0.68$
40	240	40	$0.68 < \rho < 0.87$
30	200	50	$0.87 < \rho < 1.17$
20	160	60	$1.17 < \rho < 1.82$
10	120	70	$1.82 < \rho < 5.58$
0	80	80	$5.58 < \rho$

Theoretical Framework

What enters into the objective function?

- Expected utility of lottery (when lotteries are incentivized)
- Intrinsic motivation to answer enumerators' questions
- Cognitive effort may be costly (or pleasurable)

Theoretical Framework

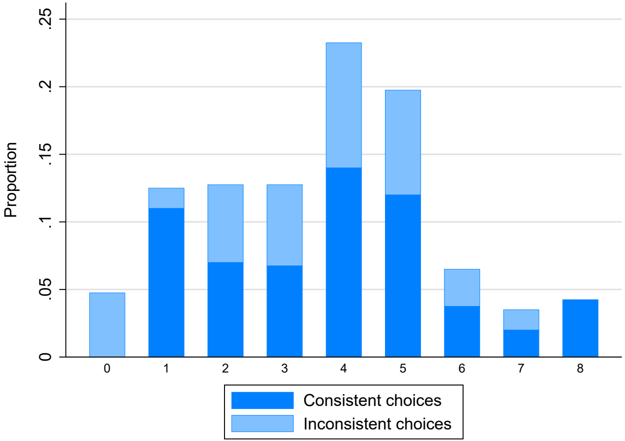
What enters into the objective function?

- Expected utility of lottery (when lotteries are incentivized)
- Intrinsic motivation to answer enumerators' questions
- Cognitive effort may be costly (or pleasurable)

Research questions:

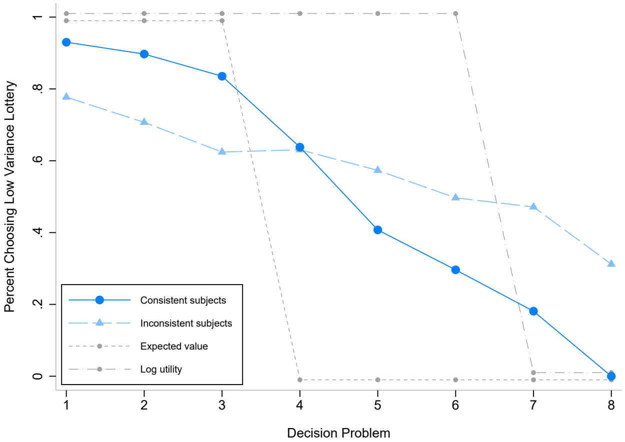
- How well do measures "work" when embedded in household surveys?
  - Do subjects understand the questions? Are responses consistent?
  - Are measures correlated with each other?
- How much do incentives matter?
  - For "harder" approaches? For consistency and comprehension?

Decisions in the MPL Task



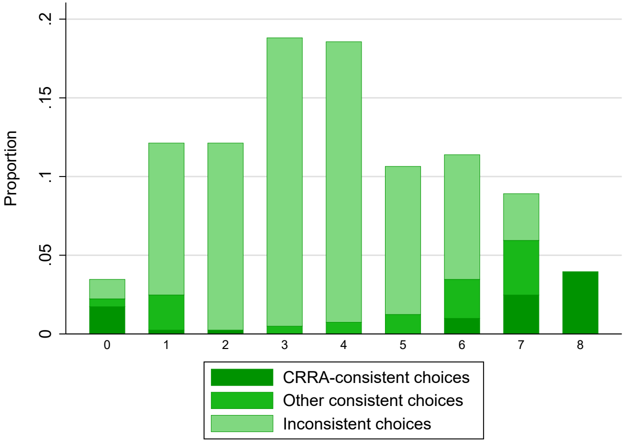
Data from non-incentivized decisions of 412 subjects. 88 percent of subjects answer all 6 comprehension questions correctly. Subjects who answered the comprehension questions correctly choose the risky lottery 3.6 times on average, while those who do not answer all the comprehension questions correctly choose the risky lottery an average of 4.3 times (p-value 0.013).

Decisions in the MPL Task



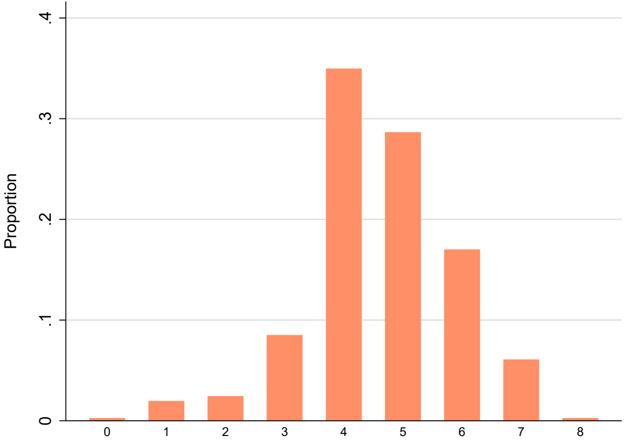
Data from non-incentivized decisions of 412 subjects.

Decisions in the MENU Task



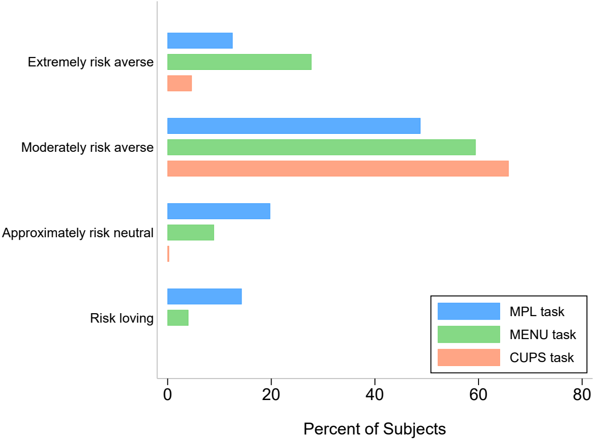
Data from non-incentivized decisions of 404 subjects. 84 percent of subjects answer both comprehension questions correctly. Subjects who answered the comprehension questions correctly choose the riskiest lottery 4.0 times on average, while those who do not answer all the comprehension questions correctly choose the riskiest lottery an average of 2.8 times (p-value < 0.001).

Decisions in the CUPS Task



Data from non-incentivized decisions of 412 subjects. 66 percent of subjects answer all four comprehension questions correctly. Subjects who do and do not answer the comprehension questions correctly both invest an average of 46 shillings in the risky cup (p-value 0.512).

Contradictions Between the Measures



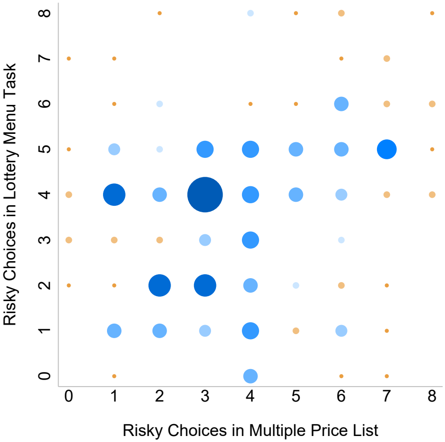
Correlations Between the Measures

	Hypothetical Tasks			Incentivized Tasks		
	MPL	MENU	CUPS	MPL	MENU	CUPS
Hypothetical MPL	·	**0.203**	**0.130**	·	**0.220**	0.094
Hypothetical MENU		·	0.013	0.076	·	−0.026
Hypothetical CUPS			·	0.040	**0.149**	·

Spearman correlations reported. **Bold text** indicates statistical significance at the 90 percent confidence level. **Red text** indicates statistical significance at the 99 percent confidence level.

Correlations are larger, more statistically significant among subjects who answered all relevant comprehension questions correctly

Correlations Between the Measures



Orange (resp. blue) dots indicate frequencies below (resp. above) what would be expected if subjects randomized across all options in each choice set.

Noisy Measures of Risk Tolerance

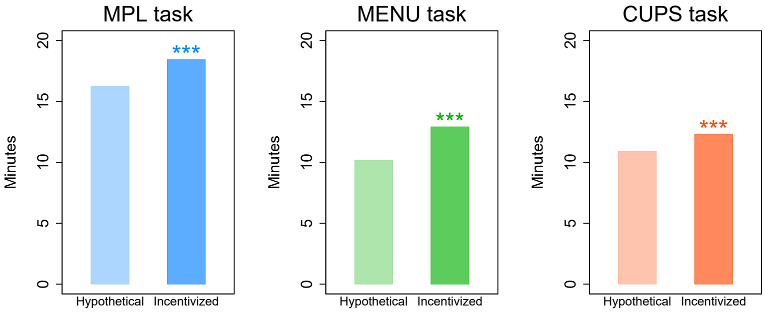
Hypothetical measures are noisy, but contain information

- Screening out inconsistent subjects biases estimates of risk aversion
- The direction of bias varies across elicitation techniques

How do incentives impact decision quality?

- Incentives tend to increase risk aversion (Camerer & Hogarth 1999)
- May also reduce the variance of responses (Smith & Walker 1993)
- Do incentives induce cognitive effort and increase comprehension?
- Do they reduce the frequency of inconsistent response patterns?

The Impacts of Incentives: Time Taken



Results:

Incentivized risk modules take between 10 and 25 percent longer

The Impacts of Incentives: Time Taken

Estimate the OLS regression:

$$Y_{im} = \alpha + \beta Incent_{im} + \gamma Later_{im} + \delta X_{im} + \varepsilon_{im}$$

where:

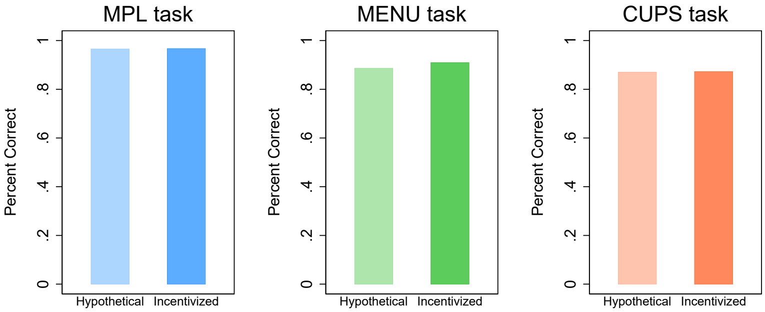
- Y_i = outcome of interest
- $Incent_{im}$ = indicator for incentivized decision tasks
- $Later_{im}$ = indicator for appearing near end of survey
- X_{im} = vector of controls (gender, schooling, math skills, enumerator)
- ε_{im} = conditionally mean zero error term

The Impacts of Incentives: Time Taken

	– MPL Task –		– MENU Task –		– CUPS Task –	
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	1.511	2.576	2.327	1.931	1.262	1.187
	(0.644)	(0.519)	(0.442)	(0.368)	(0.520)	(0.470)
	[0.019]	[$<$ 0.001]	[$<$ 0.001]	[$<$ 0.001]	[0.016]	[0.012]
End of survey	-2.880	-1.965	-1.452	-1.594	-0.497	-0.768
	(0.658)	(0.516)	(0.482)	(0.403)	(0.491)	(0.465)
	[$<$ 0.001]	[$<$ 0.001]	[0.003]	[$<$ 0.001]	[0.312]	[0.099]
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses; p-values in square brackets. Controls included in even-numbered columns: gender, education level, math skills index, enumerator fixed effects. The dependent variable is the length of the survey module (in minutes).

The Impacts of Incentives: Comprehension



Results:

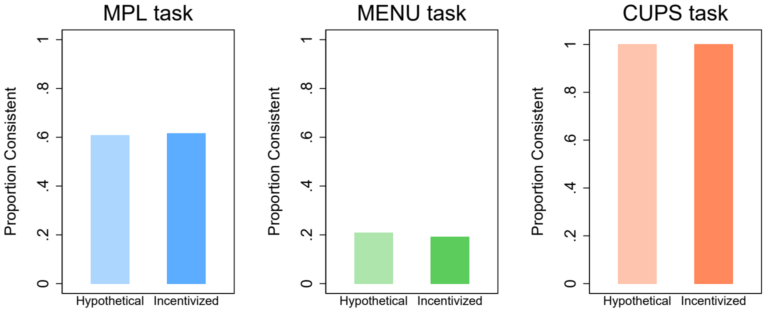
Incentivized risk modules take between 10 and 25 percent longer, but they don't impact objective measures of comprehension*

The Impacts of Incentives: Comprehension

	– MPL Task –		– MENU Task –		– CUPS Task –	
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	0.007	0.009	0.045	0.044	-0.001	0.011
	(0.009)	(0.009)	(0.022)	(0.022)	(0.018)	(0.016)
	[0.409]	[0.304]	[0.043]	[0.047]	[0.948]	[0.480]
End of survey	0.023	0.022	0.080	0.083	-0.015	-0.017
	(0.011)	(0.011)	(0.025)	(0.025)	(0.018)	(0.016)
	[0.033]	[0.036]	[0.002]	[0.001]	[0.419]	[0.313]
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses; p-values in square brackets. Controls included in even-numbered columns: gender, education level, math skills index, enumerator fixed effects. The dependent variable is the proportion of task-specific comprehension questions answered correctly (ranging from 0 to 1).

The Impacts of Incentives: Consistency



Results:

Incentivized risk modules take between 10 and 25 percent longer, but they don't impact objective measures of comprehension…much

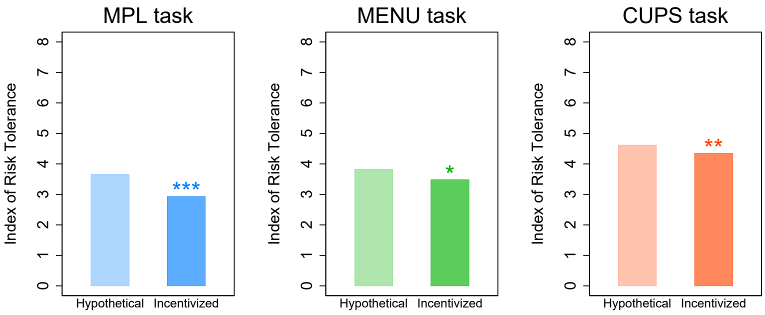
Incentives do not improve decision consistency by any measure

The Impacts of Incentives: Consistency

	Outcomes for MENU Task					
	– MPL Task –		– Consistency –		– CRRA –	
	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	0.007	0.003	-0.013	-0.024	-0.010	-0.017
	(0.043)	(0.042)	(0.036)	(0.037)	(0.026)	(0.027)
	[0.879]	[0.941]	[0.725]	[0.513]	[0.710]	[0.539]
End of survey	-0.006	0.020	0.015	0.011	0.031	0.022
	(0.043)	(0.043)	(0.036)	(0.036)	(0.025)	(0.025)
	[0.899]	[0.643]	[0.683]	[0.754]	[0.224]	[0.377]
Controls	No	Yes	No	Yes	No	Yes

Standard errors in parentheses; p-values in square brackets. Controls included in even-numbered columns: gender, education level, math skills index, enumerator fixed effects. The dependent variable in Columns 1 and 2 is an indicator for having a maximum of one switch point in the MPL task. The dependent variable in Columns 3 and 4 is an indicator for making the same choices when presented with the same lottery pairs in the MENU task. The dependent variable in Columns 5 and 6 is an indicator for making choices consistent with a CRRA utility representation in the MENU task.

The Impacts of Incentives: Risk Aversion



Results:

Incentivized risk modules take between 10 and 25 percent longer, but they don't impact objective measures of comprehension. . . much

Incentives do not improve decision consistency by any measure, but they increase risk aversion substantially

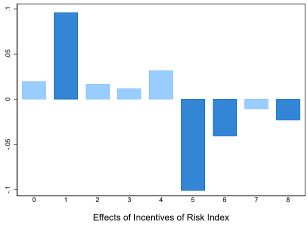
The Impacts of Incentives: Risk Aversion

	– MPL Task –		– MENU Task –		– CUPS Task –	
	OLS	**OLS**	**OLS**	**OLS**	**OLS**	**OLS**
	(1)	(2)	(3)	(4)	(5)	(6)
Incentives	-0.821	-0.793	-0.341	-0.346	-0.277	-0.266
	(0.162)	(0.160)	(0.183)	(0.184)	(0.112)	(0.108)
	[< 0.001]	[< 0.001]	[0.064]	[0.060]	[0.013]	[0.014]
End of survey	-0.402	-0.315	0.002	-0.002	-0.031	-0.054
	(0.172)	(0.164)	(0.182)	(0.174)	(0.111)	(0.111)
	[0.019]	[0.055]	[0.990]	[0.992]	[0.779]	[0.625]
Controls	No	Yes	No	Yes	No	Yes

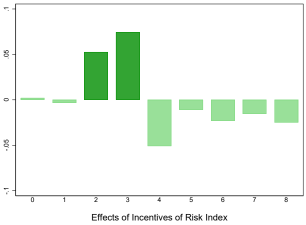
Standard errors in parentheses; p-values in square brackets. Controls included in even-numbered columns: gender, education level, math skills index, enumerator fixed effects. The dependent variable is an index of risky choices within the experimental task (ranging from 0 to 8).

The Impacts of Incentives: Risk Aversion

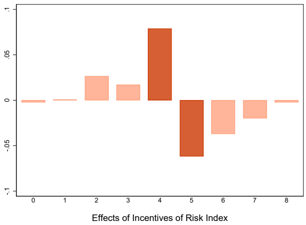
MPL Task



MENU Task

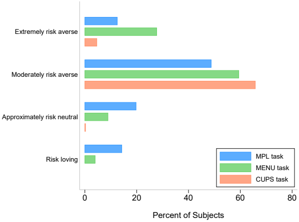


CUPS Task

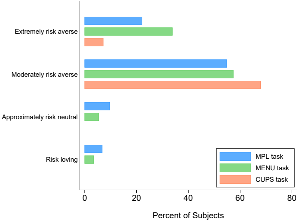


The Impacts of Incentives: Risk Aversion

Without Incentives



With Incentives



Incentives are not a panacea

- Reduce prevalence of risk-loving behavior, increase risk aversion
- Don't eliminate decision errors or reconcile CRRA estimates

Characterizing Decision Errors

Incentives impact risk aversion, but do not eliminate decision errors

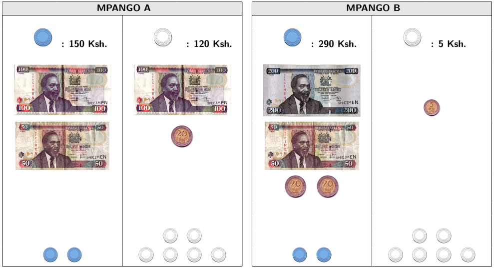
- More decisions — and more types of decisions — are needed
- Models of errors can help to separate signal from noise

Two sources of decision error:

- Trembles (Harless & Camerer 1994, von Gaudecker et al. 2010)
- Additive random utility (Hey & Orme 1994, Loomes 2005)

How well do these models characterize observed choice patterns?

Characterizing Decision Errors in the MPL Task



Discrete choice between **Lottery A** and **Lottery B**

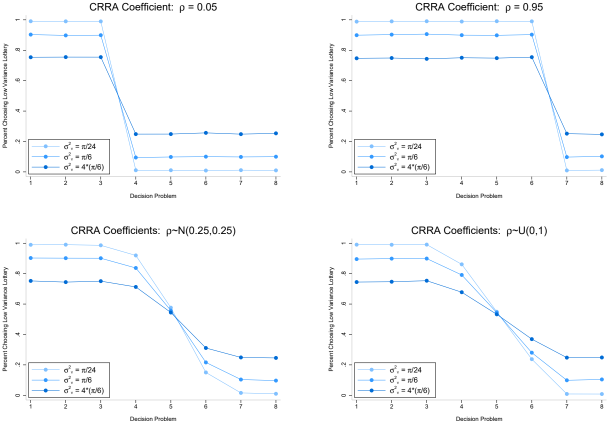
- Each lottery characterized by an expected utility (eg, CRRA)
- Decision errors occur with probability $\nu > 0$
- Choose lottery with higher expected utility with probability $1 - \nu$

Characterizing Decision Errors in the MPL Task

Consistent		One Error
SSSRRRRR	$\Rightarrow$	RSSRRRRR
		SRSRRRRR
		SSRRRRRR
		SSSSRRRR
		SSSRSRRR
		SSSRRSRR
		SSSRRRSR
		SSSRRRRS

Each consistent choice sequence associated with 8 equally likely patterns that each involve exactly one decision error; other patterns less common

Characterizing Decision Errors in the MPL Task



Proportion choosing risky (safe) lottery approaches ν $(1-\nu)$ nears ends

$\Rightarrow$ Mistakes are not more likely near middle of choice sequence

Characterizing Decision Errors in the MPL Task

Additive random utility model of lottery choices in experiment:

$$U_{nj} = EU_{nj} + \varepsilon_{nj}$$

Characterizing Decision Errors in the MPL Task

Additive random utility model of lottery choices in experiment:

$$U_{nj} = EU_{nj} + \varepsilon_{nj}$$

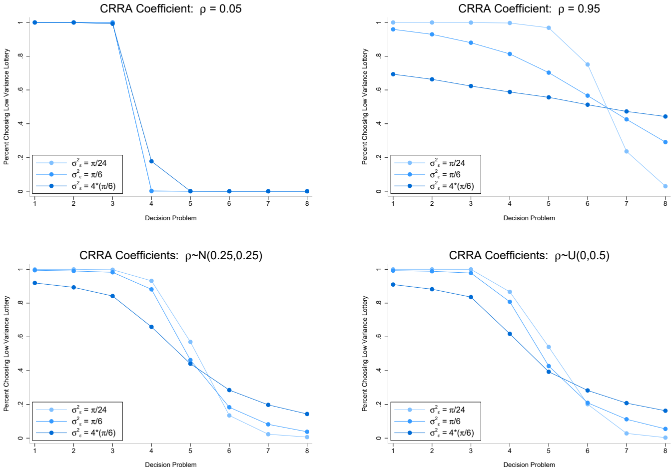
Highest (total) utility alternative is chosen:

$$\begin{aligned} P_{nj} &= \Pr\left[EU_{nj} + \varepsilon_{nj} > EU_{nk} + \varepsilon_{nk} \forall k \neq j\right] \\ &\dots \\ &= \frac{e^{EU_{nj}}}{\sum_{k \in J} e^{EU_{nk}}} \end{aligned}$$

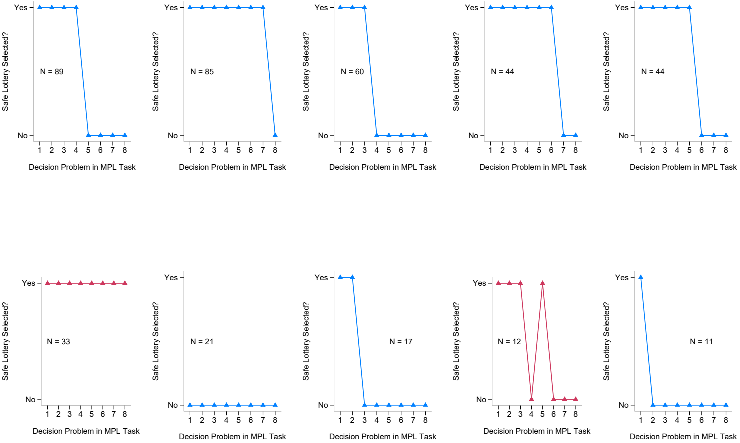
whenever ε_{nj} is EV1-distributed

Lotteries likely to be chosen “by mistake” when $EU_{nj} - EU_{nk}$ is small

Characterizing Decision Errors in the MPL Task

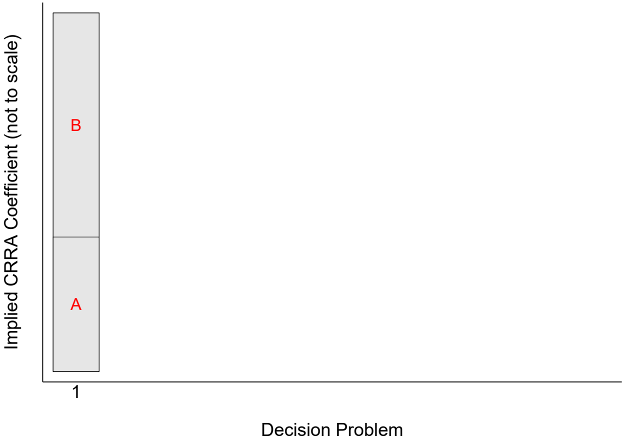


Characterizing Decision Errors in the MPL Task

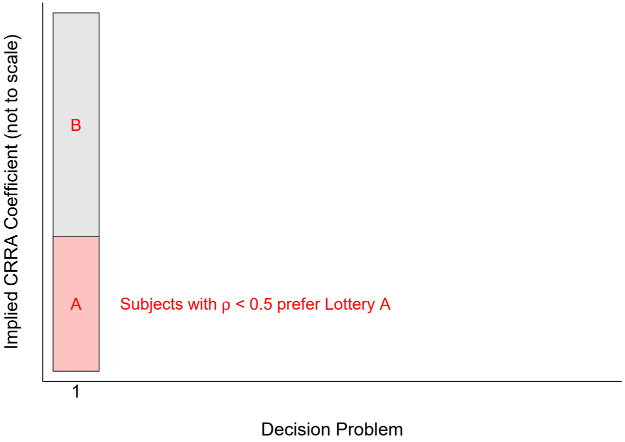


Some patterns (eg. always choosing low-variance lottery) not consistent with either model of decision errors; little support for trembles ($\nu > 0$)

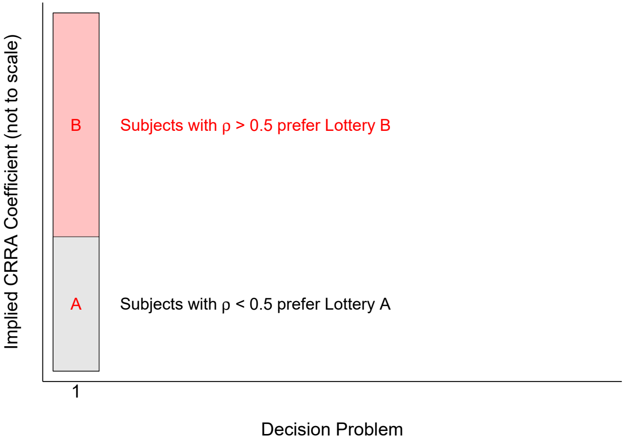
Characterizing Decision Errors in the MENU Task



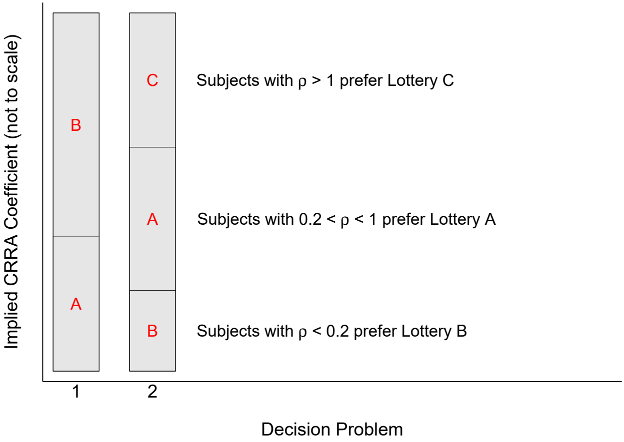
Characterizing Decision Errors in the MENU Task



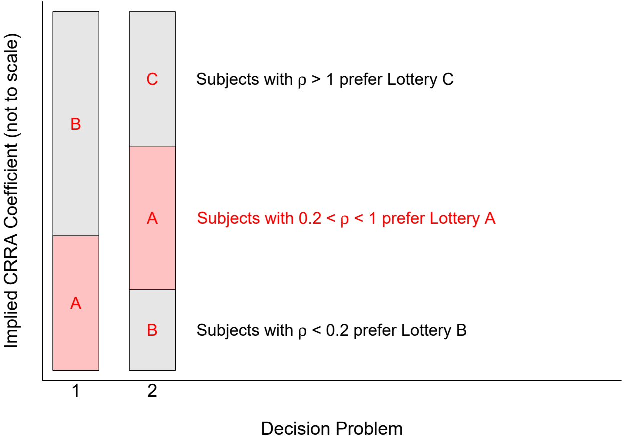
Characterizing Decision Errors in the MENU Task



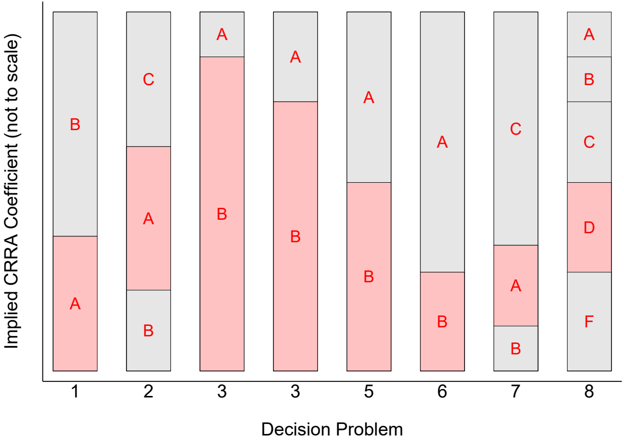
Characterizing Decision Errors in the MENU Task



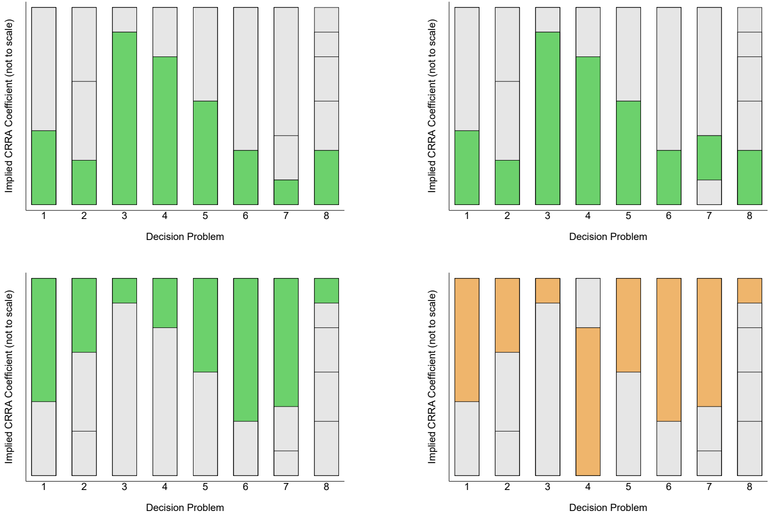
Characterizing Decision Errors in the MENU Task



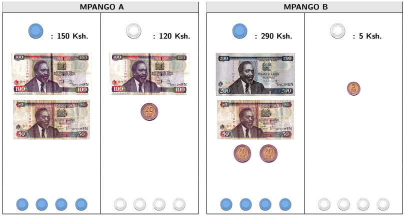
Characterizing Decision Errors in the MENU Task

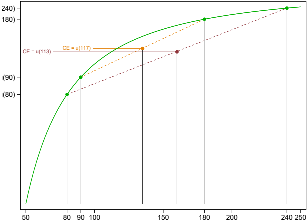


Characterizing Decision Errors in the MENU Task



Characterizing Decision Errors in the MENU Task

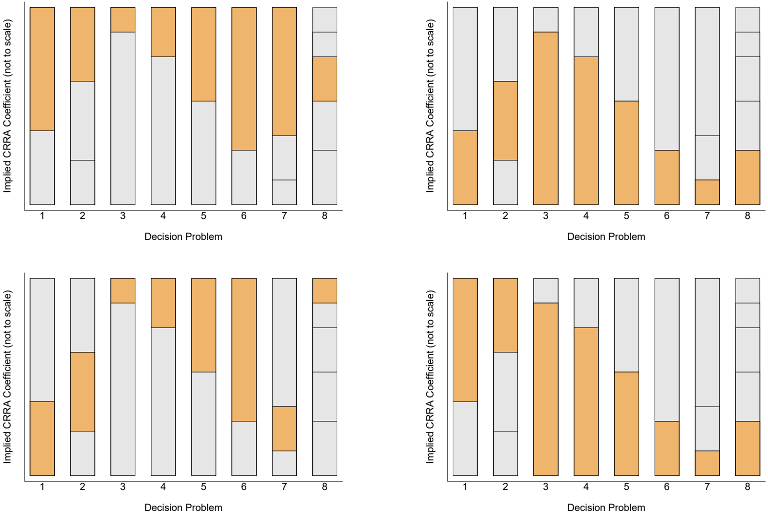




Many "mistakes" are inconsequential from an expected utility perspective

- For subject with CRRA coefficient of -1.5, certainty equivalent of riskier lottery is four shillings lower than that of safer lottery
- EU maximization, CRRA form may be approximately correct

Characterizing Decision Errors in the MENU Task



Characterizing Decision Errors in the MENU Task

Additive random utility model of lottery choices in experiment:

$$\begin{aligned} U_{nj} &= V_{nj} + \varepsilon_{nj} \\ &= \underbrace{EU_{nj} + 1_{\text{placement}} + 1_{\text{max on card}}}_{V_{nj} \;=\; \text{"deterministic utility"}} + \varepsilon_{nj} \end{aligned}$$

Characterizing Decision Errors in the MENU Task

Additive random utility model of lottery choices in experiment:

$$
\begin{aligned}
U_{nj} &= V_{nj} + \varepsilon_{nj} \\
&= \underbrace{EU_{nj} + 1_{\text{placement}} + 1_{\text{max on card}}}_{V_{nj} \;=\; \text{“deterministic utility”}} + \varepsilon_{nj}
\end{aligned}
$$

Highest (total) utility alternative is chosen:

$$
\begin{aligned}
P_{nj} &= \Pr\left[V_{nj} + \varepsilon_{nj} > V_{nk} + \varepsilon_{nk} \forall k \neq j\right] \\
&\ldots \\
&= \frac{e^{V_{nj}}}{\sum_{k \in J} e^{V_{nk}}}
\end{aligned}
$$

whenever ε_{nj} is EV1-distributed

Characterizing Decision Errors in the MENU Task

	(1)	(2)	(3)	(4)
Appears first	0.236	0.281	0.604	0.376
	(0.058)	(0.063)	(0.099)	(0.073)
	[$<$ 0.001]	[$<$ 0.001]	[$<$ 0.001]	[$<$ 0.001]
Highest amount	0.960	1.050	0.699	1.190
	(0.084)	(0.088)	(0.119)	(0.095)
	[$<$ 0.001]	[$<$ 0.001]	[$<$ 0.001]	[$<$ 0.001]
First $\times$ Z		-0.135	-0.438	-0.232
		(0.076)	(0.096)	(0.074)
		[0.077]	[$<$ 0.001]	[0.002]
Highest $\times$ Z		-0.267	0.314	-0.380
		(0.076)	(0.101)	(0.073)
		[$<$ 0.001]	[0.002]	[$<$ 0.001]
Z: Interaction with:	·	Incentives	Comprehension	Education

Conditional logit specifications reported. Standard errors in parentheses; p-values in square brackets. All specifications control for the expected value and standard deviation of each lottery plus an indicator for degenerate lotteries.

Accounting for Decision Errors

Decision errors are not the enemy

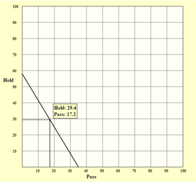
- Small deviations from utility maximization are to be expected
  - Incentives do not make decision errors disappear
  - Certain preference types (i.e. intermediate levels of risk aversion) more prone to errors, screening on consistency introduces bias
- Our objective should be to account for errors in estimation

Elicitation techniques nudge subjects toward different types of consistency

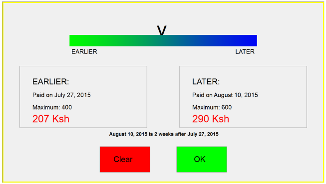
- Consistency does not mean that choices reflect preferences
- Experimental designs should separate preferences from use of procedural choice rules (e.g "always choose A")
- Screening on consistency may select **in** confused subjects

Taking Preferences Seriously

2005



2015



Data collection technology has improved a lot over the past decade

- Computer-assisted surveys/experiments can include large numbers of preference elicitation questions, randomizing at the individual level

No reason experiments in LMICs can't be on technology frontier

Taking Preferences Seriously

Decision errors are a feature, not a bug

- Mistakes provide a window into decision-making process

Experimental design, adaptation, validation can be data driven

- Preference elicitation approaches developed for university labs in WEIRD countries may not work well in field settings in LMICs
- Doesn't mean we have to measure preferences like its 1999

Measuring preferences (well) is difficult

- Experiments can't inform development discourse or policy until we develop reliable, field-tested approaches to preference elicitation

Thank you!