Applied Behavioral and Experimental Economics – Unito 2025

Eliciting risk attitudes: methods, internal & external validity; analysis

Paolo Crosetto

Let's play!

please go to https://paolocrosetto.aidaform.com/expert-template-risk-taking-test

Menu

1. **Risk aversion**: what is it? Why does it matter?
2. **Risk elicitation**: How do we measure risk attitudes?
   - (brief) overview of tasks: psychology
   - Overview of tasks: economics
3. **Does it work? a Meta-analysis**
   - Consistency
     - across tasks
     - within tasks
   - Are the tasks capturing the same thing?
     - Correlation with self-reported measures
     - Correlation with real world behavior
4. **How to fix the problems**
   - treating noise *properly*
   - risk perception
   - time to change the theory?

Measuring risk attitudes

A **difficult** *task with* **crucial** *relevance*

- directly *unobservable*
- *latent* construct ($\Rightarrow$ requires a theory)
- should we..
  - *infer* from real world data or from *ad-hoc* choices
  - ask or **t**ask?
  - elicit by *descrption* or by *experience*?

and by the way, what is our **objective***?*

What are risk attitudes?

Risk in real life

risk noun

\ ˈrisk \

Definition of *risk* (Entry 1 of 2)

1 : possibility of loss or injury : PERIL

2 : someone or something that creates or suggests a hazard

3 **a** : the chance of loss or the perils to the subject matter of an insurance contract

also : the degree of probability of such loss

b : a person or thing that is a specified hazard to an insurer

c : an insurance hazard from a specified cause or source

// war *risk*

4 : the chance that an investment (such as a stock or commodity) will lose value

Taking risks in real life

- are you willing to risk in the face of high but uncertain returns?
- do you engage in activities that could harm your long-run welfare because they provide you with a short-term gain?
- do you look for the thrill of the moment?
- do you invest in bitcoin?

Why measure?

- Risk attitudes are important throughout life
- Very important for policy (risk management, health hazards, insurance...)
- Even mandatory in some fields (finance)
- Might be one of the underlying reasons for different behavior/outcomes of groups/individuals (e.g. gender)

Risk in psychology

The act of implementing a goal-directed option qualifies as an instance of risk taking whenever two things are true: (a) the behavior in question could lead to more than one outcome and (b) some of these outcomes are undesirable or even dangerous. In essence, then, risk taking involves the implementation of options that could lead to negative consequences. (Byrnes et al 1999)

The state of the art: psychology

Risk loosely defined as **probability of harm**

Focus on **questionnaires** *and* **intuitive tasks**

- **Questionnaires**:
  - directly ask
  - over different domains
  - tackle risk perception
- **Tasks**
  - hand in cold water
  - card/gambling tasks

Metrics of success: **convergent validity** + **predictive validity**

Risk in economics

- how do people make decisions given a probability distribution over outcomes? (risk)
- how do people make decisions given a (partially or not known) probability distribution over outcomes? (ambiguity)

Key assumptions:

- Constant over time (preferences are hardwired, *in a sense*)
- Constant across domains.

Definitions

Risk is a situation in which outcomes happen with *known* probabilities; i.e. a situation in which returns are uncertain, *but*:

- all possible outcomes are known;
- a probability distribution across all outcomes is known

Ambiguity is a situation in which all outcomes are known, but the probability distribution across each outcome (or part of it) is not.

Deep (Knightian) Uncertainty is a situation in which neither the full range of outcomes nor probabilities are known. (Keynes): The future is fundamentally uncertain,

Different concepts of uncertainty

Risk

10	100	1000
50%	10%	40%

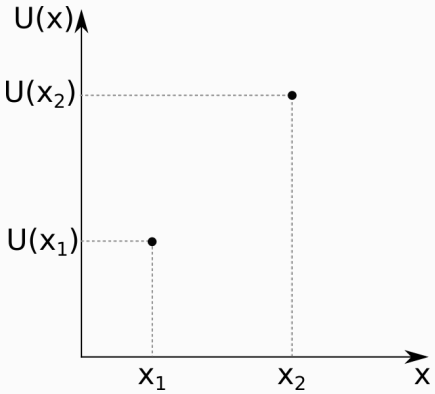
Ambiguity

10	100	1000
50%	50%	

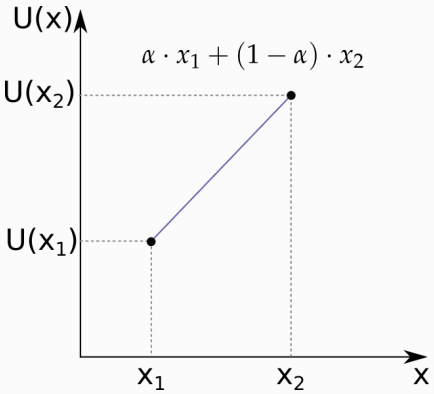
Deep (Knightian) uncertainty

10	??	1000	??	...
50%	30%		??	

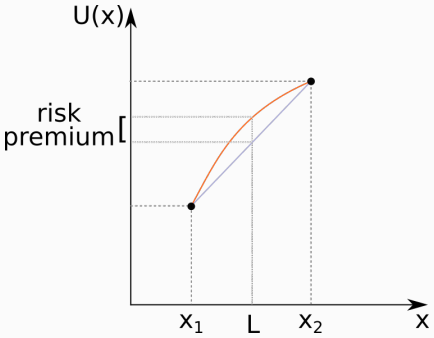
Risk: expected utility framework



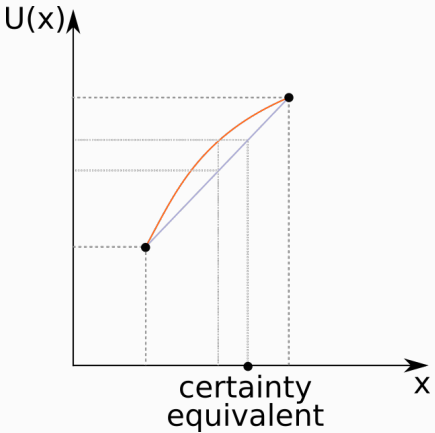
Risk: expected utility framework



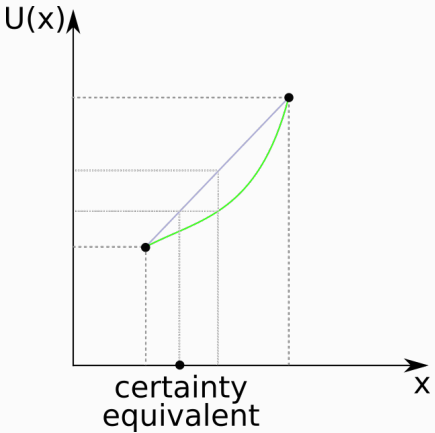
Risk: expected utility framework



Risk: expected utility framework



Risk: expected utility framework



The state of the art: economics

Risk formally defined as **uncertainty over outcomes**

Focus on **decontextualized tasks** *(and questionnaires)*

- **The lottery paradigm**
  - incentives
  - risk task = choice over lotteries
  - different formats, cover stories, contexts
  - strong theoretical underpinning
  - estimation of utility functions ($\Rightarrow$ models)

Metric of success: **internal validity** *(task* $\Longleftrightarrow$ *theory)*

Measuring risk attitudes?

Assumptions needed to yield a measure

Risk attitudes are never observed directly. If we want to *elicit* them (i.e.: make them appear through their consequences), we need them to have some properties.

Existence as a psychological trait

Stability risk prefernces must be stable. This stability could hold

- overall: just *one* risk attitudes for all domains
- over domains: e.g. lots of gambling but no career risks
- always: same risk attitude from cradle to grave
- over reasonable periods: child/young/middle-aged/old

Consistence if asked several times, roughly same answer

Appreciable features of a RET

Accurate
actually representing
true preferences

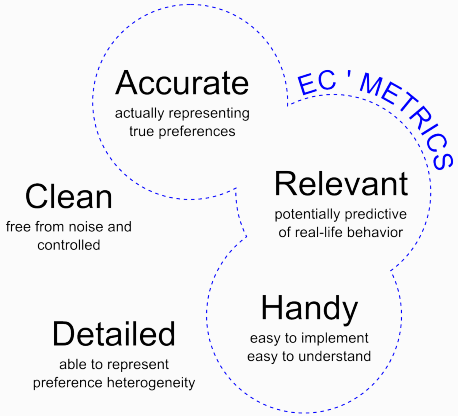
Clean
free from noise and
controlled

Relevant
potentially predictive
of real-life behavior

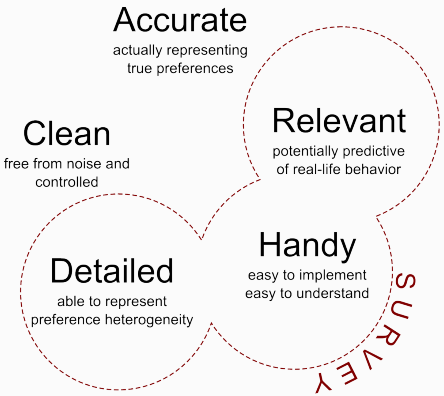
Detailed
able to represent
preference heterogeneity

Handy
easy to implement
easy to understand

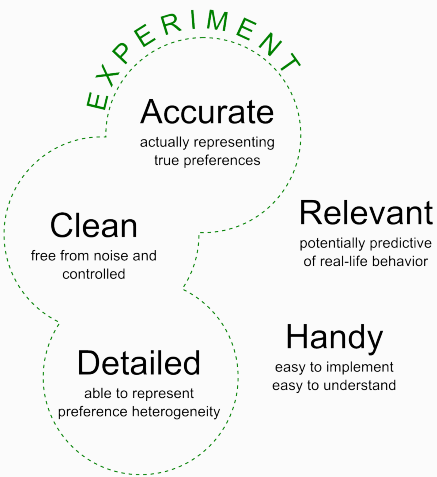
Methods compared



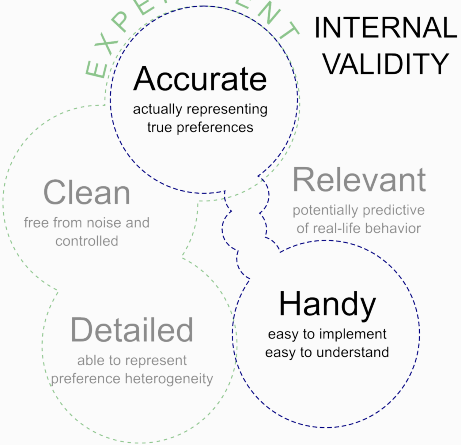
Methods compared



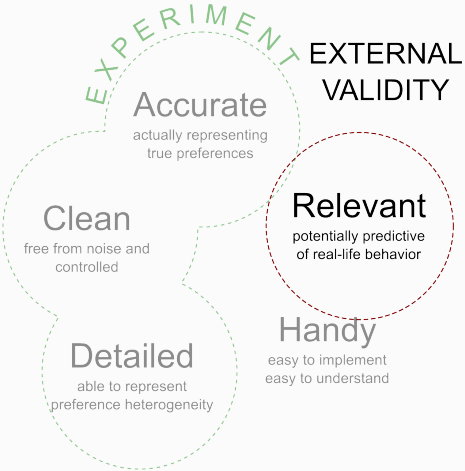
Methods compared



Experiments: internal validity



Experiments: external validity



How do we evaluate a task?

A good risk elicitation control task should be...

Complete: identifying a wide range of preferences (risk-averse/risk-loving);

Precise: delivering a *fine* estimate of risk attitudes (many categories);

Parsimonious: needing a *low* number of choices from subjects;

Intuitive: easy to understand, *easy* to implement (lab, field);

Sound: theoretically sound, not *ambiguous*, not exposed to commonly observed biases.

Correlated: with questionnaires, real-life situations, self-assessments.

Risk Elicitation Tasks

Psychology tasks

Psychologists have come up with a list of risky tasks that are tailored to the specific needs of each manipulation / experiment / theory:

- left-turn in high traffic
- hand in freezing water
- guessing given little information
- (various forms of) gambling
- playing (various forms of) card games
- Deal or No Deal game
- ...

(see Byrnes JP, Miller DC, Schafer WD (1999) Gender differences in risk taking: A meta-analysis. Psych. Bull. 125(3):367–383. for a list of tasks seen from a gender perspective)

Pros and cons

Pros

- external validity
- real worlds behavior
- losses

Cons

- no or little theory
- generalizability dubious
- really, hand in freezing water?

Questionnaires: SOEP

How likely are you to take risks in general, one a scale from 0 (not taking any risks) to 10 (taking many risks)?

Questionnaires: DOSPERT

Domain **Spe**cific **R**isk **T**aking Scale

- 6 domains: investing, gambling, health/safety, recreational, ethical, and social
- 1 to 7 scale: *how likely are you to engage in X?*

Examples:

- Riding a motorcycle without a helmet.
- Engaging in unprotected sex.
- Investing 10% of your annual income in a moderate growth diversified fund.

Questionnaire: Pros and cons

Pros

- external validity
- real world behavior
- "near" to the obect of interest

Cons

- map - territory
- results not suitable to be plugged into models
- averaging over items is a dubious exercise
- (what do you think)?

Auctions

In the 1970s and 80s it was proposed to use *auctions* to elicit the risk aversion of subjects.

- suppose you are bidding for an object worth 10 euro to you
- against a computerized opponent that bids uniformly on $[0; 10]$.
- if you bid more than the opponent, you win the auction and get the object.
- if you bid less than the opponent, you lose the auction.

What should you do?

- Your earnings are

$$\Pi = \begin{cases} 10 - \text{bid} & \text{if bid} \geq U(0;10) \\ 0 & \text{if bid} < U(0;10) \end{cases}$$

- this means that if you bid 1 you get 9 with probability 10%; if 2 you get 8 with probability 20%, and so on...
- turns out that your opt imal strategy if risk neutral is to bid $10/2 = 5$
- but if you are *risk averse* you should bid *more*.

Overbidding = risk aversion

Pros and cons

Pros

- robust theory
- incentivized – monetary consequences

Cons

- much can be going on other than risk aversion
- lots of instructions
- feels very artificial

Choice over lotteries

Under EUT, there is *one* risk attitude and can be identified with *one* continuous certainty equivalent choice or a *small set* of lottery choices.

Natural for economists under EUT to *directly* use lotteries to elicit risk attitudes.

- lotteries are simple objects
- incentivizable
- less bulk than auctions
- portable and easy to ask
- to allow for noise, just ask *many* lottery choices

RETs, I: Holt and Laury

Ten binary lottery choices – risk attitude as switching point

	Option A				Option B			
1	1/10	4 €	9/10	3.2 €	1/10	7.7 €	9/10	0.2 €
2	2/10	4 €	8/10	3.2 €	2/10	7.7 €	8/10	0.2 €
3	3/10	4 €	7/10	3.2 €	3/10	7.7 €	7/10	0.2 €
4	4/10	4 €	6/10	3.2 €	4/10	7.7 €	6/10	0.2 €
5	5/10	4 €	5/10	3.2 €	5/10	7.7 €	5/10	0.2 €
6	6/10	4 €	4/10	3.2 €	6/10	7.7 €	4/10	0.2 €
7	7/10	4 €	3/10	3.2 €	7/10	7.7 €	3/10	0.2 €
8	8/10	4 €	2/10	3.2 €	8/10	7.7 €	2/10	0.2 €
9	9/10	4 €	1/10	3.2 €	9/10	7.7 €	1/10	0.2 €
10	10/10	4 €	0/10	3.2 €	10/10	7.7 €	0/10	0.2 €

Risk neutral should switch after 5 choices. $>$ 5 safe $\rightarrow$ risk averse

Pros and cons

Pros

- robustly linked to EUT
- incentivized – monetary consequences

Cons

- might be difficult to parse by subjects
- (what do you think?)

RETs, II: Binswanger

A single choice among 50-50 lotteries – chosen lottery is played.

	Event	Probability	Outcome
1	A	50%	4 €
	B	50%	4 €
2	A	50%	6 €
	B	50%	3 €
3	A	50%	8 €
	B	50%	2 €
4	A	50%	10 €
	B	50%	1 €
5	A	50%	12 €
	B	50%	0 €

Risk neutral should choose lottery 5. Extreme risk aversion to choose lottery 1.

Pros and cons

Pros

- robustly linked to EUT
- incentivized – monetary consequences
- easier than HL

Cons

- only 50-50 lotteries
- risk lovers?
- (what do you think?)

RETs III: Certainty equivalent price lists

A	B	
100%	50%	50%
0		
10		
20		
30		
40		
50	100	0
60		
70		
80		
90		
100		

Risk-neutral chooses 50.

Pros and cons

Pros

- robustly linked to EUT
- incentivized – monetary consequences

Cons

- Might be easier to parse than HL
- in a way, a bridge between HL and Binswanger
- central bias?
- (what do you think?)

RETs,IV: Gneezy and Potters

Endowment X

How much would you like to invest?

Safe account
1 : 1

Risky investment
1 : {1/2: 2.5; 1/2: 0}

Risk-neutral should invest all, as $E(\text{risky}) = 1.25 > 1$.

Pros and cons

Pros

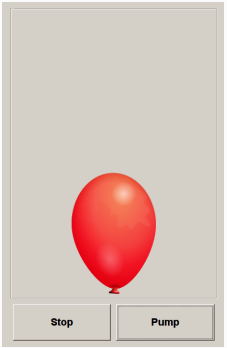
- robustly linked to EUT
- incentivized – monetary consequences
- investment context

Cons

- only one lottery, sensitive to parameter choice
- risk lovers?
- (what do you think?)

RETs, V: Balloon

Inflating a balloon with increasing probability of explosion



Risk-neutral should stop halfway – but not enough information

Pros and cons

Pros

- intuitive
- might be fun – might be related to gambling

Cons

- ambiguity!
- serial correlation
- (what do you think?)

RETs, VI: BRET

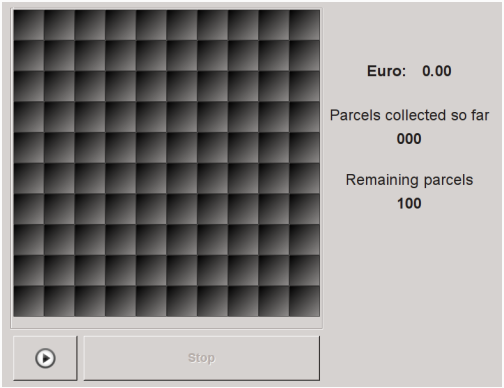


Figure: The BRET interface at the start of the experiment

BRET: interface, I

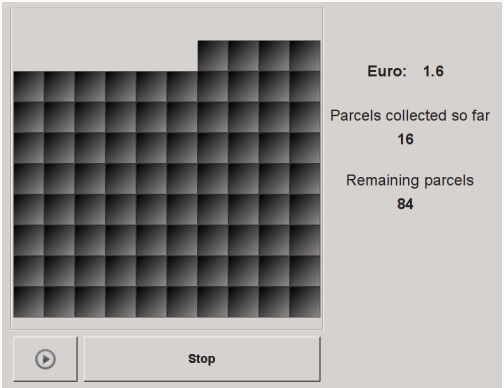


Figure: The BRET interface after 16 seconds

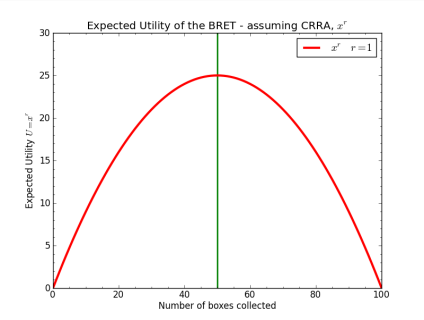
BRET: under the hood

- Theoretically, the task amounts to choosing the preferred among 101 lotteries.
- Each lottery is characterized as

$$L^k = \begin{cases} 0 & \frac{k}{100} \\ k & \frac{100-k}{100} \end{cases}$$

- The 101 lotteries are all summarized by the parameter k...
- ...that is also governing probabilities.
- Example: at $k = 20$, $L = \{20\% : 0\,;\, 80\% : 20\}$

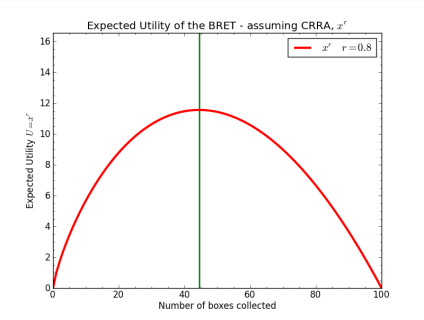
BRET: solution for the expected value maximizer



The expected value is maximized at $k^* = 50$.
Assuming a power CRRA utility function x^r, the optimal stopping point is:

$$k^* = 100\frac{r}{1+r}.$$

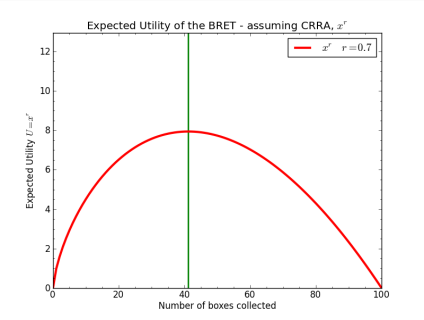
BRET: Risk averse subject



The expected value is maximized at $k^* = 50$.
Assuming a power CRRA utility function x^r, the optimal stopping point is:

$$k^* = 100\frac{r}{1+r}.$$

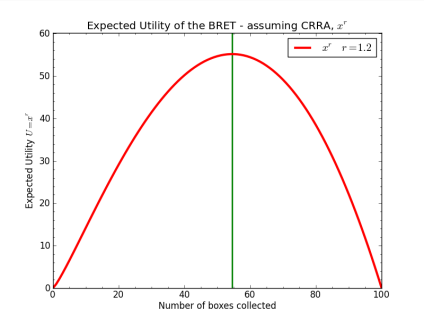
BRET: Risk averse subject



The expected value is maximized at $k^* = 50$.
Assuming a power CRRA utility function x^r, the optimal stopping point is:

$$k^* = 100\frac{r}{1+r}.$$

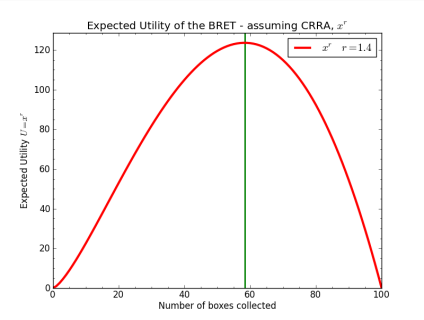
BRET: Risk lover subject



The expected value is maximized at $k^* = 50$.
Assuming a power CRRA utility function x^r, the optimal stopping point is:

$$k^* = 100\frac{r}{1+r}.$$

BRET: Risk lover subject



The expected value is maximized at $k^* = 50$.
Assuming a power CRRA utility function x^r, the optimal stopping point is:

$$k^* = 100 \frac{r}{1+r}.$$

Pros and cons

Pros

- intuitive (?)
- might be fun – might be related to gambling
- strongly related to theory

Cons

- artificial
- might be misunderstood
- (what do you think?)

How do measures fare?

Slovic (1962)

TABLE 1

INTERCORRELATIONS AMONG RISK TAKING MEASURES
($N = 82$)

Variable	1	2	3	4	5	6	7	8
Response sets								
1 Dot Estimation								
2 Word Meanings	−.17							
3 Test Risk	.16	.05						
Questionnaires								
4 Life Experience Inventory	.05	.27**	−.04					
5 Job Preference Inventory[a]	.07	−.14	−.19	−.06				
Gambling preferences								
6 Self-Crediting Test	−.08	.19*	−.24*	.05	.09			
7 Variance preferences	.32**	.03	−.07	.23*	.07	.04		
8 Probability preferences	.16	−.03	−.07	−.03	−.35*	−.20	−.17	
Ratings								
9 Risk rating	.05	.00	−.24*	.34**	.10	−.02	.02	.18[a]

"... future research must carefully consider the problem of adequately **defining** *and* **assessing** *risk taking behavior."*

METARET

- **Part 1: state of the art**
  - a *detailed map* of elicited risk attitudes
  - an assessment of *convergent validity*
  - an assessment of *predictive validity*
- **Part 2: moving forward**
  - theoretical: what are we measuring?
  - empirical: develop a better tool

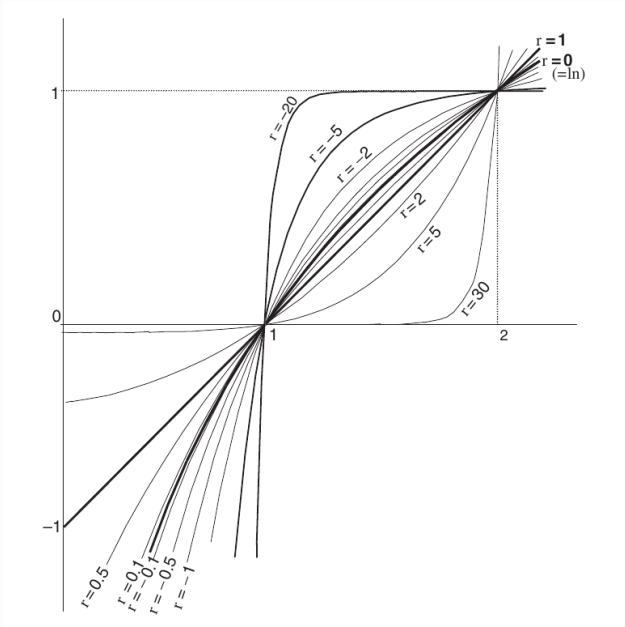
METARET resources

- data from experimental economists worldwide
- preregistration on OSF
- transparent data collection & analysis on gitHub
- live data exploration on a shiny app

CRRA (à la Wakker)

$$u(x) = x^r$$

- simple
- captures risk aversion
- makes different tasks comparable


r = 1
r = 0
(=ln)
r=-20
r=-5
r=-2
r=2
r=5
r=30
r=0.5
r=0.1
r=-0.1
r=-0.5
r=-1
1
0
−1
1
2

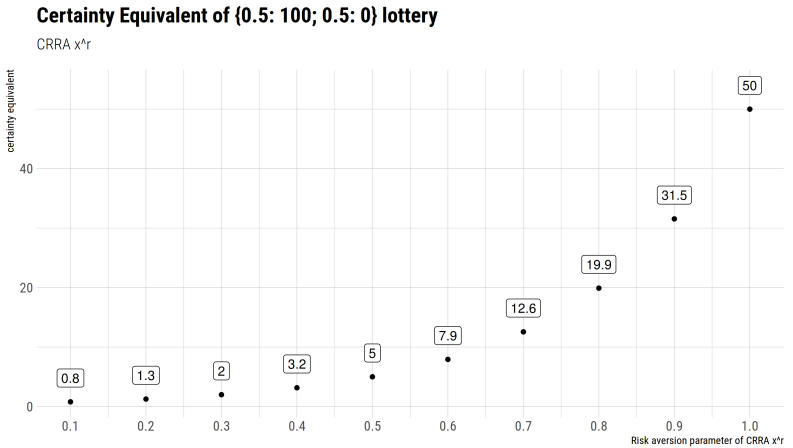
Elicited attitudes

What is the elicited level and dispersion of risk attitudes?

- **low** consistency across tasks
- surprisingly, **low** consistency also *within* tasks
- but **heterogeneity** by task is large
- only result that holds: most people are *risk averse*

  possible explanation: between-subjects variation.

How big are these differences?



Questionnaires

Another approach: ask the subjects!

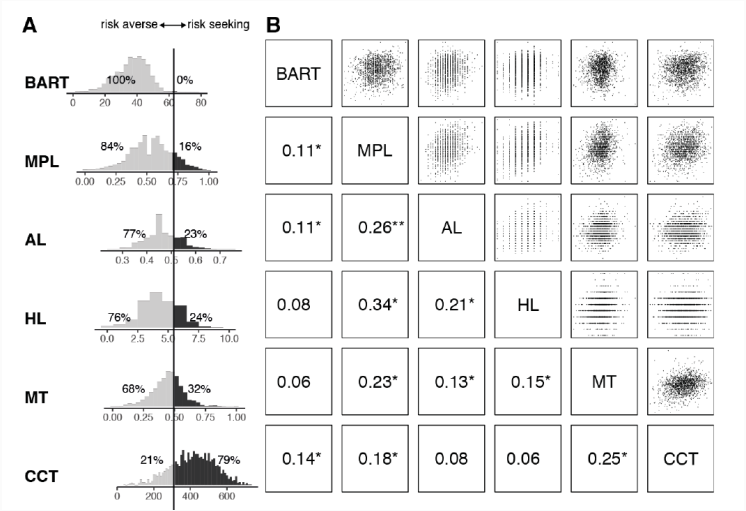
- **better** consistency across samples
- a tendency to report *'in the middle'*
- we do not really know what those number mean

Convergent validity: correlations among RETs

Within-subjects, do measures correlate?

- we replicate Slovic 1962 (!!)
- no correlation higher than .35
- when translating into r things get *worse*

Convergence: more evidence

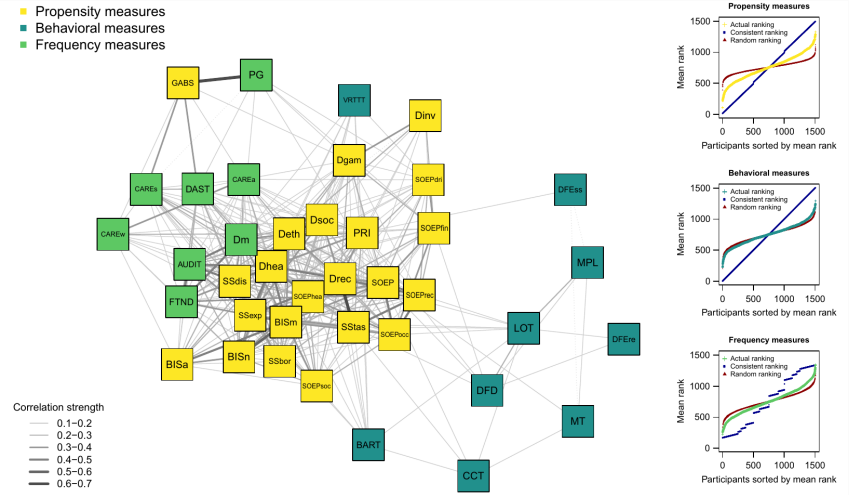


Convergent validity: correlations among questionnaires

Within-subjects, do measures correlate?

- questionnaires by and large more stable
- some exceptions
- is it surprising – middle of the range?

Predictive validity: more evidence



Predictive validity: quest$<->$RETs

Do RETs predict questionnaires (and vice versa)?

- **low** correlations with questionnaires
- across questionnaires and tasks
- Beauchamp et al JRU 2016: questionnaires are rather predictive

we have a problem

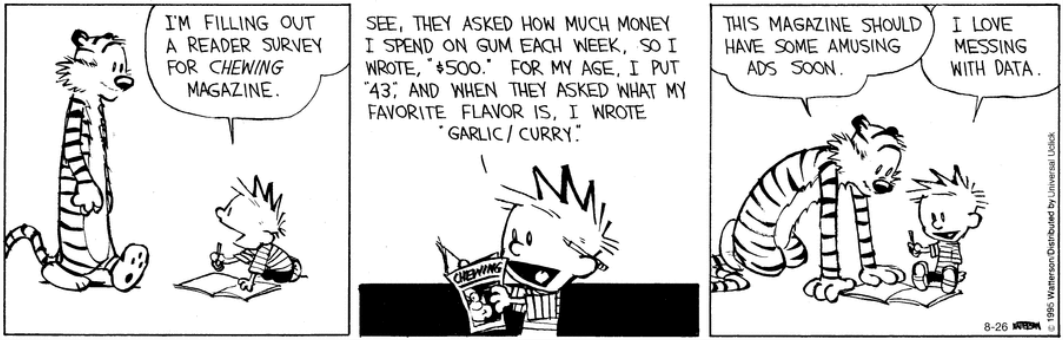
How do we fix it?

Possible causes

- Hypothetical bias
- Task-specific error: characteristics of the tasks
- Noise
- Risk perception
- Theory

Hypothetical Bias

Hypothetical bias



But again...

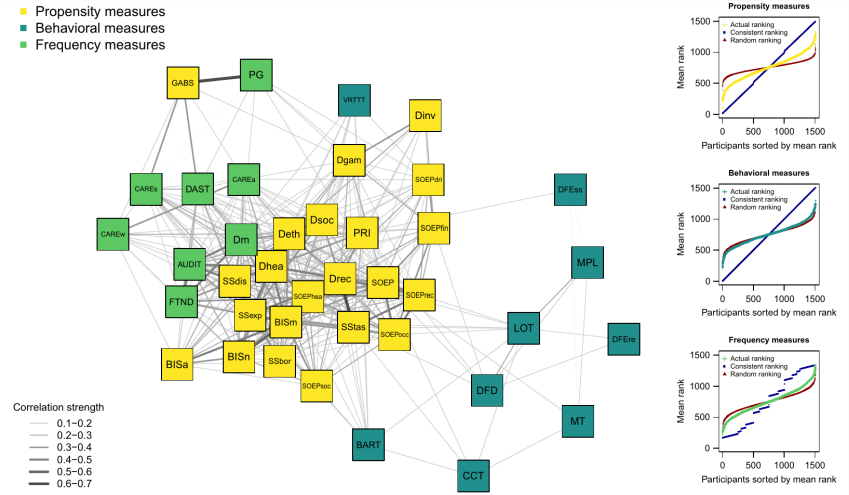


Figure: Brick et al., Science Advances 2017

Task-specific error

Impact of the tasks

- characteristics of the task
- complexity of the task

Characteristics of the task

Crosetto and Filippin, Experimental Economics 2016

Between-subjects comparison across 4 tasks (Jena, DE):

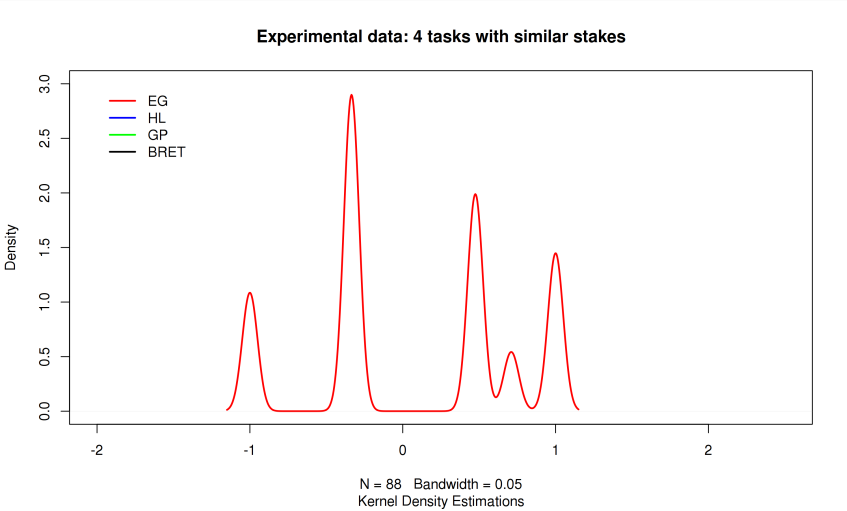
1. Holt & Laury
2. Eckel & Grossman
3. Charness & Gneezy
4. The 'Bomb' Risk Elicitation Task (BRET)

Experimental design

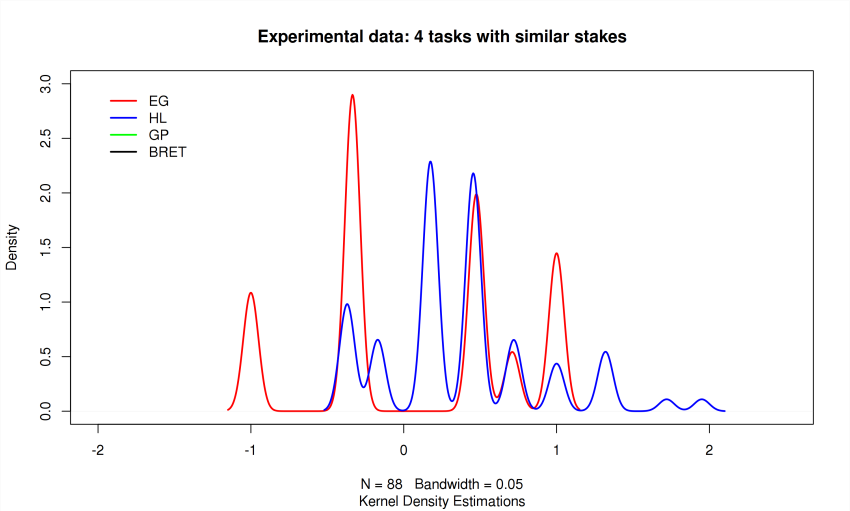
- Max Planck Institute for Economics Lab, Jena, March-May 2012.
- A total of 350 subjects (151 M, 199 F).
- Pure between subjects design, one shot.
- 'Live' resolution of uncertainty at the session level.
- Task + SOEP and DOSPERT

	HL	EG	CGP	BRET	Total
Males	37	45	37	32	**151**
Females	51	43	49	56	**199**
Total	**88**	**88**	**86**	**88**	**350**

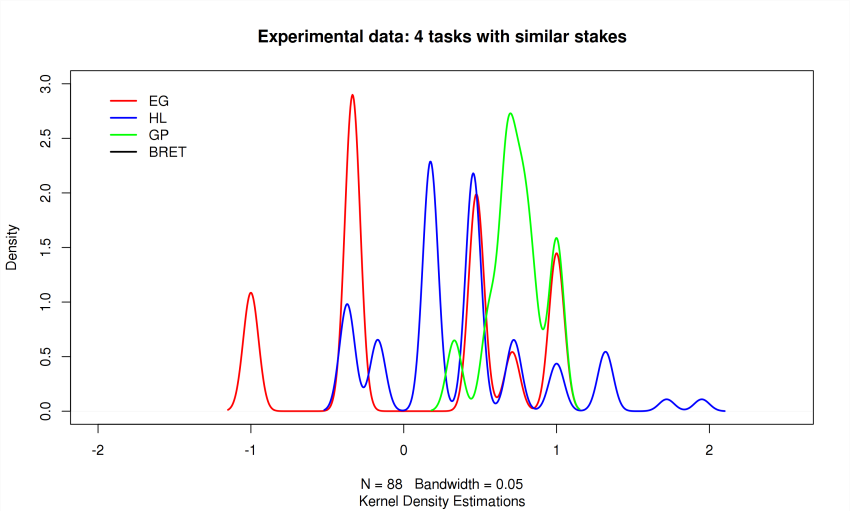
Distribution of r



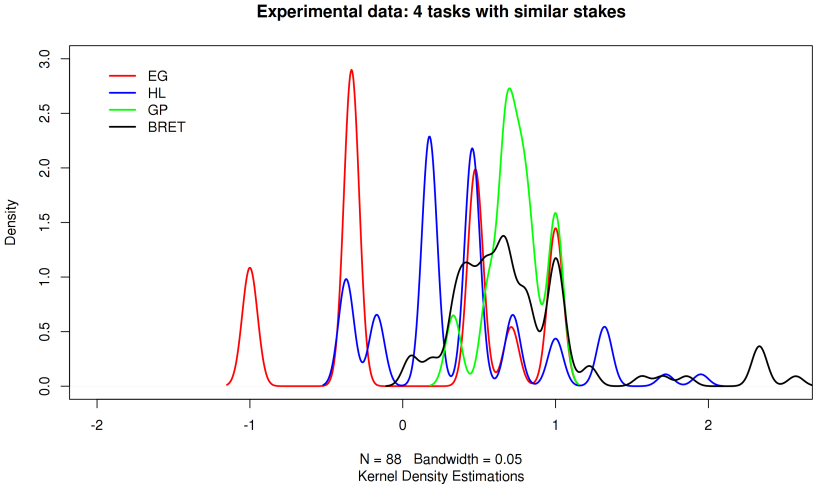
Distribution of r



Distribution of r



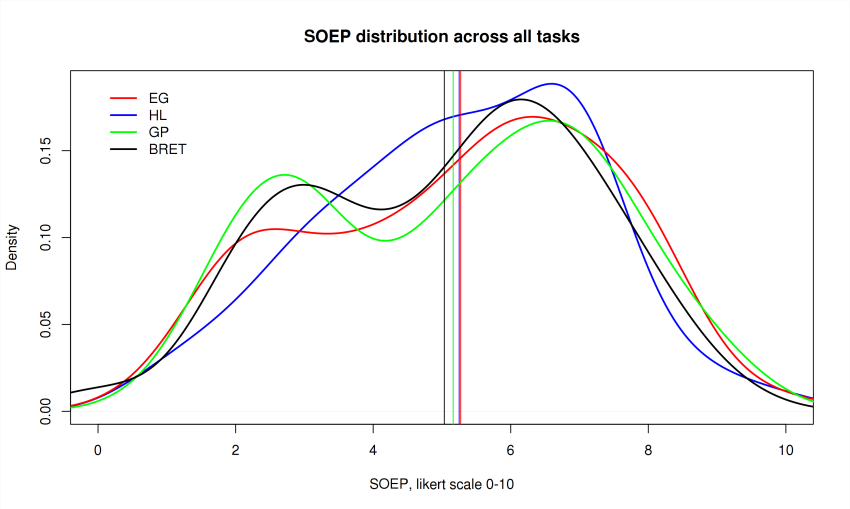
Distribution of r



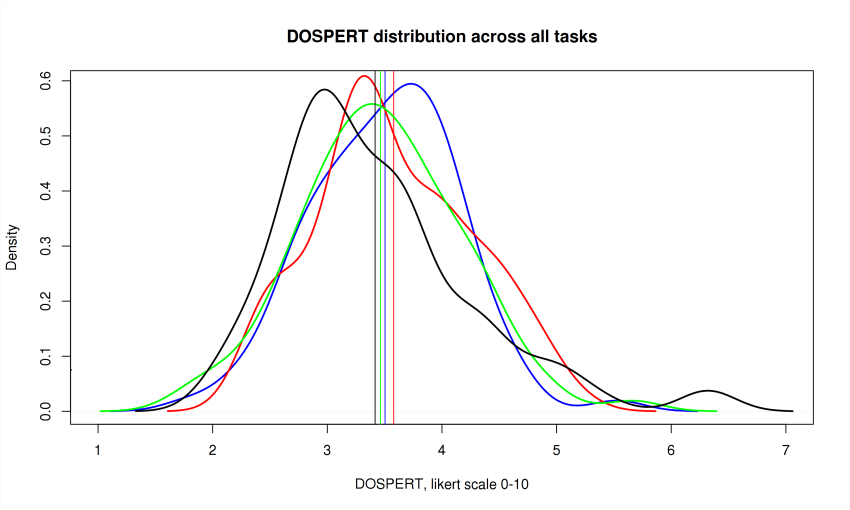
Just because we are between-subjects? SOEP

treatment	N	soep.mean	soep.stdev	dospert.mean	dospert.stdev
hl	88	5.25	1.95	3.50	0.64
eg	88	5.27	2.16	3.58	0.69
gp	86	5.16	2.18	3.46	0.71
bret	88	5.03	2.10	3.42	0.85

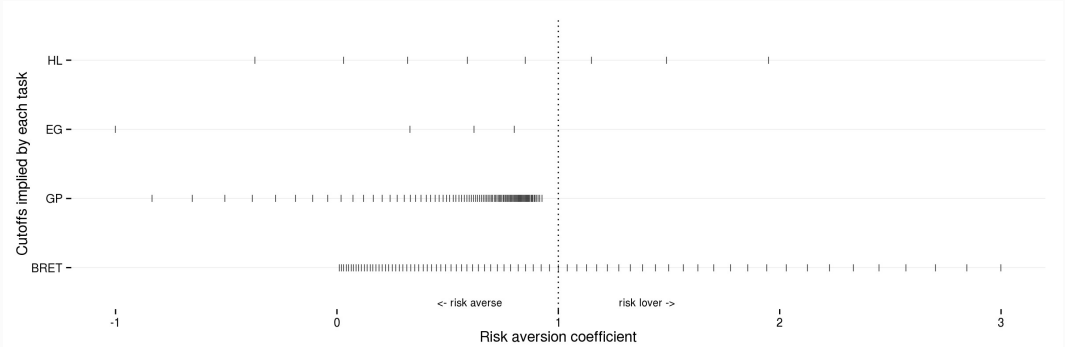
Just because we are between-subjects? SOEP



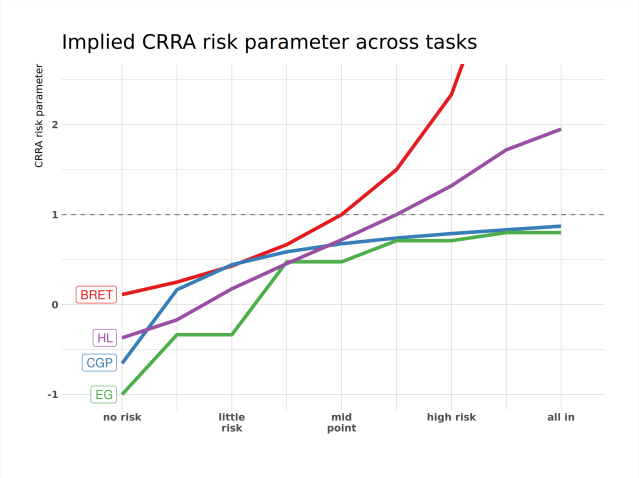
Just because we are between-subjects? DOSPERT



Mapping choices to r: cutoffs



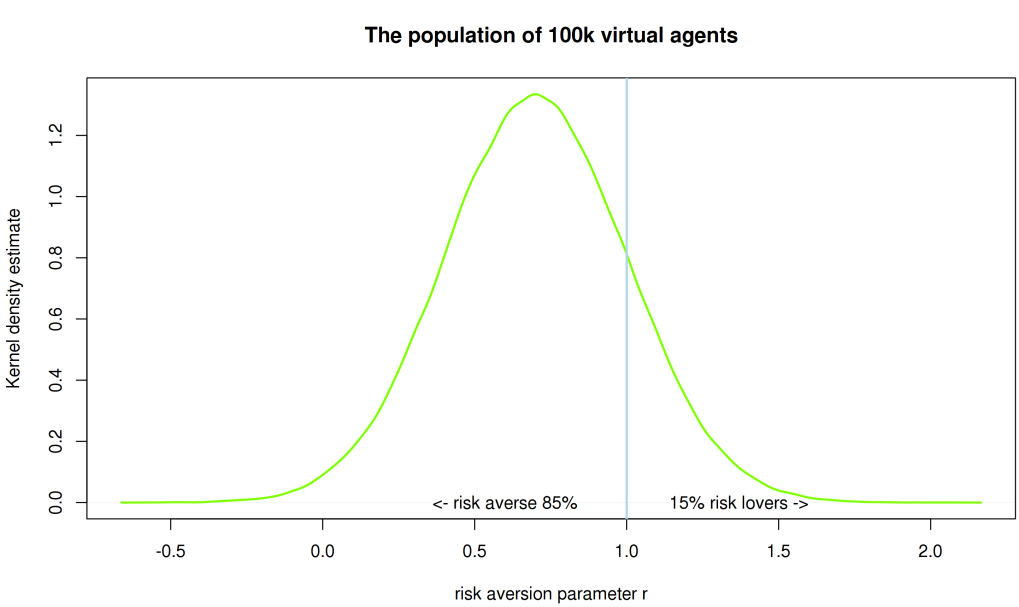
Mapping choices to r: risk levels



Clean test: simulations

How does the mere mechanics of each task affect the outcome?

- Simulation exercise:
  - Generate 100k virtual agents
  - for each agent, $r \sim N(0.7, 0.3)$
  - let the agents play each of the 4 tasks
  - collect results, run statistics
  - analyze the retrieved $\hat{r}$
- a good task should be able to recreate the starting distribution, if no error.

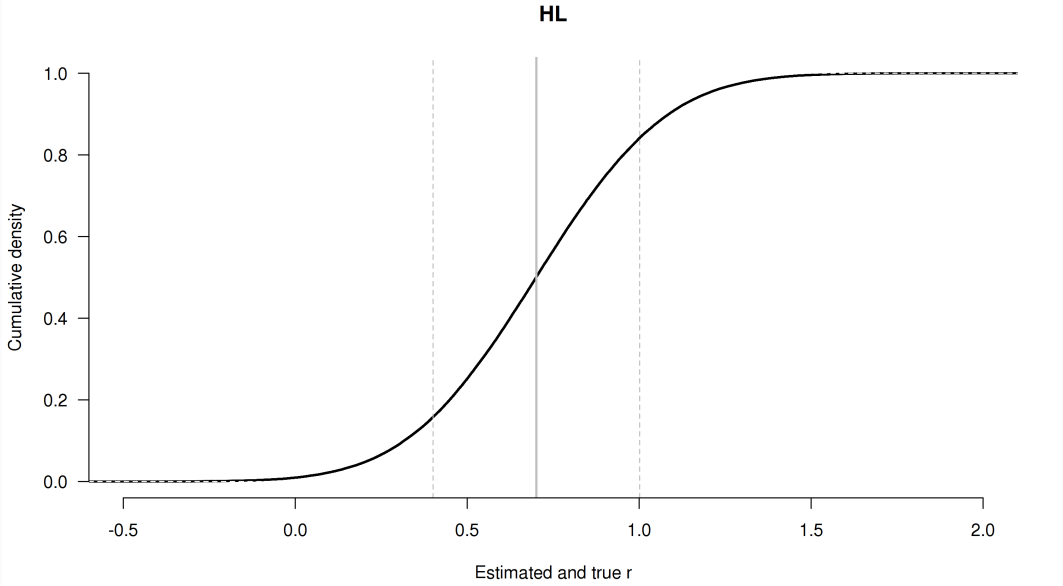

The population of 100k virtual agents
Kernel density estimate
0.0
0.2
0.4
0.6
0.8
1.0
1.2
<- risk averse 85%
15% risk lovers ->
-0.5
0.0
0.5
1.0
1.5
2.0
risk aversion parameter r

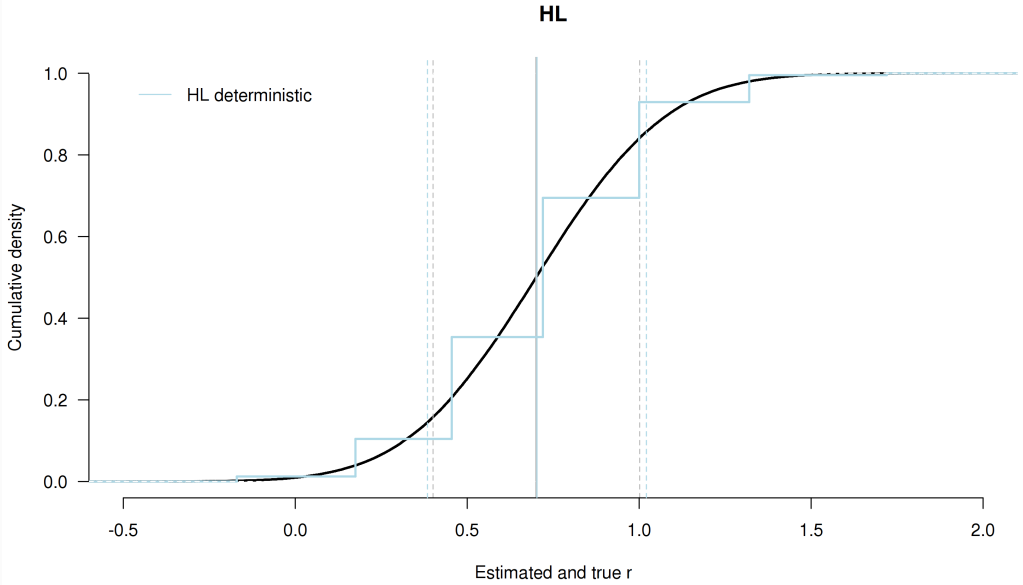
Simulations, II

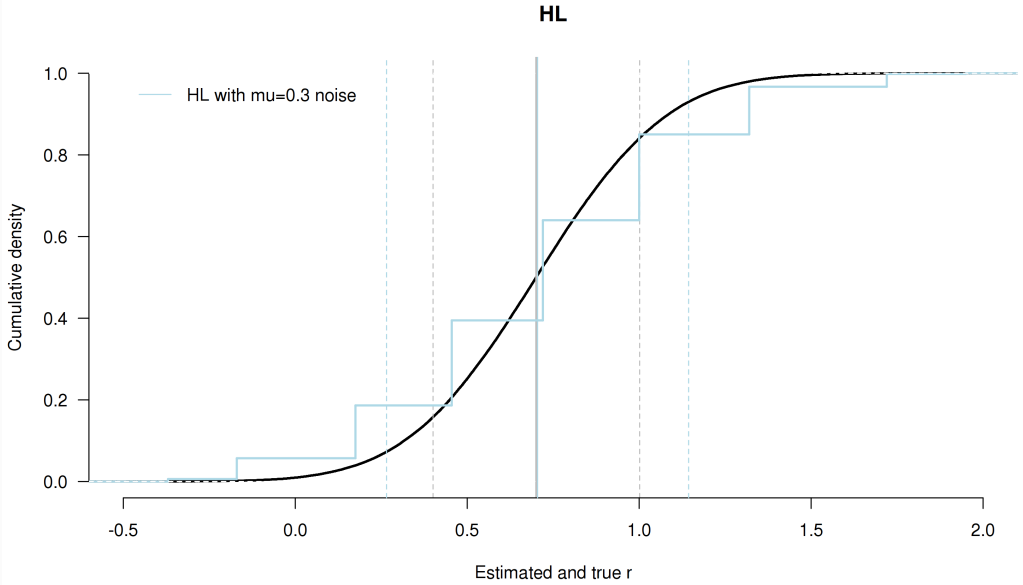
Three types of simulations:

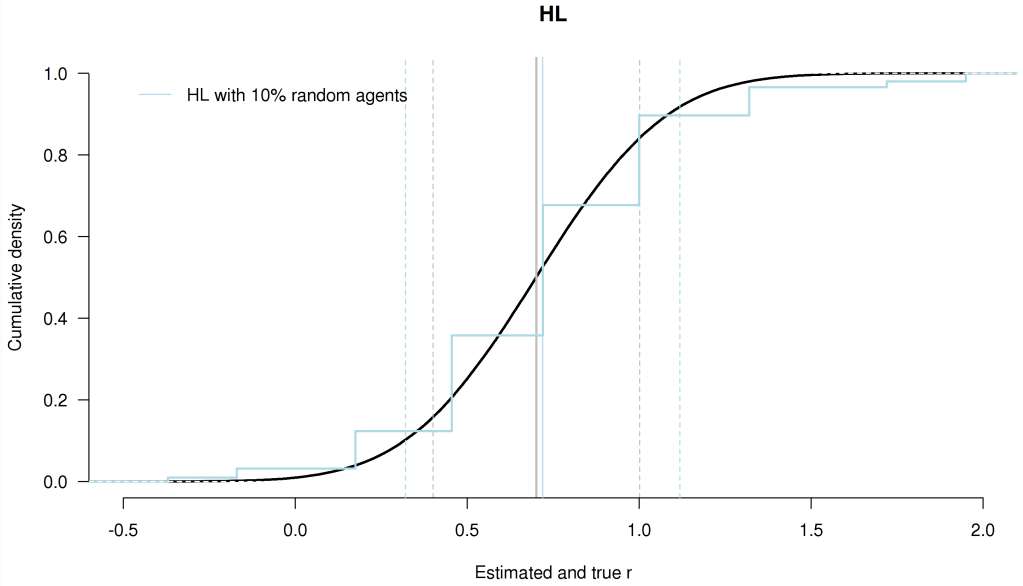
1. Deterministic: virtual subjects play according to their **true r**
2. Random parameter model:
   - for each agent, $r_n = r + \varepsilon, \varepsilon \sim N(0, \mu)$
   - that is, the agent deviates from her true preferences with a white noise
   - $\mu = 0.3$ or 0.6
3. Trembling hand: behaviorally random:
   - a 10% share of subjects just chooses uniformly random
   - on the *task space*: i.e., same likelihood of switching in row 1 as in row 10 in HL.
   - models both error and (extreme) frame effects

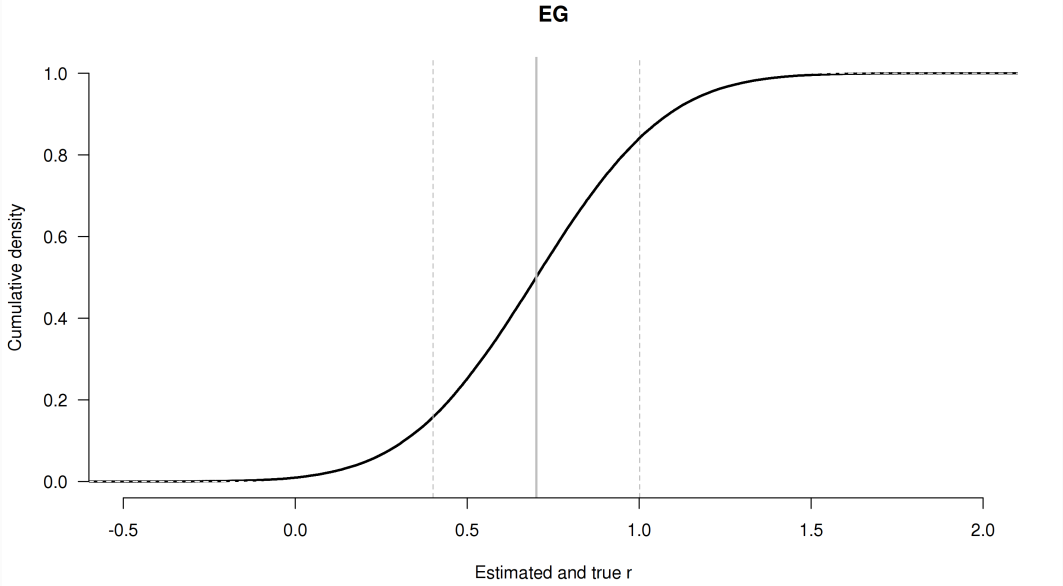
Task by task...

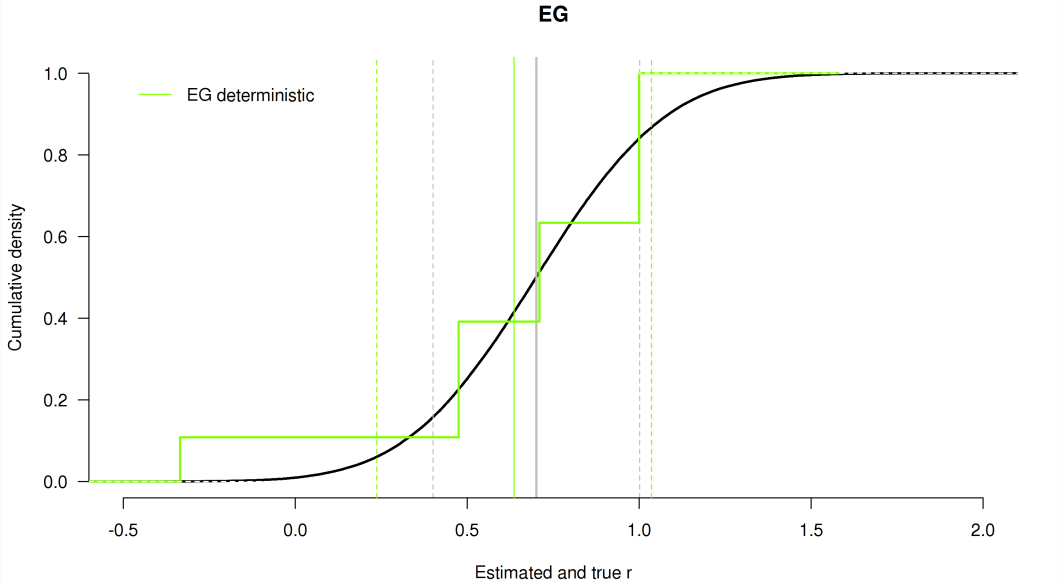

HL
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

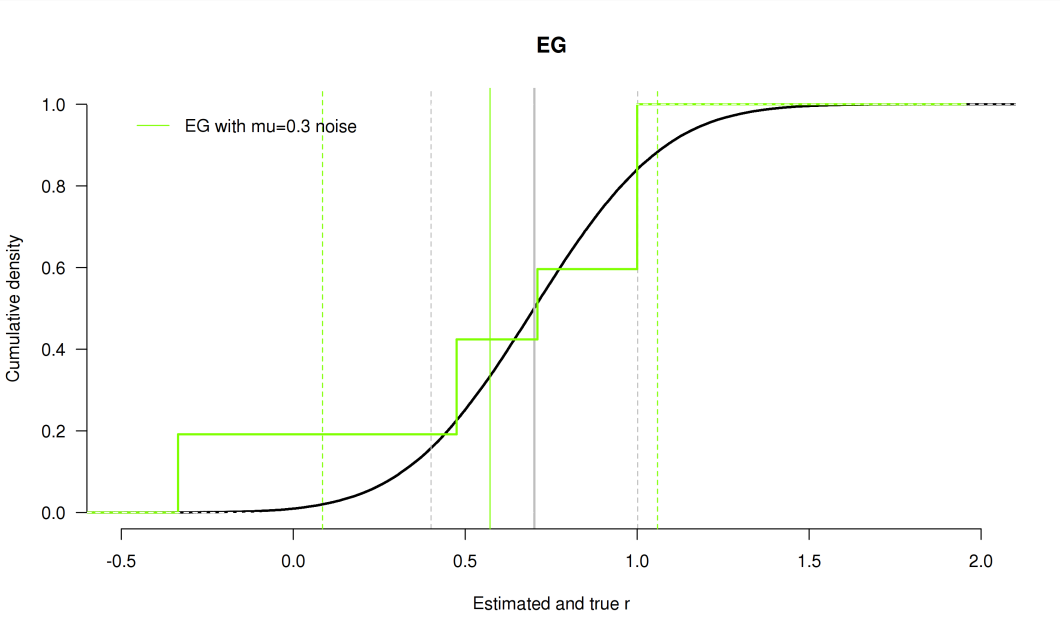

HL
HL deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

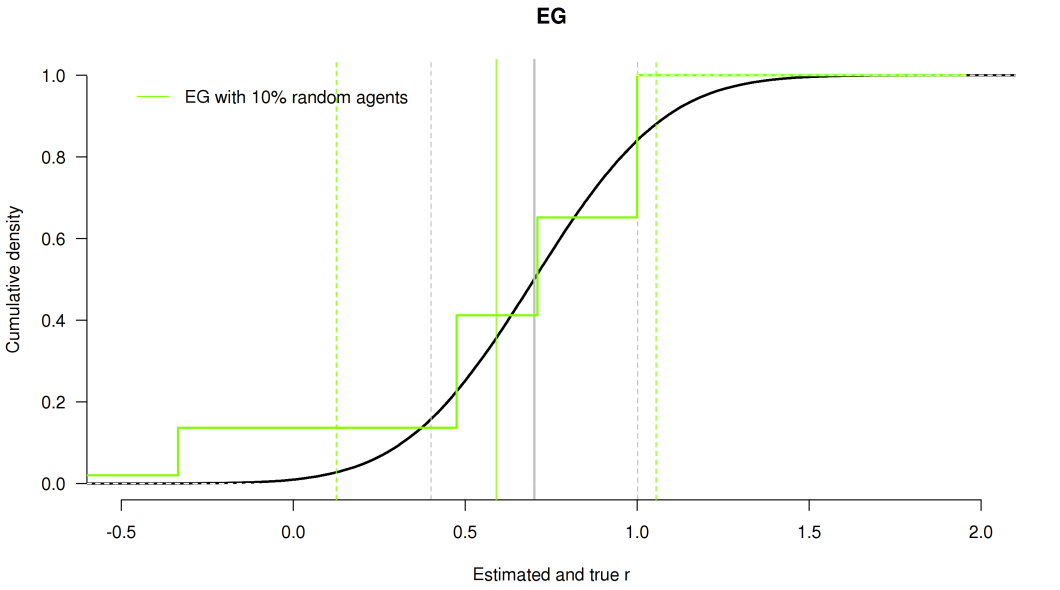

HL
HL with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

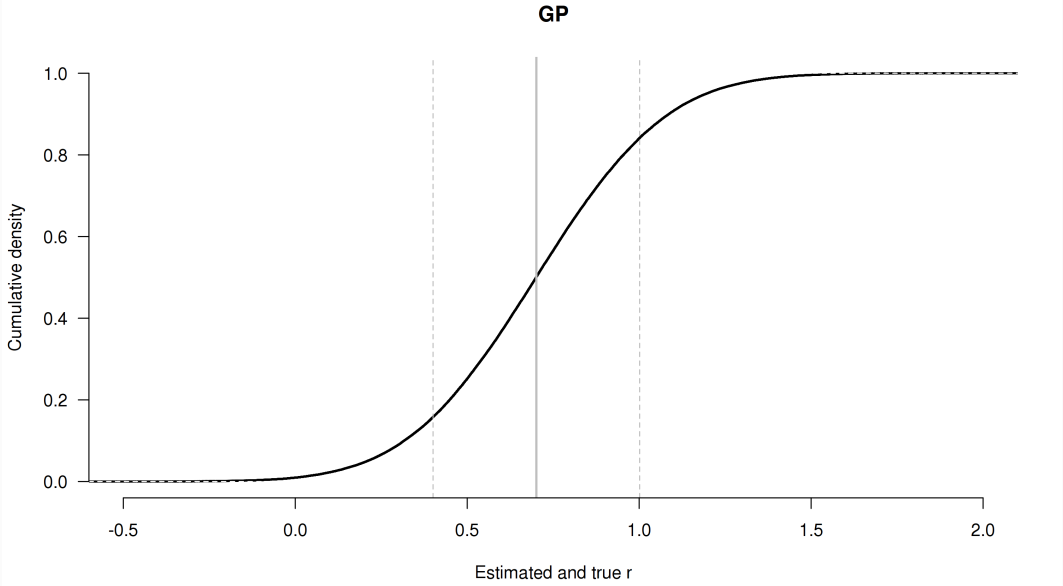

HL
HL with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

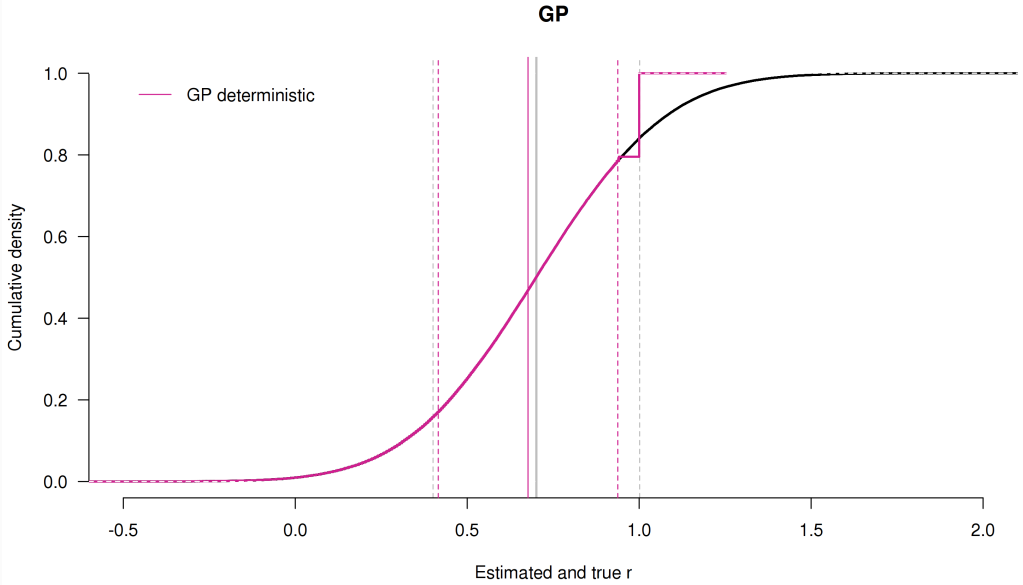

EG
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

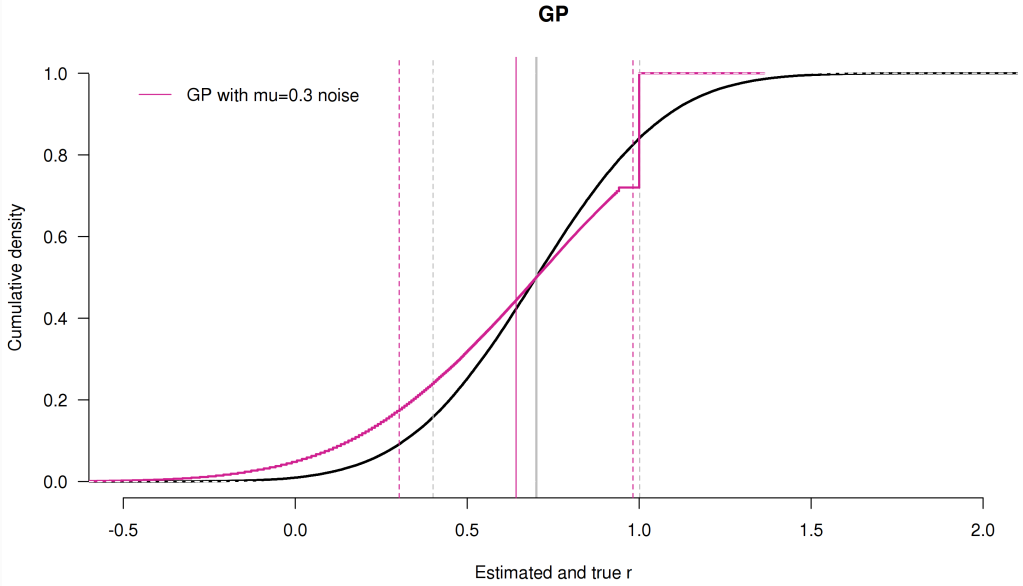

EG
EG deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

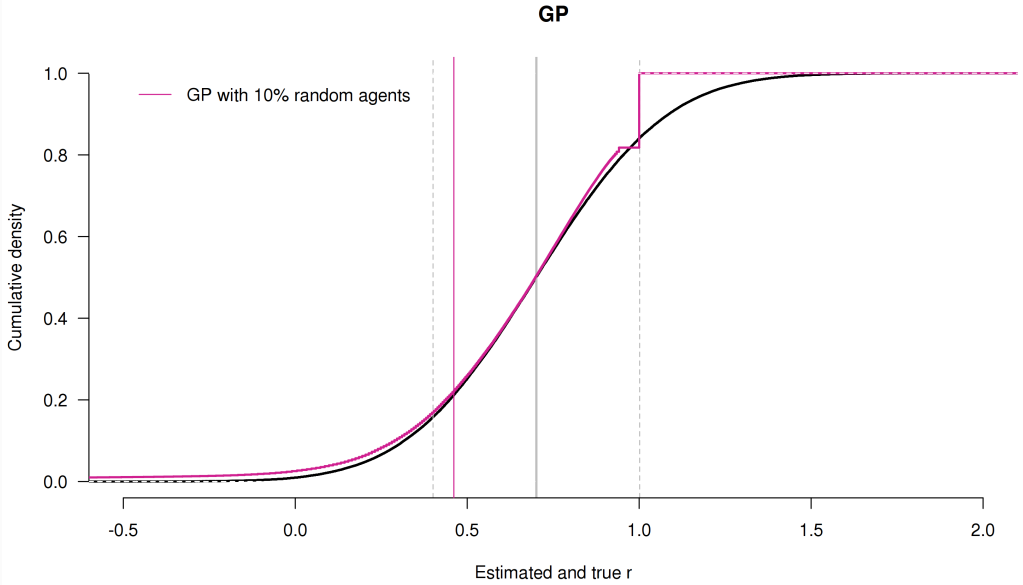

EG
EG with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

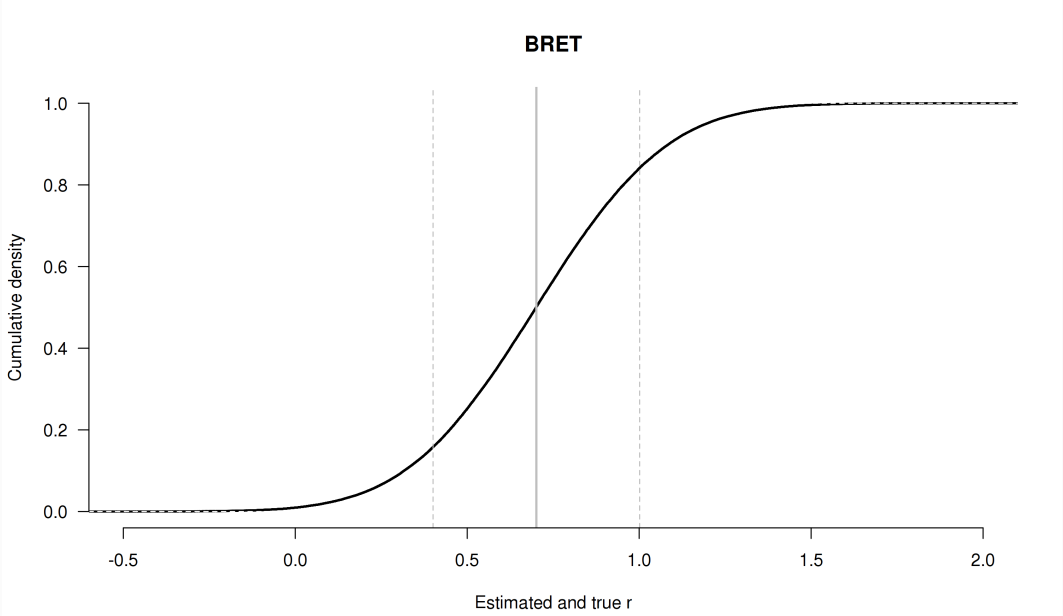

EG
EG with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

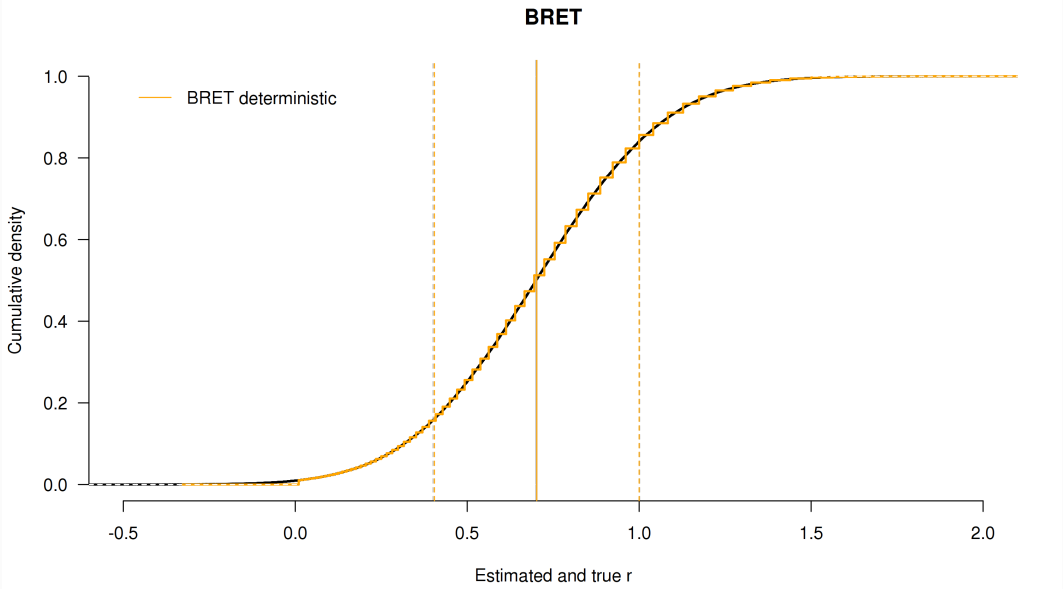

GP
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

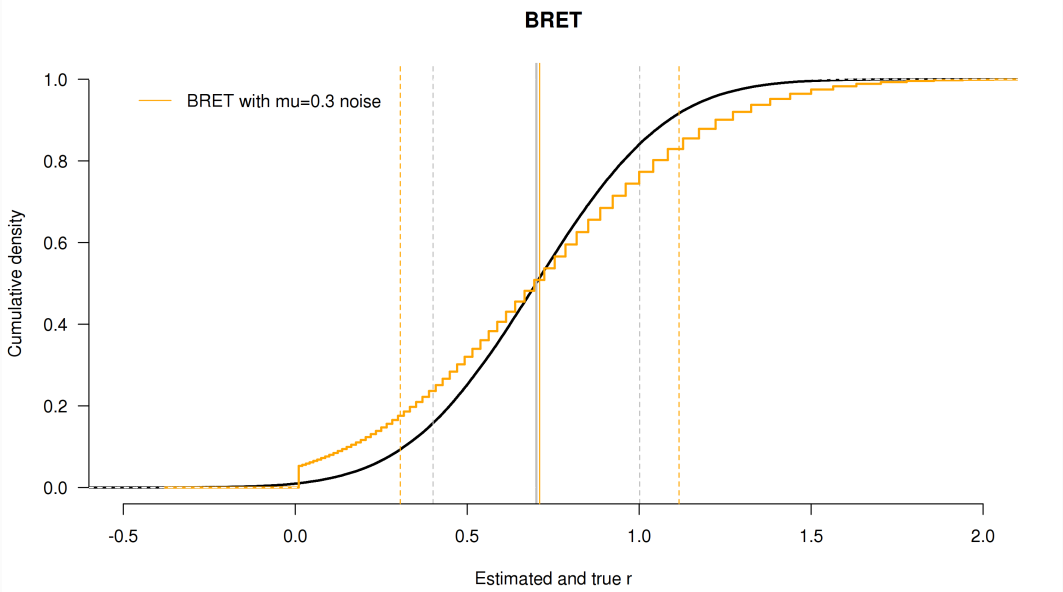

GP
GP deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

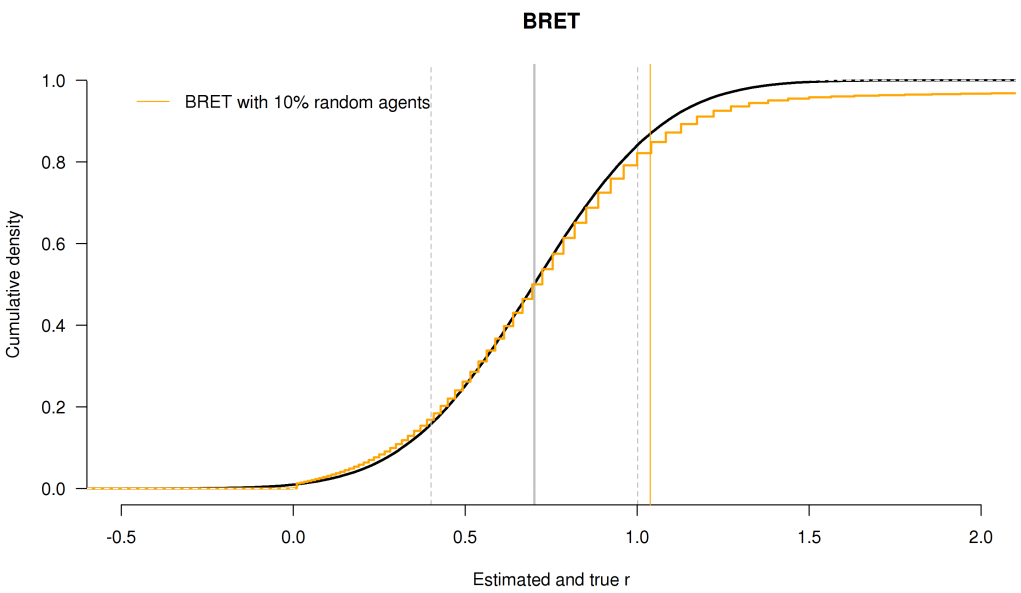

GP
GP with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


GP
GP with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


BRET
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

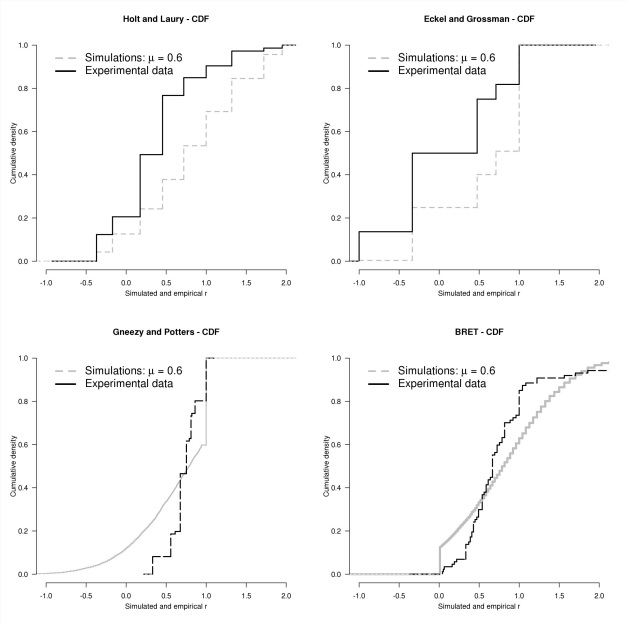

BRET
BRET deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


BRET
BRET with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


BRET
BRET with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

Simulation results: table

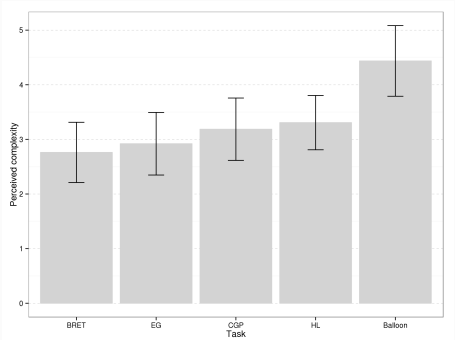
	$\mu = 0$			$\mu = 0.3$			$\mu = 0.6$		
	median	*mean*	*stdev*	*median*	*mean*	*stdev*	*median*	*mean*	*stdev*
True r	**.700**	**.701**	**.300**	**.700**	**.701**	**.300**	**.700**	**.701**	**.300**
HL	.72	.7026	.318	.72	**.704**	.439	.72	**.712**	.627
EG	.71	.636	.399	.71	.572	.487	.71	.490	.572
CGP	**.700**	.676	.261	**.700**	.642	.339	**.704**	.558	**.488**
Bret	.694	**.7022**	**.297**	.694	.710	**.405**	.694	.754	.584


Holt and Laury - CDF
Simulations: μ = 0.6
Experimental data
Cumulative density
Simulated and empirical r
Eckel and Grossman - CDF
Simulations: μ = 0.6
Experimental data
Cumulative density
Simulated and empirical r
Gneezy and Potters - CDF
Simulations: μ = 0.6
Experimental data
Cumulative density
Simulated and empirical r
BRET - CDF
Simulations: μ = 0.6
Experimental data
Cumulative density
Simulated and empirical r

Does all this explain task differences?

- No. Some of the differences across task are accounted by mechanics.
- especially for EG/BRET
- others are not, especially for GP.
- What else might be driving the differences?

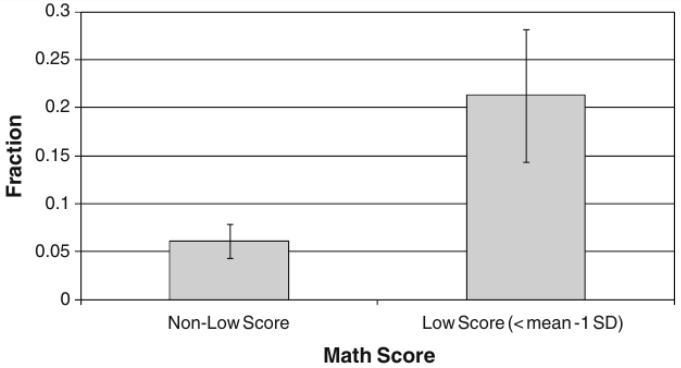
Complexity



- self-reported complexity based on a 0-10 complexity scale.
- HL perceived as more difficult
- BRET, CGP and EG lower, indistinguishable
- Excluding inconsistents in HL, all four tasks undistinguishable

Task complexity? Dave et al 2010 JRU

Fig. 1 Fraction of Subjects Making Inconsistent Choices in the HL Task*



Treating Noise

Noisy preferences

If preferences are noisy, then there will be noise in the data

How to treat noise?

- Average over different measures / questionnaires
- Use econometrics and get population estimate + noise
- Use structural modeling and get population estimate + theory + noise
- Use structural modeling at the individual level (needs many choices)

Averaging over tasks

For RETs:
Menkhoff & Sakha, *Estimating risky behavior with multiple-item risk measures* , Jo Econ Psy 2017
For questionnaires:
Beauchamp et al., *The psychometric and empirical properties of measures of risk preferences*, JRU 2017

Structural modeling

- Assume a theory (e.g., EUT, PT, ...)
- Set up the equations describing the theory
- Link the equations to the data
- Estimate parameter via maximum likelihood
- Let parameters vary on demographics
- Let parameters depend on noise

HL

	Option A				Option B			
1	1/10	4 €	9/10	3.2 €	1/10	7.7 €	9/10	0.2 €
2	2/10	4 €	8/10	3.2 €	2/10	7.7 €	8/10	0.2 €
3	3/10	4 €	7/10	3.2 €	3/10	7.7 €	7/10	0.2 €
4	4/10	4 €	6/10	3.2 €	4/10	7.7 €	6/10	0.2 €
5	5/10	4 €	5/10	3.2 €	5/10	7.7 €	5/10	0.2 €
6	6/10	4 €	4/10	3.2 €	6/10	7.7 €	4/10	0.2 €
7	7/10	4 €	3/10	3.2 €	7/10	7.7 €	3/10	0.2 €
8	8/10	4 €	2/10	3.2 €	8/10	7.7 €	2/10	0.2 €
9	9/10	4 €	1/10	3.2 €	9/10	7.7 €	1/10	0.2 €
10	10/10	4 €	0/10	3.2 €	10/10	7.7 €	0/10	0.2 €

Structural modeling

- assume $U(x) = x^r$
- assume subjects evaluate left and right lotteries $EU(L)$; $EU(R)$

$$EU(L) = \frac{1}{10} \cdot (4^r) + \frac{9}{10} \cdot (3.2^r)$$

$$EU(L) = \frac{1}{10} \cdot (7.7^r) + \frac{9}{10} \cdot (0.2^r)$$

- subjects compare utilities and choose accordingly:

$$\text{Decision} = \begin{cases} L & \text{if} \quad EU(L) > EU(R) \\ R & \text{if} \quad EU(L) < EU(R) \end{cases}$$

Adding noise

There are two main ways to add noise:

1. Random utility model (Fechner error)

$$EU(x) = x^r$$

$$\text{Prob}(L) = Prob(EU(L) - EU(R) + \varepsilon > 0);$$

2. Random parameter model

$$EU(x) = x^{r+\varepsilon}$$

$$\text{Prob}(L) = Prob(EU(L) - EU(R) > 0).$$

Error structures

There are several noise specifications possible:

1. probit

$$\varepsilon \sim N(0, \mu^2)$$
$$Pr(L) = \Phi(\frac{EU_L - EU_R}{\mu}).$$

2. logit

$$\varepsilon \sim \Lambda(0, \mu)$$
$$Pr(L) = \frac{1}{1 + e^{-\frac{1}{\mu}(EU_L - EU_R)}}.$$

3. Luce / HL

$$\text{Prob}(L) = Prob(log(EU(L)) - log(EU(R)) + \varepsilon > 0);$$
$$\varepsilon \sim \Lambda(0, \mu)$$
$$Pr(R) = \frac{EU_R^{\frac{1}{\mu}}}{EU_L^{\frac{1}{\mu}} + EU_R^{\frac{1}{\mu}}}.$$

MLE, results

	Log-likelihood	Coefficient	Estimate	St.Err.	p-value
HL	-391.25	r	.427	.064	.000
		r_{female}	-.061	.060	.310
		σ	.433	.090	.000
EG	-194.62	r	.694	.035	.000
		r_{female}	-.262	.057	.000
		σ	.206	.020	.000
CGP	-1546.79	r	.863	.014	.000
		r_{female}	-.093	.023	.000
		σ	.010	.001	.000
Balloon	-2243.81	r	1.13	.066	.000
		r_{female}	-.103	.042	.013
		σ	.345	.078	.000
BRET	-2584.71	r	.696	.089	.000
		r_{female}	.034	.049	.488
		σ	.104	.037	.006

- MLE confirming previous results
- σ interpreted as decision noise $\sim$ complexity.

Dave et al 2010 – similar results

Table 6 Estimates of risk parameter "r" and noise parameter "μ" as functions of characteristics, separate specifications for EG and HL instruments

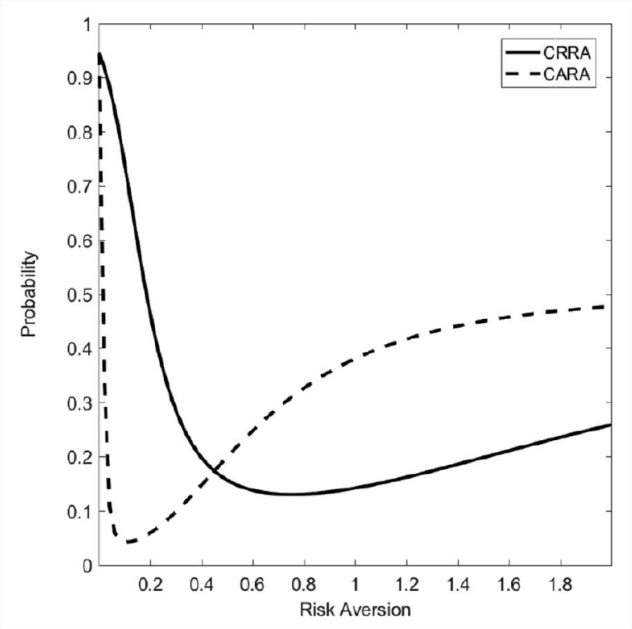
Parameter/Variable	EG Instrument			HL Instrument		
	Estimate	Std. Err.[a]	*p*-value	Estimate	Std. Err.[a]	*p*-value
r						
Female	0.2584	0.0658	0.000	0.1275	0.0386	0.001
Low Income	0.0202	0.0511	0.693	−0.0271	0.0388	0.485
Young	−0.0190	0.0552	0.731	0.0415	0.0384	0.280
Low Math Score	0.2344	0.1570	0.136	0.0218	0.0921	0.813
Constant	0.3260	0.0382	0.000	0.5479	0.0356	0.000
μ						
Female	−0.0134	0.0080	0.092	−0.0337	0.0128	0.009
Low Income	0.0002	0.0063	0.979	0.0113	0.0117	0.332
Young	−0.0074	0.0069	0.283	−0.0390	0.0115	0.001
Low Math Score	0.0123	0.0227	0.587	0.1023	0.0466	0.028
Constant	0.0462	0.0045	0.000	0.1302	0.0117	0.000
Log*L*:	2490.3			3172.2		
Obs.:	4405			8810		

Problem with RUM - Fechner

Apesteguia and Ballester, *MONOTONE STOCHASTIC CHOICE MODELS: THE CASE OF RISK AND TIME PREFERENCES*, Jo Pol Econ 2018

Problem: the probability of choosing the safe lottery should monotonically increase in your risk aversion.

This is **not** the case for RUMs


CRRA
CARA
Probability
Risk Aversion
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8

why?

Two problems

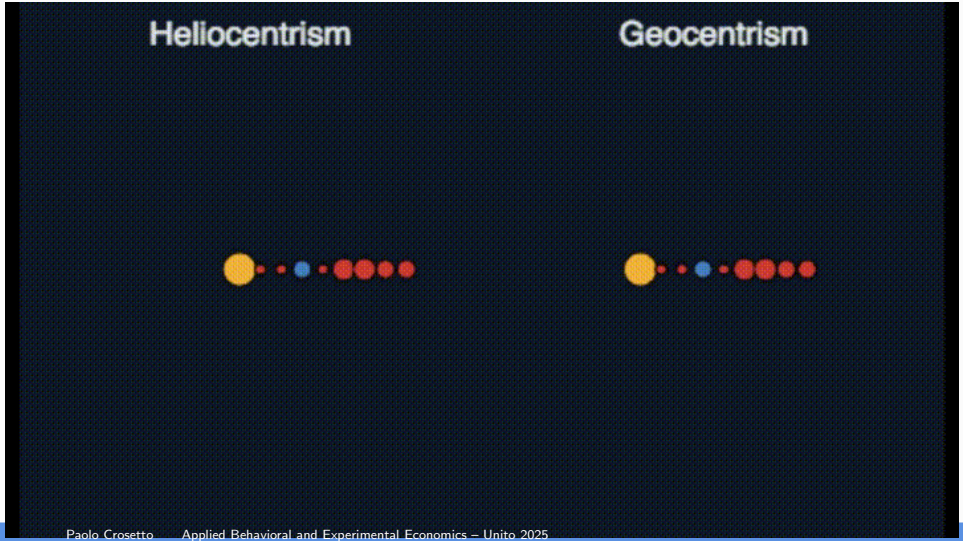
1. probability of safe choice goes to 0.5 as risk aversion increases;
2. curve describing probability non-monotone

1/2 : as r decreases, EU() becomes very small, difference too, so error more important than difference

non-monot : Utilities are ordinal and non linear. Adding an error to a difference implicitly assumes utilities to be cardinal, and linear: i.e. assumes that a difference in utility of 1 is *bigger* than a difference of 0.5; and assumes that EU(diff) = EU(L) - EU(R).

take-home: do not run RUMs (I plead guilty)

Time to change theory?



Risk perception

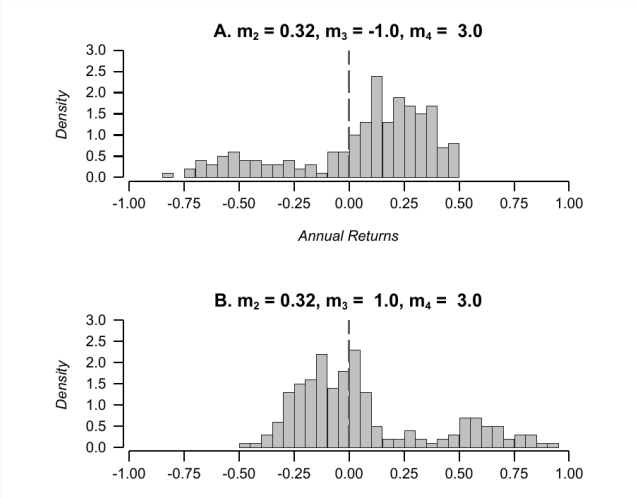
Risk perception: a mismatch

- economists *assume* subjects share the same risk *definition*
- namely:
  - risk as a distribution of **probability** over outcomes
  - *EV* as the average across all possible states of the world
  - risk aversion as diminishing marginal utility of money
  - subjects care about **variance**
- but subjects think of risk as *probability of a loss*
- *do subjects find our tasks risky?*
- We **do not know** because we **assume** they do

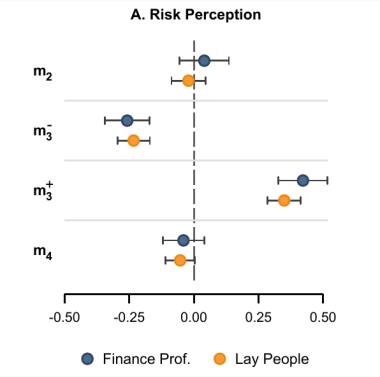
Experimenting on risk perception

- Holzmeister et al Working Paper
- gave description of return from an asset to subjects
- ∼7000 subjects
- including ∼2500 **traders**
- asked to rate **perceived risk of each asset**

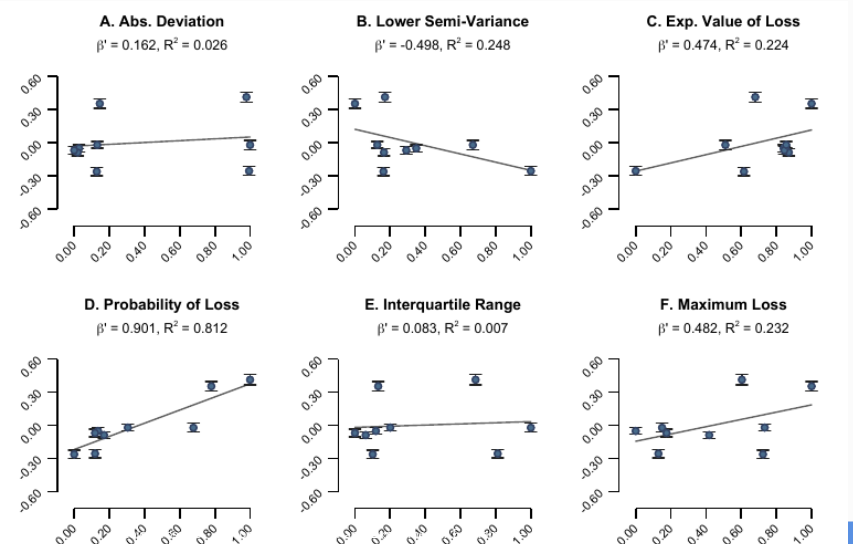
Holzmeister et al: design



results - skewness



results - aggregate risk measures



Theory

- Spiliopoulos & Hertwig: *different* **decision rules** for different contexts
- Schneider and Sutter: **higher moments** matter
- Sunder et al: *curvature of utility* function **not** a valid theory
- **Ergodicity** economics (peters et al): drop EV, use time-means
- ...

Data Analysis

Gender effects

Are women more risk averse than men?

yes, *of course*

statistical significance $\Rightarrow$ stylized fact $\Rightarrow$ fact

Are evolution and biology dictating that women are more risk averse than men? Or is the gender gap in risk aversion an outcome of child-rearing practices? (Bertrand 2011)

Our subject pool is atypical in the sense that the female subjects were generally less risk averse than the male subjects.
(Anderson and Mellor, 2009)

Are women more risk averse than men?

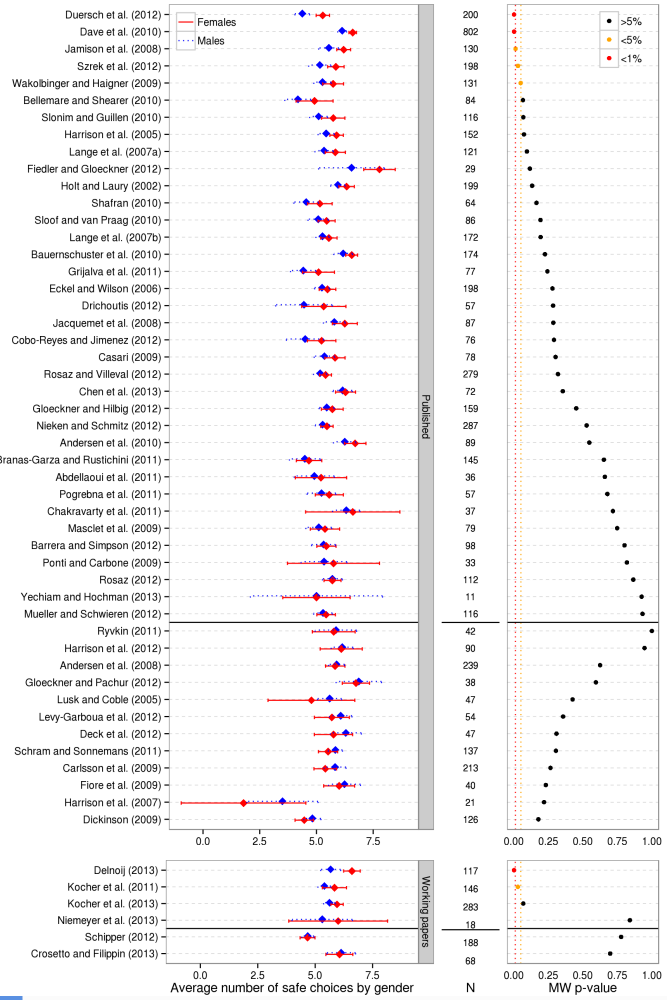
yes *but...*

Are women more risk averse than men?

yes *but...*

...it depends on the **task** used.

Eckel & Grossmann	always
Gneezy & Potters Investment Game	$\sim$always
Bomb Risk Elicitation Task	never
$\sim$ 100 experiments in psychology	about 50%
Holt & Laury	???

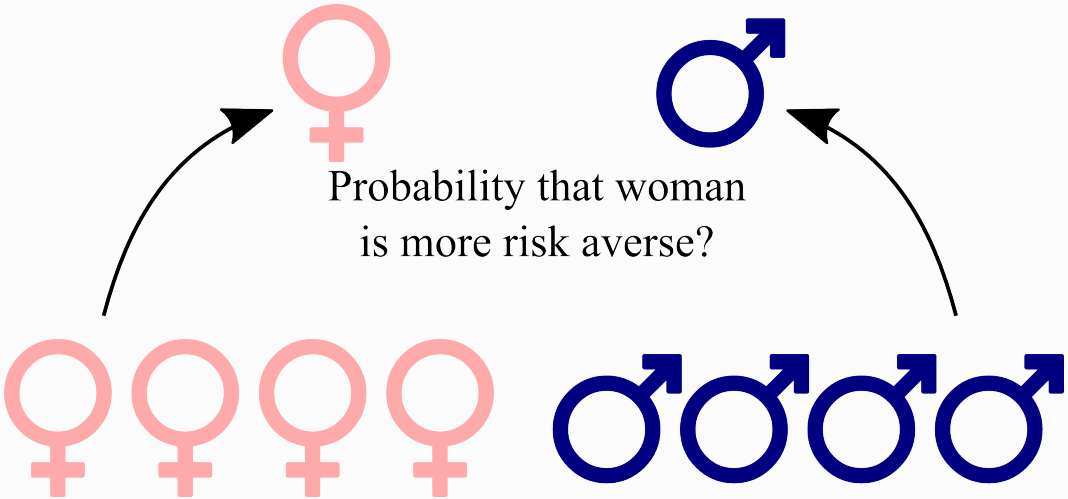

Females
Males
Duersch et al. (2012)
Dave et al. (2010)
Jamison et al. (2008)
Szrek et al. (2012)
Wakolbinger and Haigner (2009)
Bellemare and Shearer (2010)
Slonim and Guillen (2010)
Harrison et al. (2005)
Lange et al. (2007a)
Fiedler and Gloeckner (2012)
Holt and Laury (2002)
Shafran (2010)
Sloof and van Praag (2010)
Lange et al. (2007b)
Bauernschuster et al. (2010)
Grijalva et al. (2011)
Eckel and Wilson (2006)
Drichoutis (2012)
Jacquemet et al. (2008)
Cobo-Reyes and Jimenez (2012)
Casari (2009)
Rosaz and Villeval (2012)
Chen et al. (2013)
Gloeckner and Hilbig (2012)
Nieken and Schmitz (2012)
Andersen et al. (2010)
Branas-Garza and Rustichini (2011)
Abdellaoui et al. (2011)
Pogrebna et al. (2011)
Chakravarty et al. (2011)
Masclet et al. (2009)
Barrera and Simpson (2012)
Ponti and Carbone (2009)
Rosaz (2012)
Yechiam and Hochman (2013)
Mueller and Schwieren (2012)
Ryvkin (2011)
Harrison et al. (2012)
Andersen et al. (2008)
Gloeckner and Pachur (2012)
Lusk and Coble (2005)
Levy-Garboua et al. (2012)
Deck et al. (2012)
Schram and Sonnemans (2011)
Carlsson et al. (2009)
Fiore et al. (2009)
Harrison et al. (2007)
Dickinson (2009)
Published
0.0
2.5
5.0
7.5
200
802
130
198
131
84
116
152
121
29
199
64
86
172
174
77
198
57
87
76
78
279
72
159
287
89
145
36
57
37
79
98
33
112
11
116
42
90
239
38
47
54
47
137
213
40
21
126
>5%
<5%
<1%
0.00
0.25
0.50
0.75
1.00
Delnoij (2013)
Kocher et al. (2011)
Kocher et al. (2013)
Niemeyer et al. (2013)
Schipper (2012)
Crosetto and Filippin (2013)
Working papers
117
146
283
18
188
68
Average number of safe choices by gender
N
MW p-value

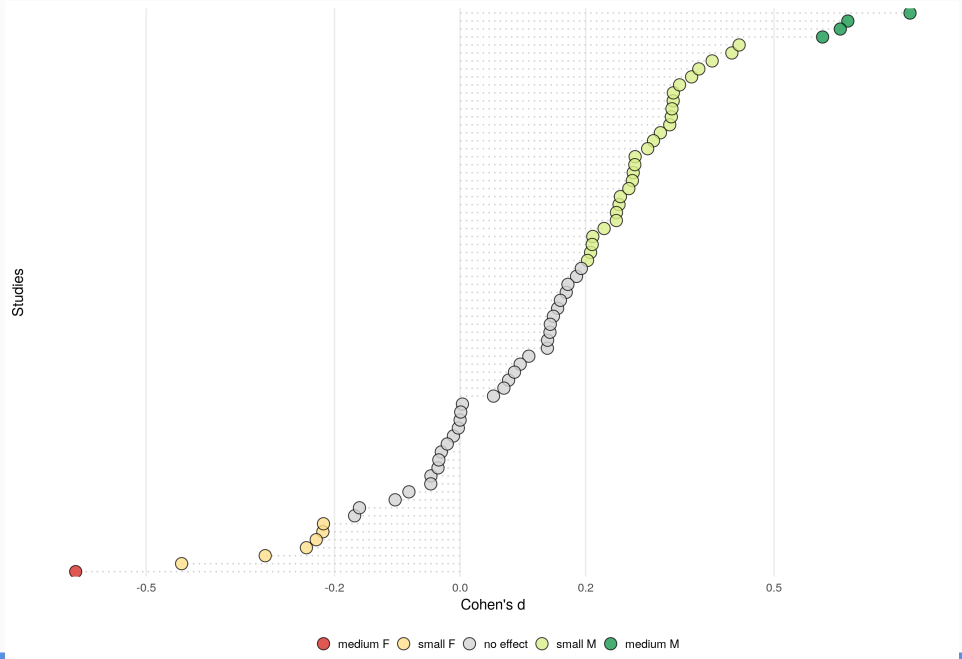
Only **8** out of 63 papers show significant differences

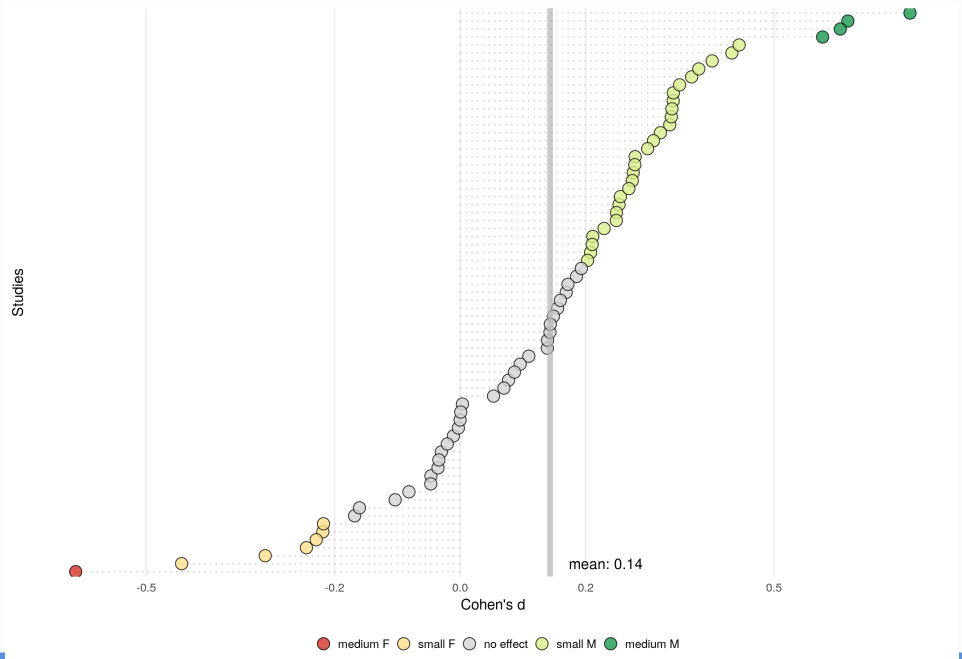
Only **5** out of 54 for published studies

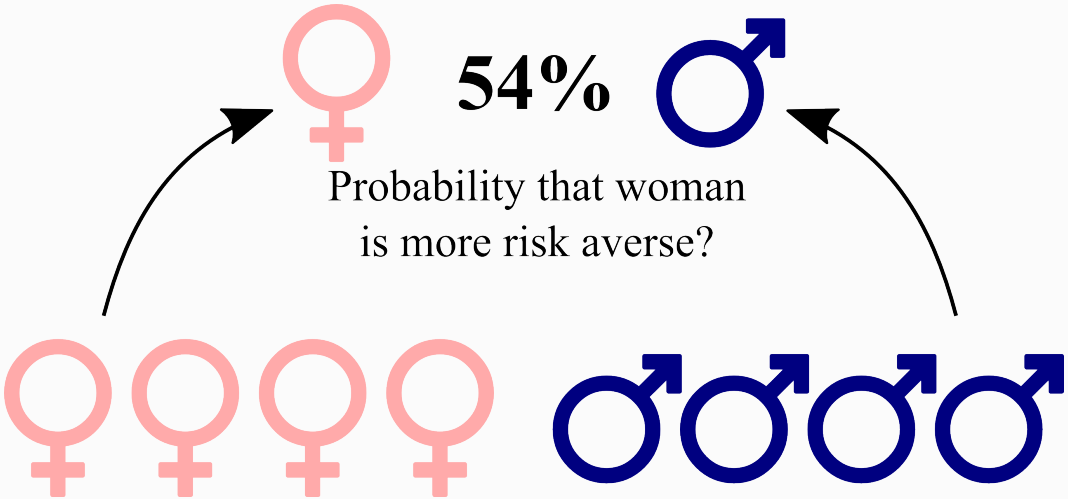
Cohen's **d**

$$d = \frac{\bar{X}_f - \bar{X}_m}{s}$$


Probability that woman
is more risk averse?


Studies
-0.5
-0.2
0.0
0.2
0.5
Cohen's d
medium F
small F
no effect
small M
medium M


Studies
mean: 0.14
-0.5
-0.2
0.0
0.2
0.5
Cohen's d
medium F
small F
no effect
small M
medium M


54%
Probability that woman
is more risk averse?