Eliciting risk attitudes

Methods, internal & external validity, competing approaches

Paolo Crosetto

PSE Frontiers in Beh Econ – Dec 5^{th}, 2025

Let's play!

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

What's on the **menu**?

1. Risk **attitudes**: what are they? Why do they matter?
2. Risk **elicitation**: How do we measure risk attitudes?
3. **Does it work?** A meta-analysis of elicited risk attitudes
4. Proceeding by **FIAT** – Fix it Again, Tony!
   - measurement error
   - task specific bias
5. **Changing** paradigm
   - layers of uncertainty
   - risk perception
   - cognitive turn

1. What are risk attitudes?

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

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 in 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**

How do people make decisions given a probability distribution over outcomes?

Key assumptions:

- Constant over time (preferences are hardwired, *in a sense*)
- Constant across domains.
- Further (usually parametric) assumptions on the utility model (EUT, PT...)

Different layers of uncertainty

Risk

<table>
<tr><td>10</td><td>100</td><td>1000</td></tr>
<tr><td>50%</td><td>10%</td><td>40%</td></tr>
</table>

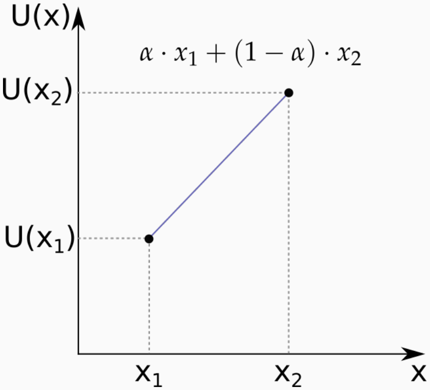
Ambiguity

<table>
<tr><td>10</td><td>100</td><td>1000</td></tr>
<tr><td>50%</td><td colspan="2">50%</td></tr>
</table>

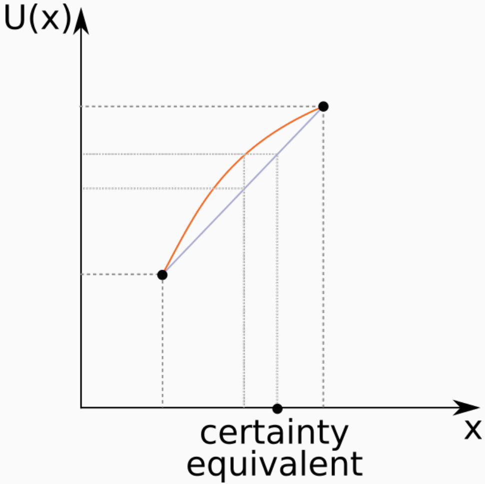
Deep (Knightian) uncertainty

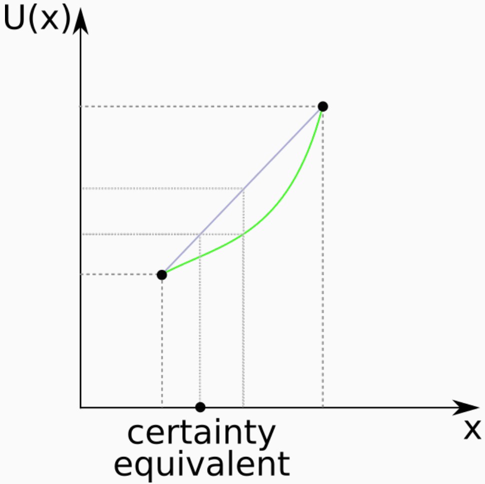
<table>
<tr><td>10</td><td>??</td><td>1000</td><td>??</td><td>...</td></tr>
<tr><td>50%</td><td colspan="2">30%</td><td colspan="2">??</td></tr>
</table>

The expected utility framework



The expected utility framework





The state of the art in economics

Risk formally defined as uncertainty over outcomes

Focus on **decontextualized tasks**

- **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 $\iff$ theory)

2. Measuring risk attitudes

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)

We will need some assumptions...

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

. . . and some methodological choices

Risk attitudes are **elicited** in different ways:

- infer from real world data **vs.** build ad-hoc choices
- survey measures via questionnaires **vs.** incentivized tasks
- binary choices + structural model **vs.** structured choice lists
- elicitation by descrption **vs.** by experience

. . . and some methodological choices

Risk attitudes are **elicited** in different ways:

- infer from real world data **vs.** build ad-hoc choices
- survey measures via questionnaires **vs.** incentivized tasks
- binary choices + structural model **vs.** structured choice lists
- elicitation by descrption **vs.** by experience

Focus on **ad-hoc structured tasks** elicited by **description**:

Risk Elicitation Tasks (**RET**)

A good RET should be...

Accurate: theoretically sound, not *ambiguous*, unbiased...

Relevant: predictive of real-life behavior

Handy: easy to implement, understand, deploy (lab, field)

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

Clean: with low noise and allowing control

…and of course there are trade-offs

Accurate
actually representing true preferences

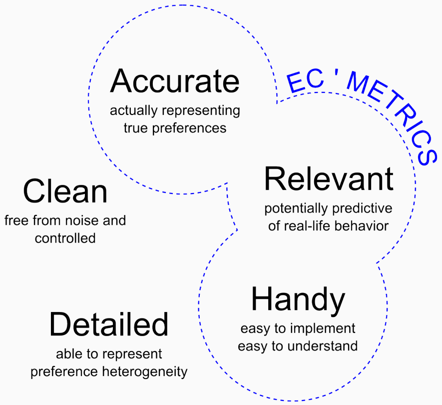
Clean
free from noise and controlled

Relevant
potentially predictive of real-life behavior

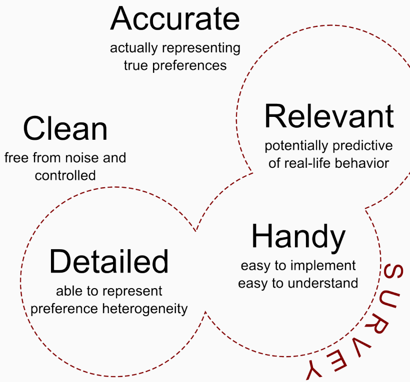
Handy
easy to implement
easy to understand

Detailed
able to represent preference heterogeneity

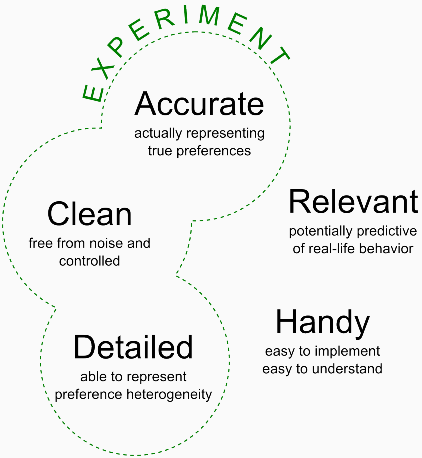
...and of course there are trade-offs



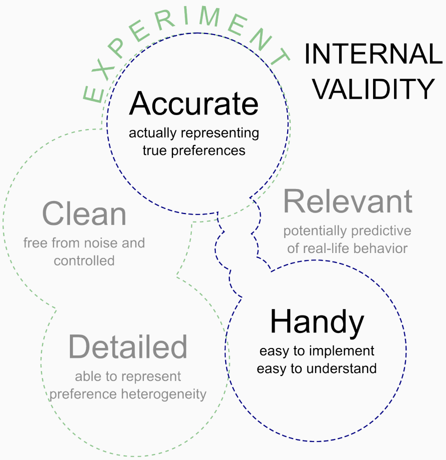
...and of course there are trade-offs



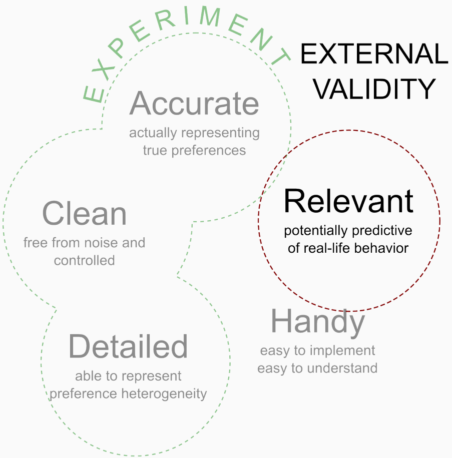
...and of course there are trade-offs



Experiments: internal validity



Experiments: external validity



Some RETs

Tasks from psychology

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)

Tasks from psychology: pros and cons

Pros

- external validity
- real worlds behavior
- losses

Cons

- no or little theory
- generalizability dubious

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.

Questionnaires: 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)?

Early tasks from economics: auctions

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

- you bid for an object worth 10 euro
- against a computerized opponent $\sim U[0;10]$

Early tasks from economics: auctions

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

- you bid for an object worth 10 euro
- against a computerized opponent $\sim U[0;10]$

What should you do?

- Your earnings are

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

- bid 1 $\Rightarrow$ get 9 with probability 10%, and so on...
- optimal strategy if risk neutral: bid $= 10/2 = 5$
- if *risk averse*: bid *more*.

Early tasks from economics: auctions

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

- you bid for an object worth 10 euro
- against a computerized opponent $\sim U[0; 10]$

What should you do?

- Your earnings are

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

- bid 1 $\Rightarrow$ get 9 with probability 10%, and so on...
- optimal strategy if risk neutral: bid $= 10/2 = 5$
- if *risk averse*: bid *more*.

risk aversion = overbidding

Eliciting risk attitudes via auctions: 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

The workhorse of economics RETs: 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 €

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**

Holt and Laury: 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 €

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.

Binswnanger: 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		

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.

CEPL: 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}

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 $E(\text{risky}) = 1.25 > 1$.

Gneezy and Potters: pros and cons

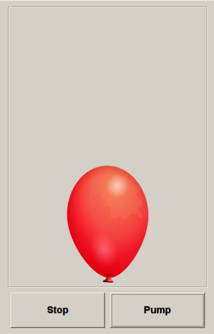
Pros

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

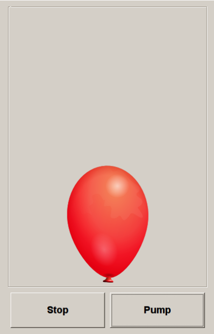
Cons

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

Inflating a balloon with increasing probability of explosion



Inflating a balloon with increasing probability of explosion



Risk-neutral should stop halfway – but not enough information

Balloon: pros and cons

Pros

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

Cons

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

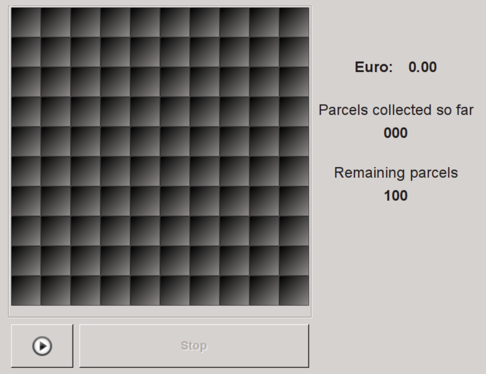


Figure 1: The BRET interface at the start of the experiment

BRET: interface

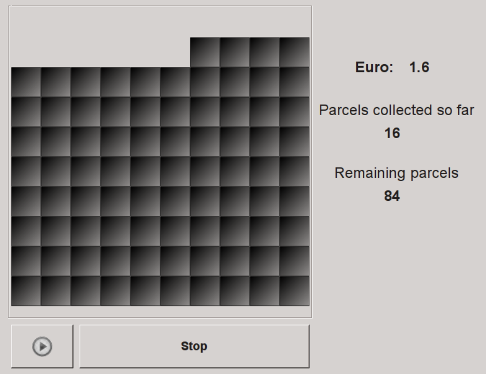


Figure 2: The BRET interface after 16 seconds

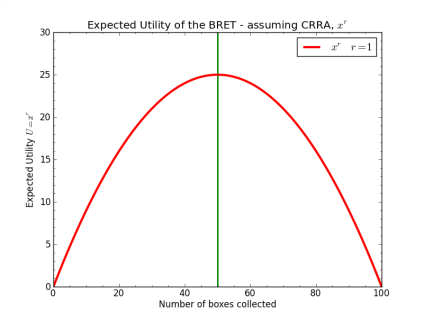
BRET: under the **hood**

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

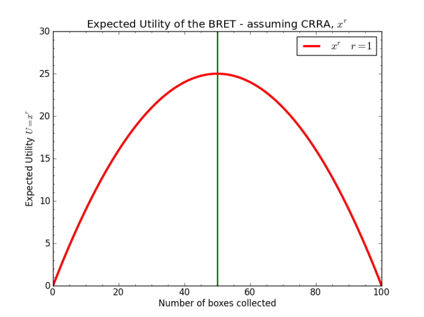
$$L^k = \begin{cases} 0 & \frac{k}{100} \\ \\ k & \frac{100-k}{100} \end{cases}$$

- 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



BRET: solution for the expected value maximizer

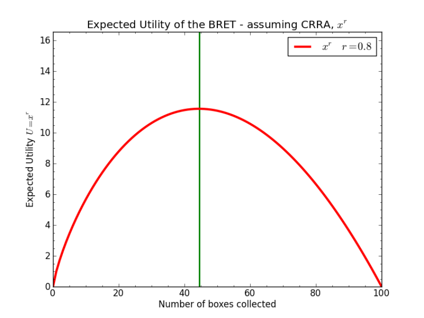


The expected value is maximized at $k^* = 50$.

Assuming a power CRRA utility function x^r, the optimal stopping point is:

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

BRET: Risk averse subject

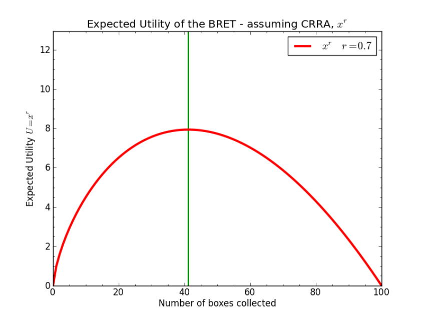


The expected value is maximized at $k^* = 50$.

Assuming a power CRRA utility function x^r, the optimal stopping point is:

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

BRET: Risk averse subject

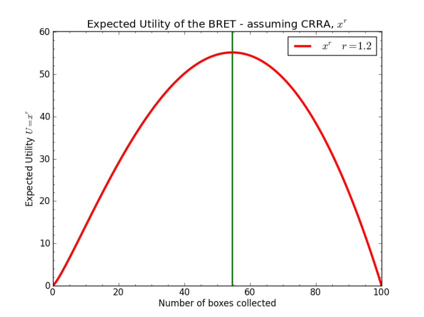


The expected value is maximized at $k^* = 50$.

Assuming a power CRRA utility function x^r, the optimal stopping point is:

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

BRET: Risk lover subject

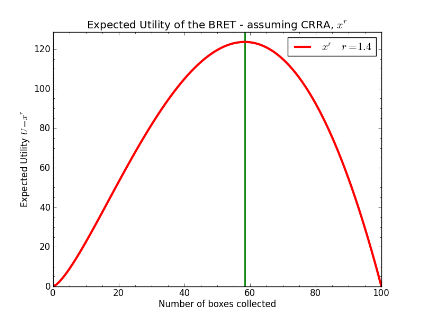


The expected value is maximized at $k^* = 50$.

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:

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

BRET: 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?)

3a. Do measures work? your data

Let's have a look at your data...

Head over to R!

3b. Do measures work?

meta-anaysis

Our forefathers: 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*

Explore the data!

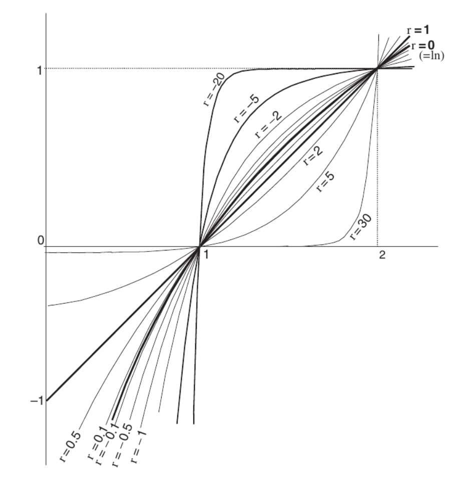
Meta-Analysis of Risk Elicitation Tasks

- data from experiments worldwide
- convergent & predictive validity
- preregistration on OSF
- data & scripts on gitHub
- live exploration on shiny app

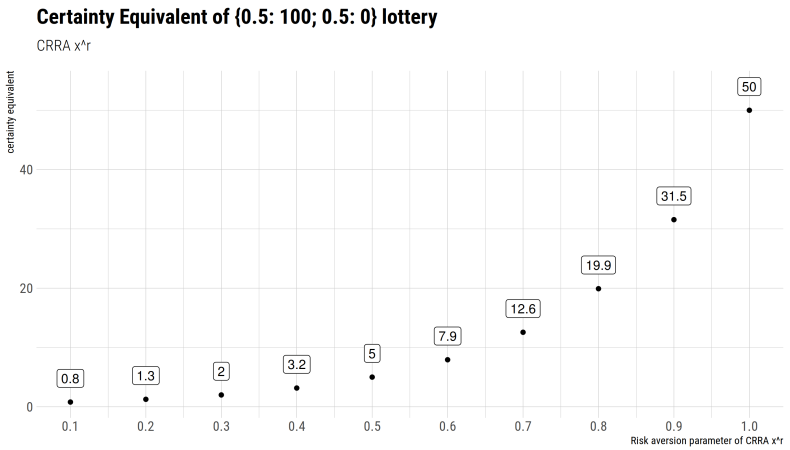
METARET assumptions: CRRA (à la Wakker)

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

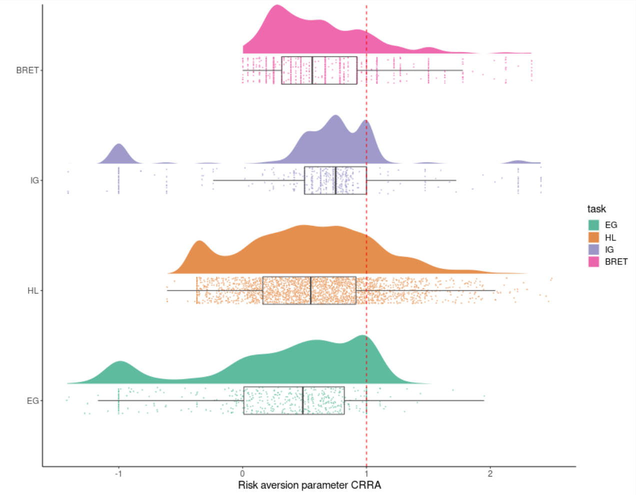
- simple
- captures risk aversion
- makes different tasks comparable



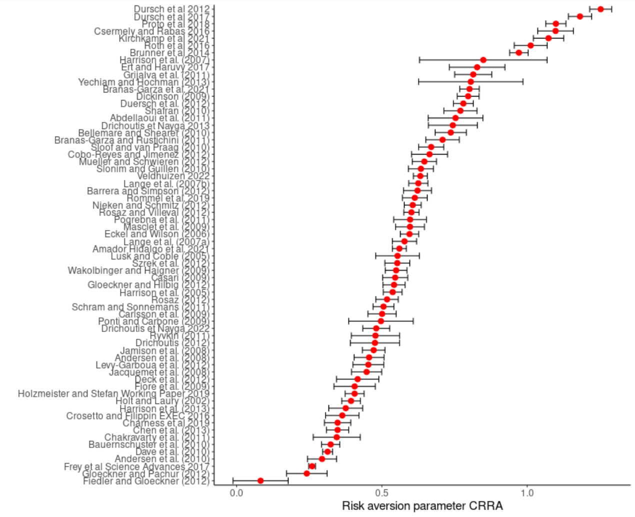
How to interpret parameter differences



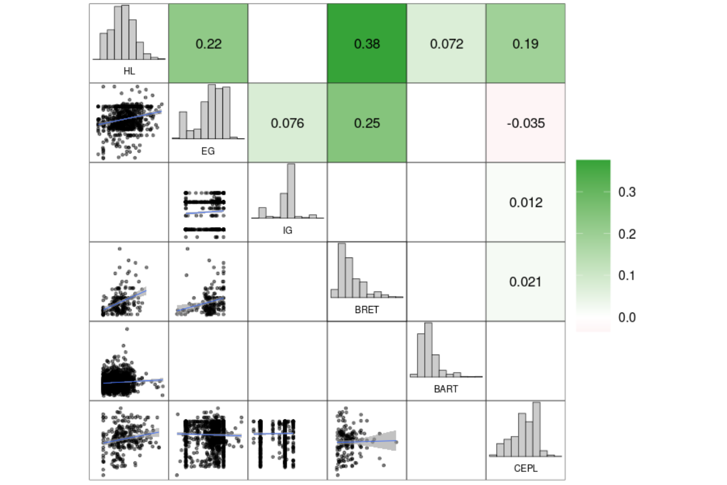
Low consistency across tasks



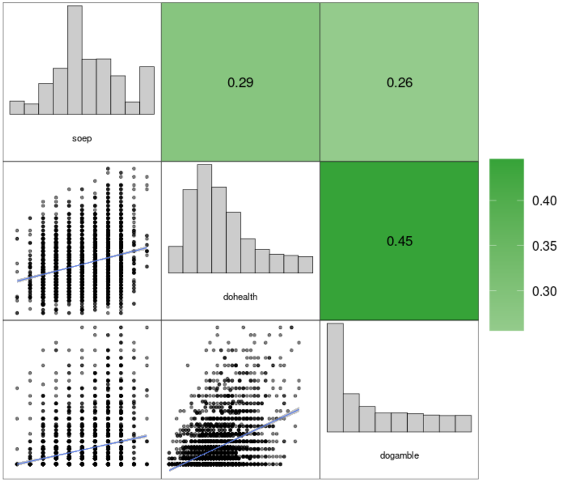
Low consistency within tasks



Convergence: tasks



Convergence: questionnaires



Tasks $\Longleftrightarrow$ Questionnaires

Convergent validity: more evidence

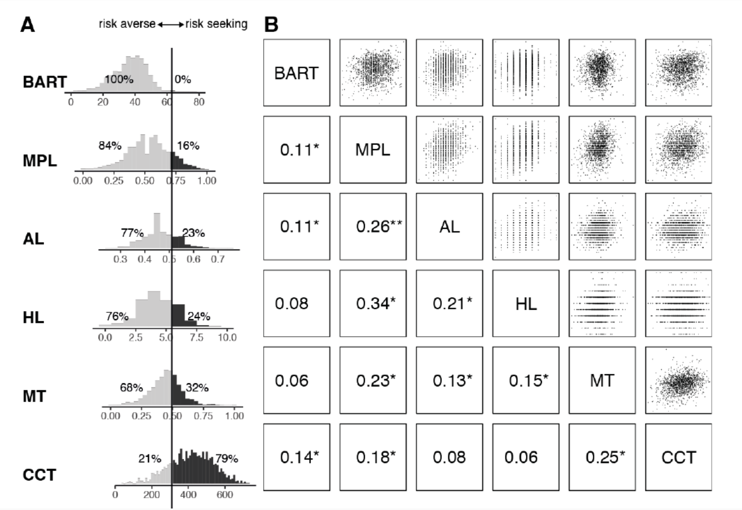


Figure 3: Pedroni et al. Nature Human Behavior 2017

Predictive validity: more evidence

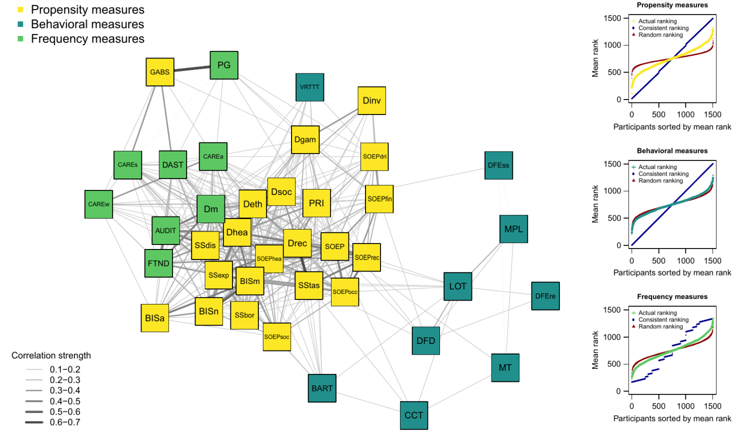


Figure 4: Frey et al. Science Advances 2017


WELL THIS SUCKS

4. Fix It Again, Tony!

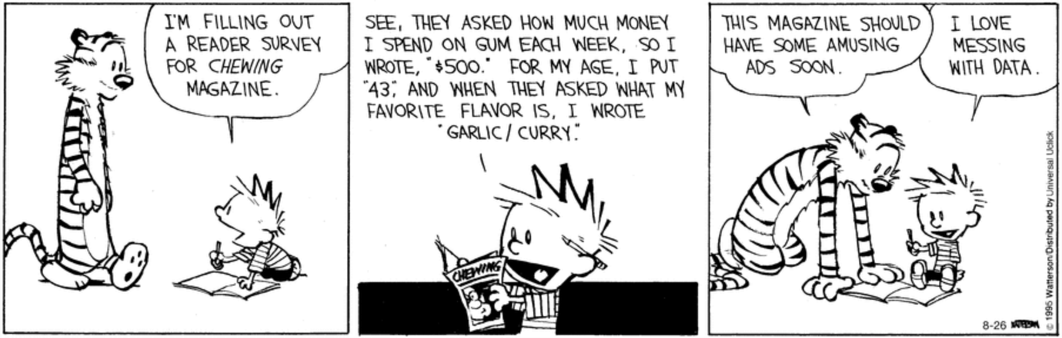
FIAT

Potential fixes

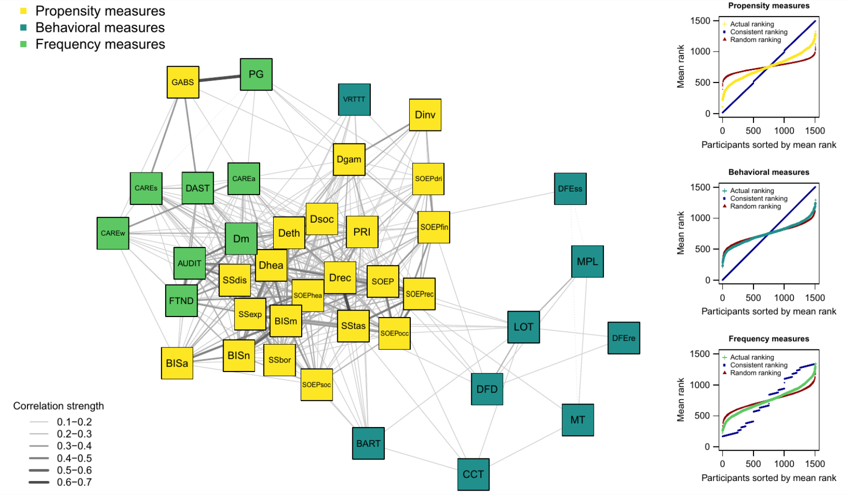
- hypothetical bias
- measurement error
- task specific bias

Hypothetical Bias

Hypothetical bias: subjects love messing with data



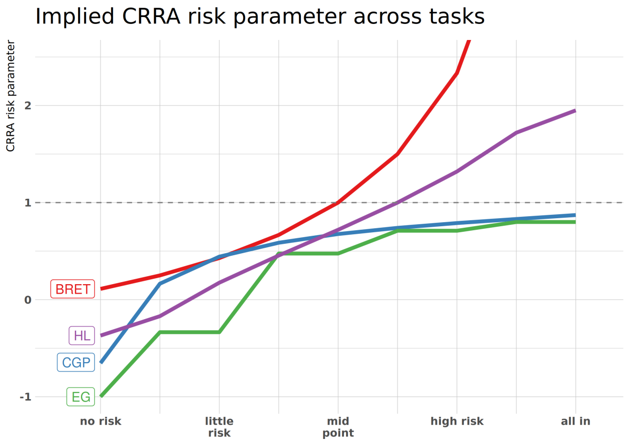
But again...



Propensity measures
Behavioral measures
Frequency measures
GABS
PG
VRTTT
Dinv
Dgam
CAREs
DAST
CAREa
SOEPdri
DFEss
CAREw
Dsoc
Deth
PRI
SOEPfin
Dm
MPL
AUDIT
Dhea
SSdis
Drec
SOEP
SOEPrec
FTND
SOEPhea
LOT
SSexp
BISm
SStas
SOEPocc
DFEre
BISa
BISn
SSbor
SOEPsoc
DFD
MT
BART
CCT
Correlation strength
0.1–0.2
0.2–0.3
0.3–0.4
0.4–0.5
0.5–0.6
0.6–0.7
Propensity measures
Behavioral measures
Frequency measures
Actual ranking
Consistent ranking
Random ranking
Mean rank
Participants sorted by mean rank
0
500
1000
1500

Task-specific error

Mapping choices to r: risk levels

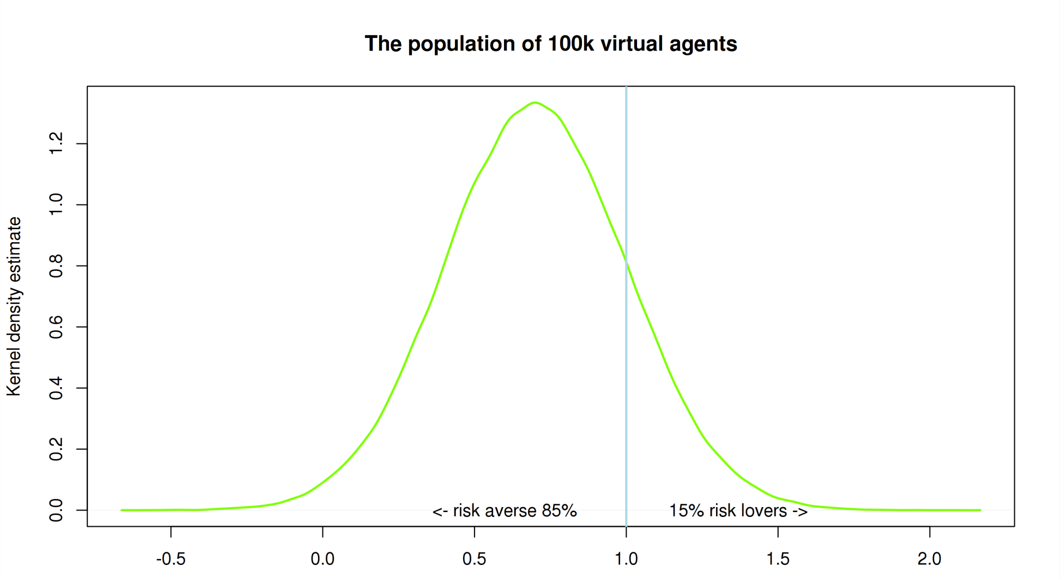


Clean test: simulations

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.

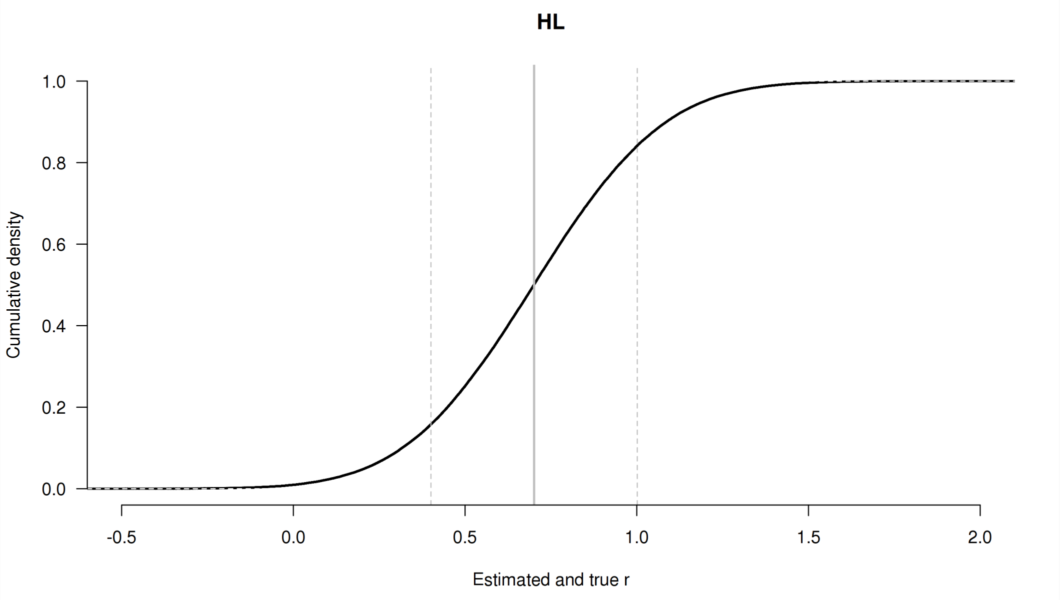
Putting the cart before the horse: simulations

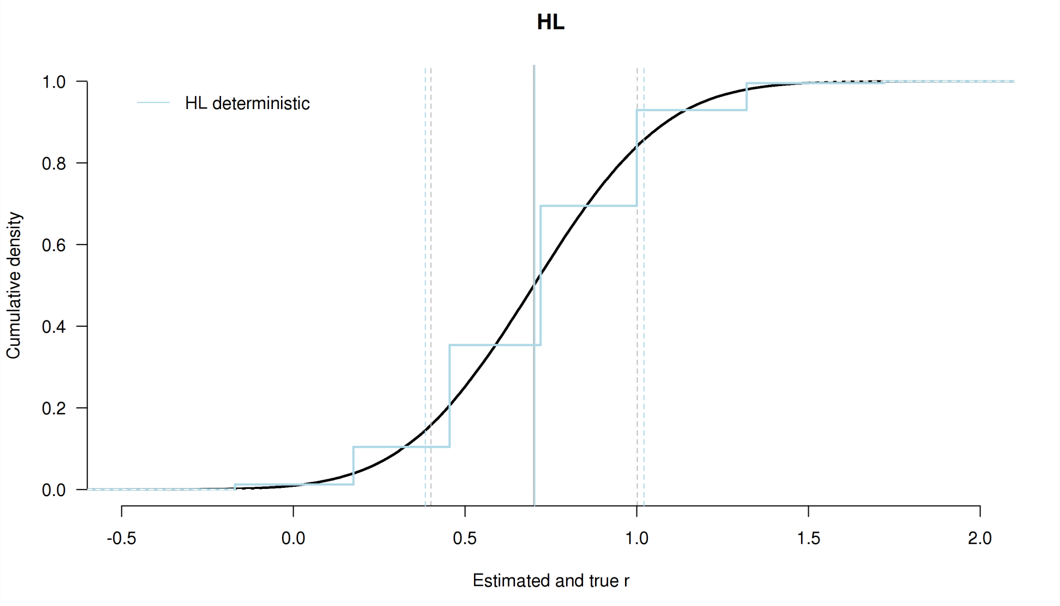


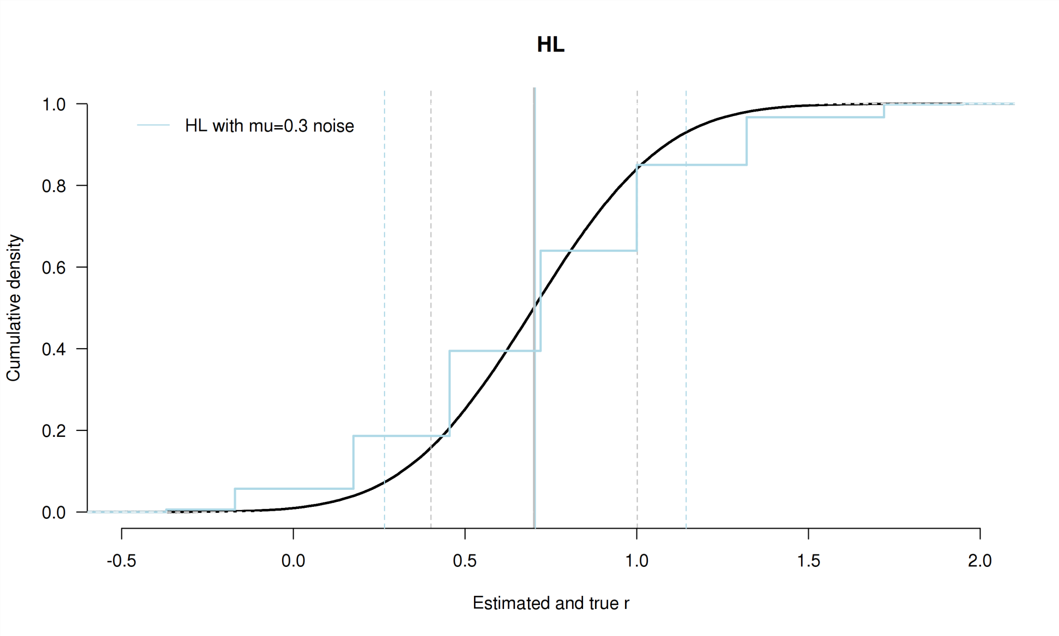
Simulations, II

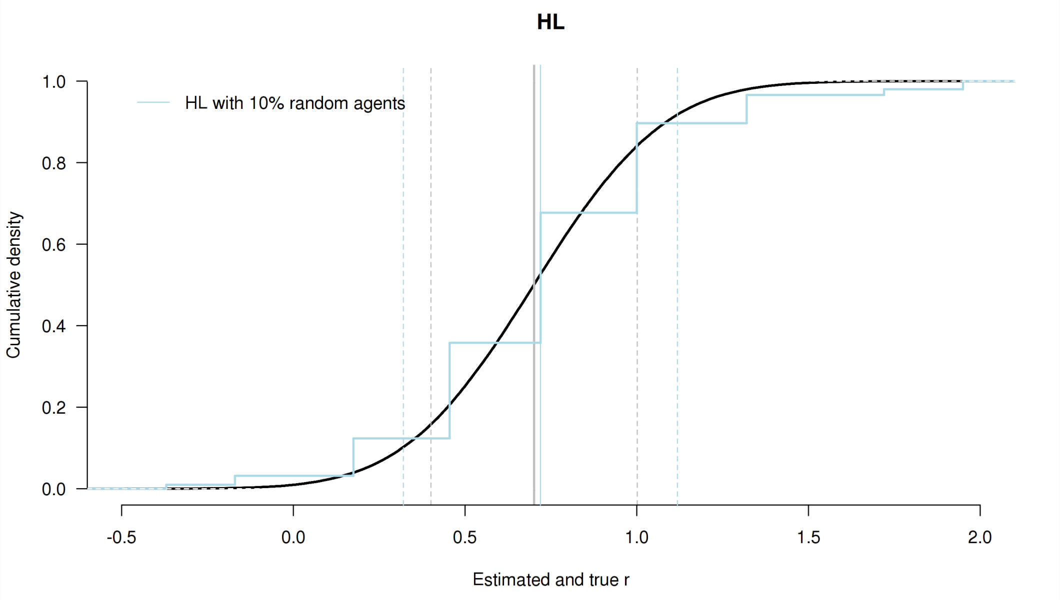
Three types of simulations:

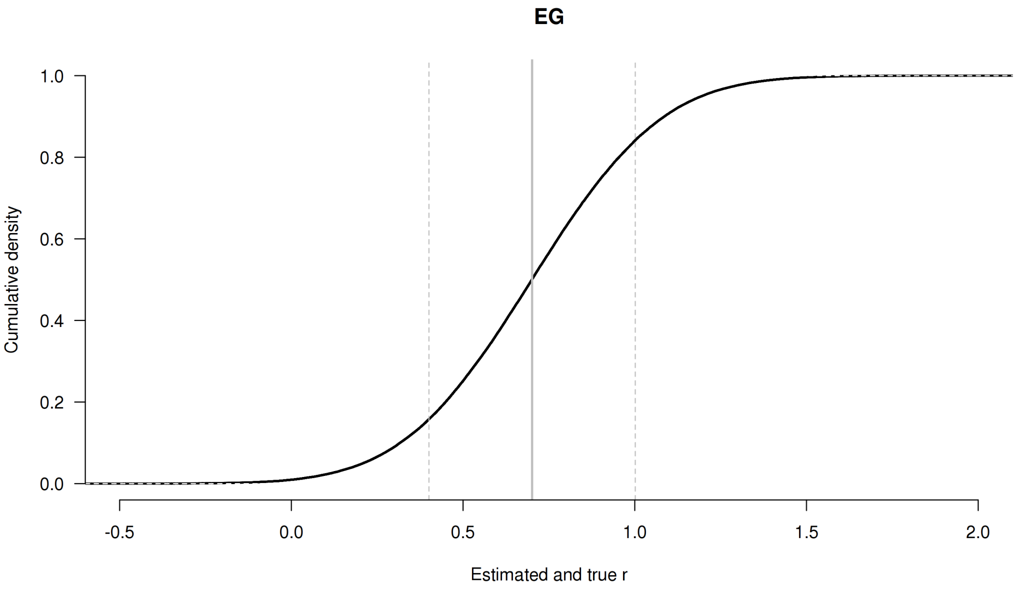
1. Deterministic: virtual subjects play according to their **true r**
2. Random parameter model:
   - for each agent, $r_n = r + \varepsilon, \varepsilon \sim N(0, \mu)$
   - that is, the agent deviates from her true preferences with a white noise
   - $\mu = 0.3$ or 0.6
3. Trembling hand: behaviorally 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

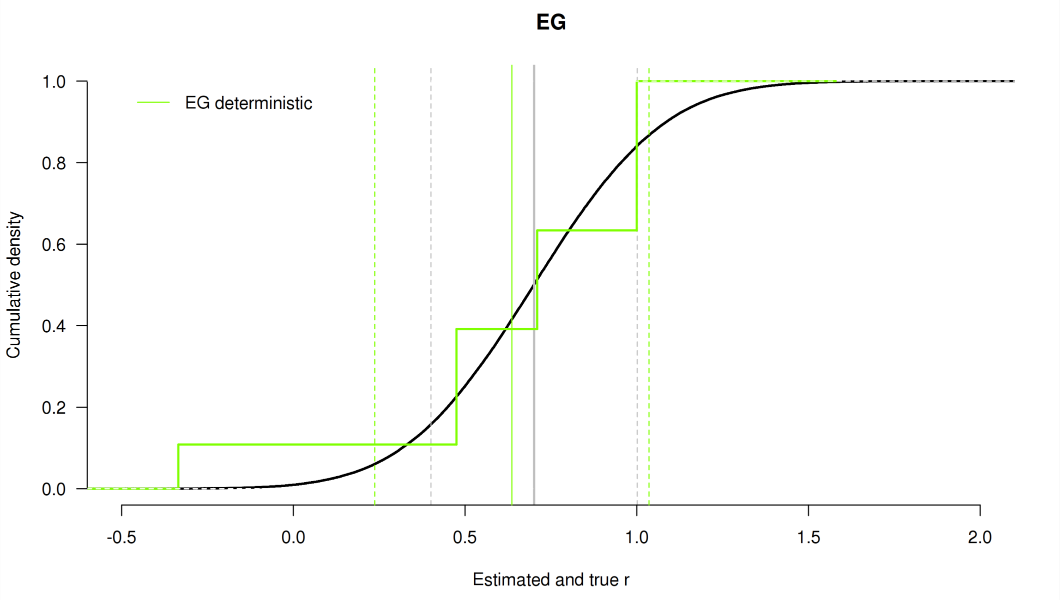

HL
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative density
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

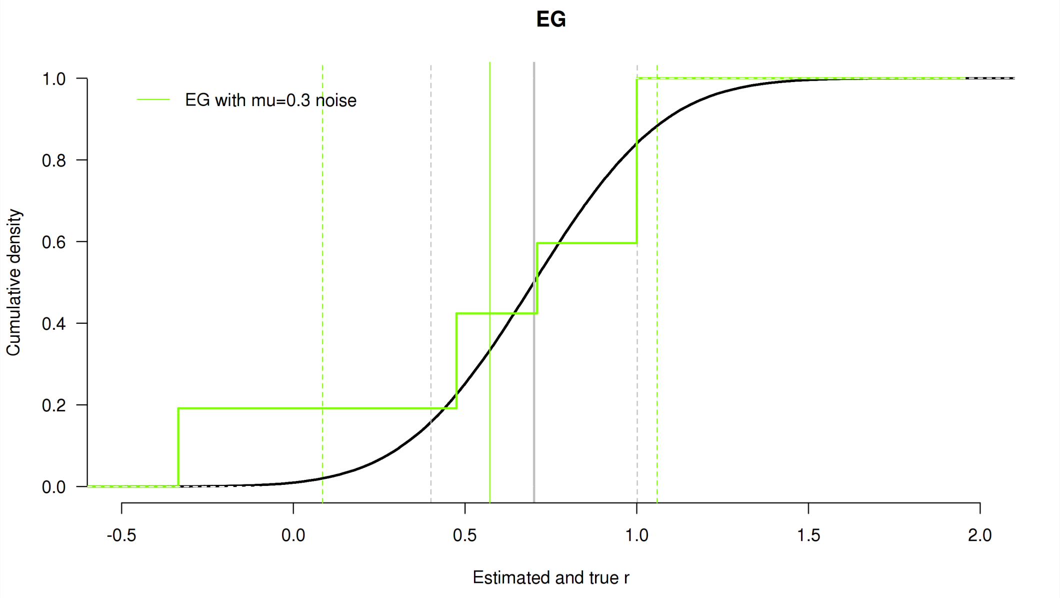

HL
HL deterministic
Cumulative density
0.0
0.2
0.4
0.6
0.8
1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

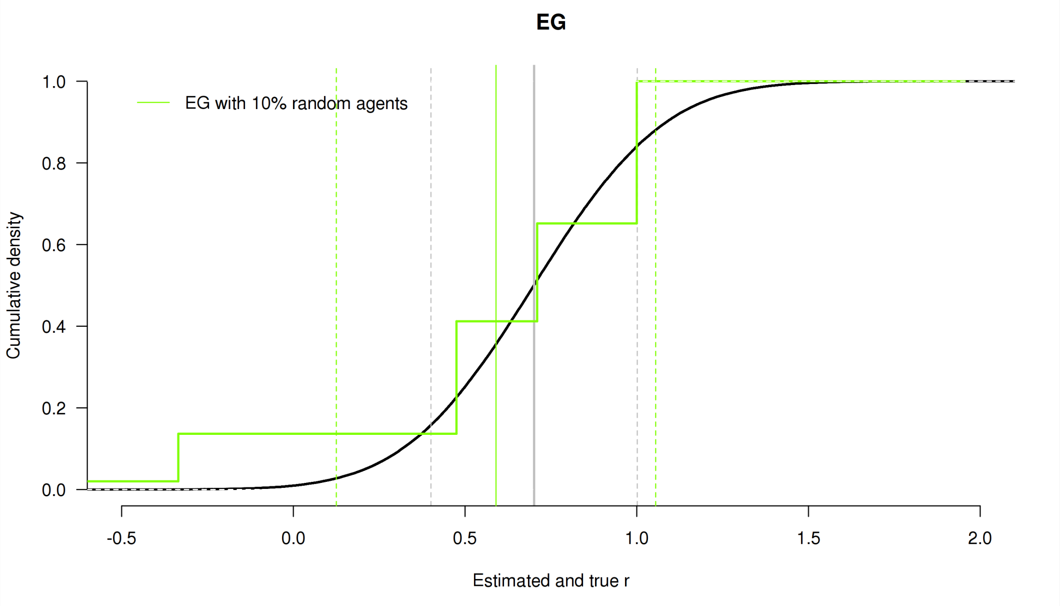

HL
HL with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


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

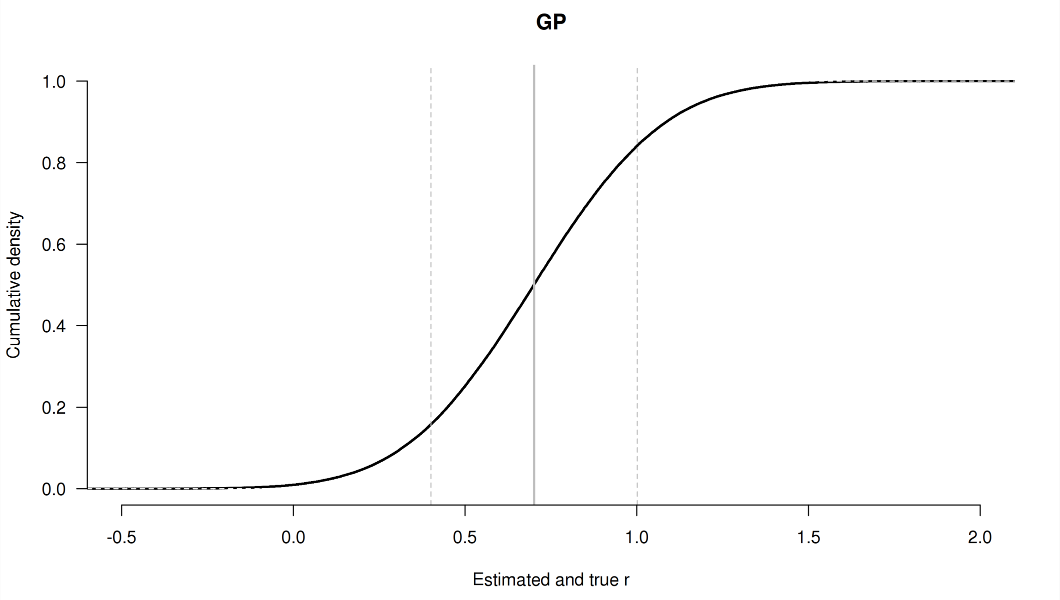

EG
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative density
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

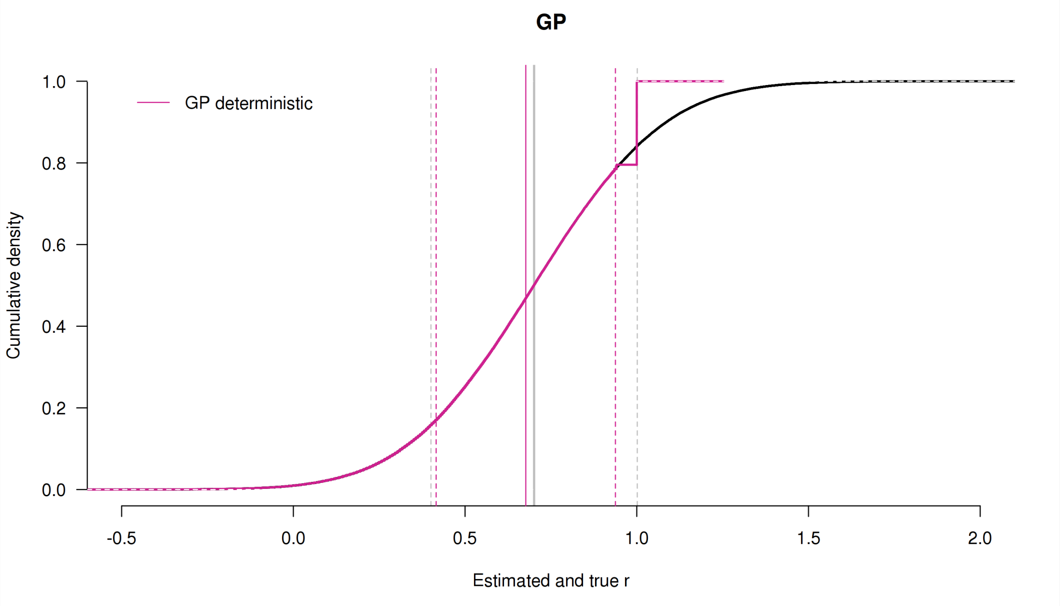

EG
EG deterministic
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative density
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

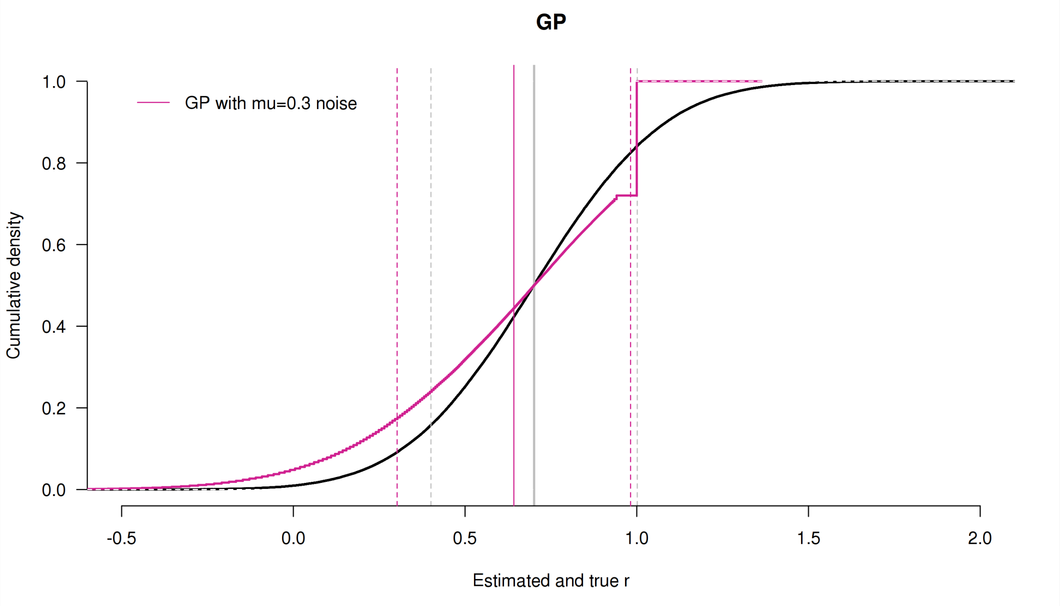

EG
EG with mu=0.3 noise
Cumulative density
0.0
0.2
0.4
0.6
0.8
1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

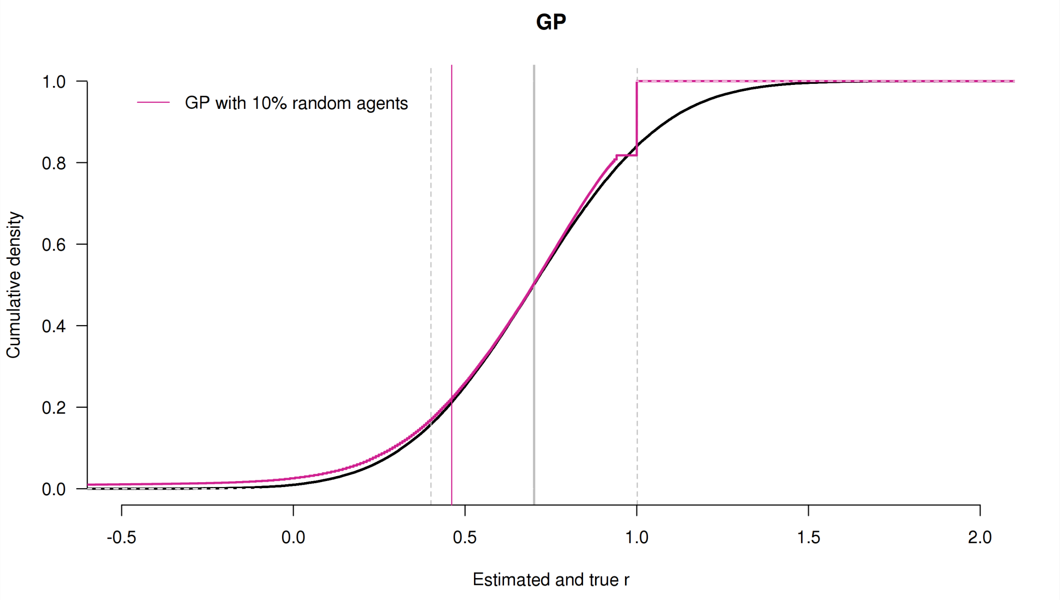


EG
EG with 10% random agents
Cumulative density
0.0
0.2
0.4
0.6
0.8
1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

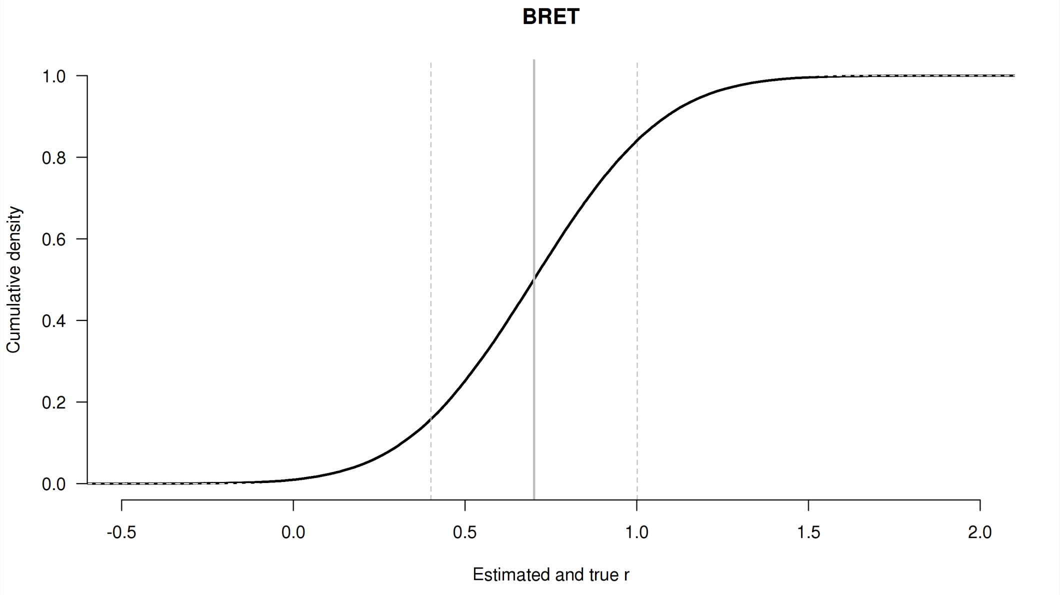

GP
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative density
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

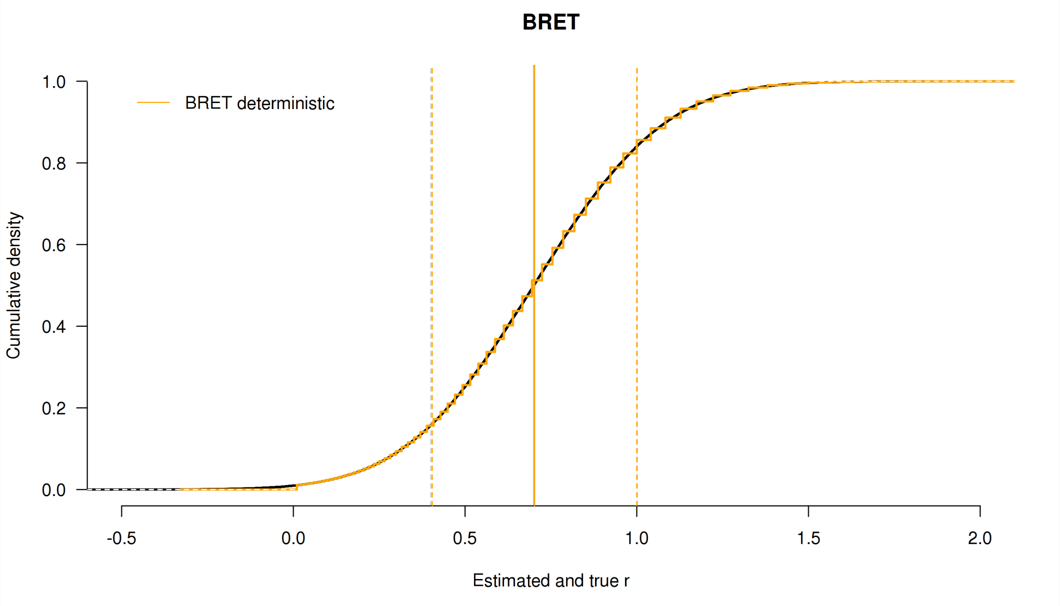

GP
GP deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

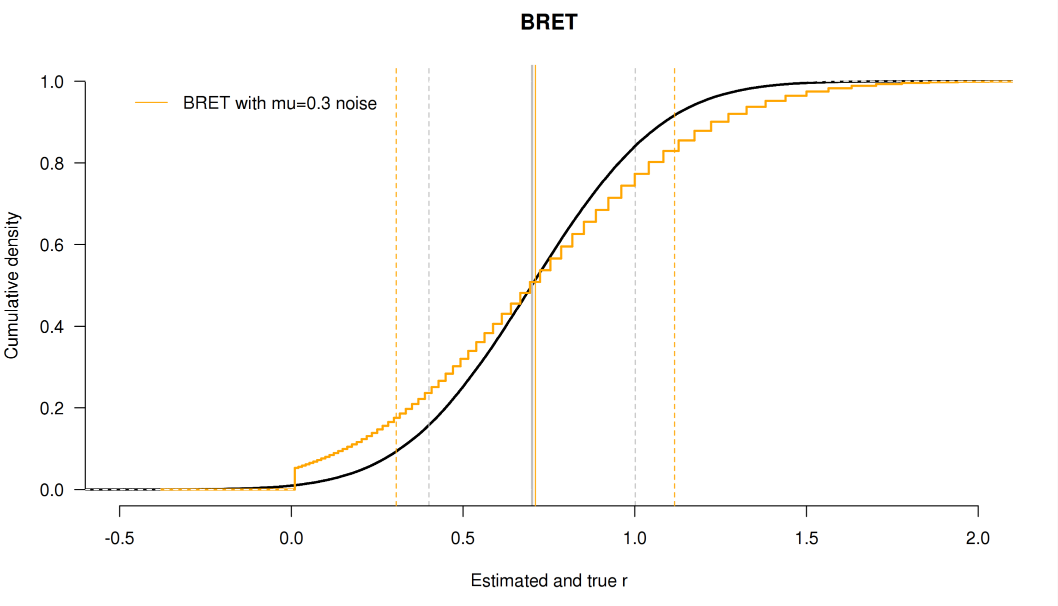

GP
GP with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r



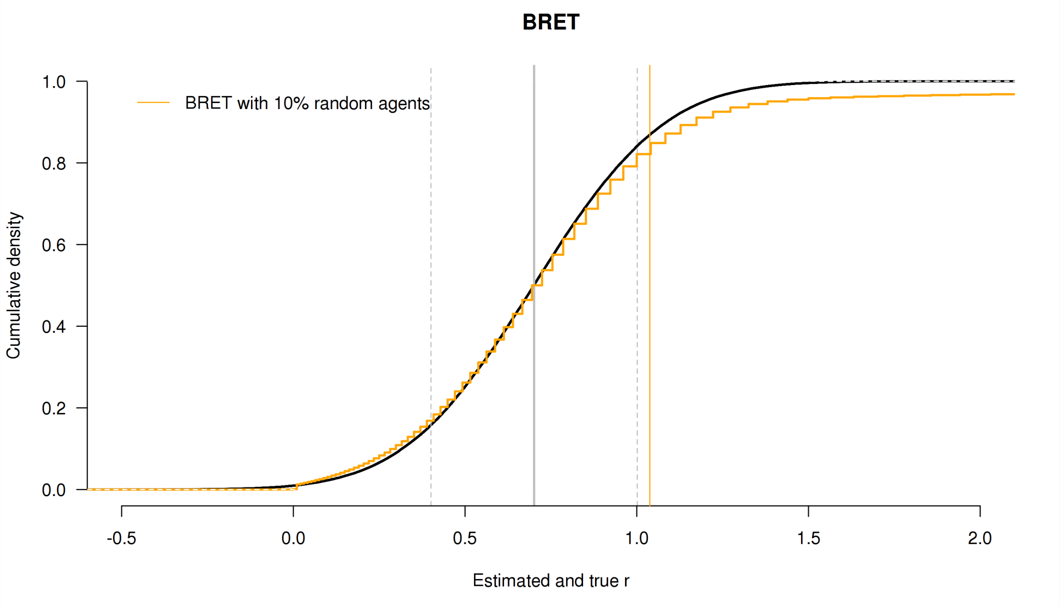
GP
GP with 10% random agents
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative density
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


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
1.0
0.8
0.6
0.4
0.2
0.0
Cumulative density
-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

Does all this explain all measured 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?

Pure noise: measurement error

Noisy preferences lead to measurement error. How do we fix it?

The experiment-intensive way:

Average over different measures / questionnaires

The econometrics-intensive way:

Structural modeling and get estimate + theory + noise

The elbow-grease way: averaging over tasks

For RETs:

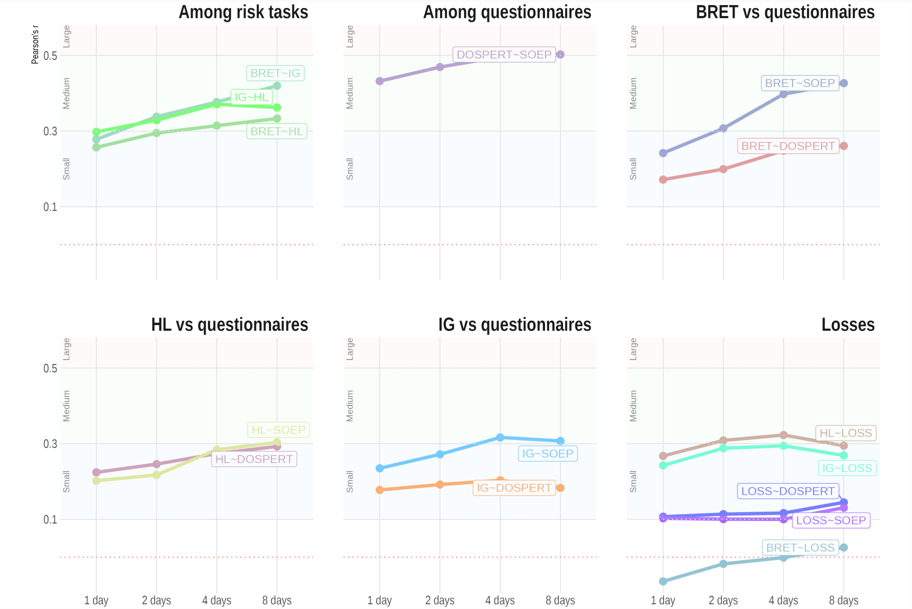
Menkhoff & Sakha, *Estimating risky behavior with multiple-item risk measures* , Jo Econ Psy 2017

For questionnaires:

Beauchamp et al., *The psychometric and empirical properties of measures of risk preferences*, JRU 2017

For both:

Crosetto et al, *Measurement Error in Risk Elicitation*, WP 2025 (maybe)



Among risk tasks
Among questionnaires
BRET vs questionnaires
HL vs questionnaires
IG vs questionnaires
Losses
Pearson's r
Large
Medium
Small
0.5
0.3
0.1
BRET~IG
IG~HL
BRET~HL
DOSPERT~SOEP
BRET~SOEP
BRET~DOSPERT
HL~SOEP
HL~DOSPERT
IG~SOEP
IG~DOSPERT
HL~LOSS
IG~LOSS
LOSS~DOSPERT
LOSS~SOEP
BRET~LOSS
1 day
2 days
4 days
8 days

The stat-intensive way: 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

Explaining structural modeling: Holt and Laury

	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 €

Explaining structural modeling: Holt and Laury

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

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

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

Explaining structural modeling: Holt and Laury

- 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 to the model

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**

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

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

You can play around with error structures

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**

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

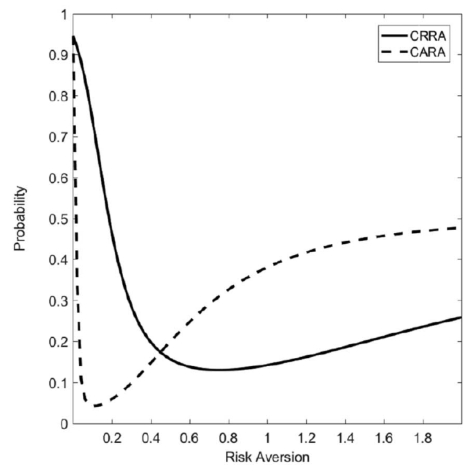
MLE, some results (Crosetto & Filippin, ExEc 2015)

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

Why you should never run a RUM

- P(safe) monotonic increasing in risk aversion
- ... but it doesn't!
- Why?
  - $\lim_{x \to -\infty} EU(L) - EU(R) = 0 < \varepsilon$
  - Working on ΔEU assumes cardinality

Don't run RUMs



Apesteguia and Ballester, JPE 2018

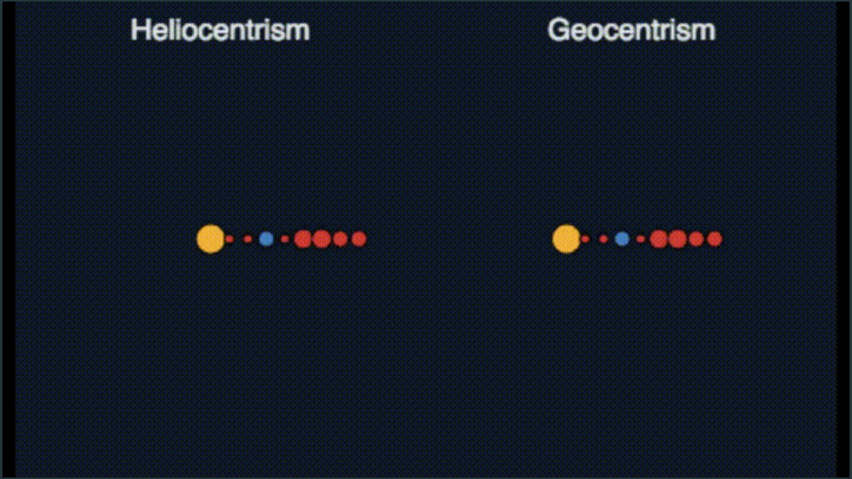
Potential fixes: how are we doing?

- ~~hypothetical bias~~
- measurement error: helps marginally
- task specific bias: helps marginally


WELL THIS SUCKS

5. Deeper fixes: changing paradigm

Are we looking at the problem the right way?



Potential changes of paradigm

- uncertainty layers: risk, ambiguity, or deep uncertainty?
- risk perception
- have we got the right theory? A cognitive turn

Layers of uncertainty: risk, ambiguity, deep...

Remember? Different layers of uncertainty

Risk

10	100	1000
50%	10%	40%

Ambiguity

10	100	1000
50%	50%	

Deep (Knightian) uncertainty

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

Have we got the right representation of risk?

In the lab: ”risk”

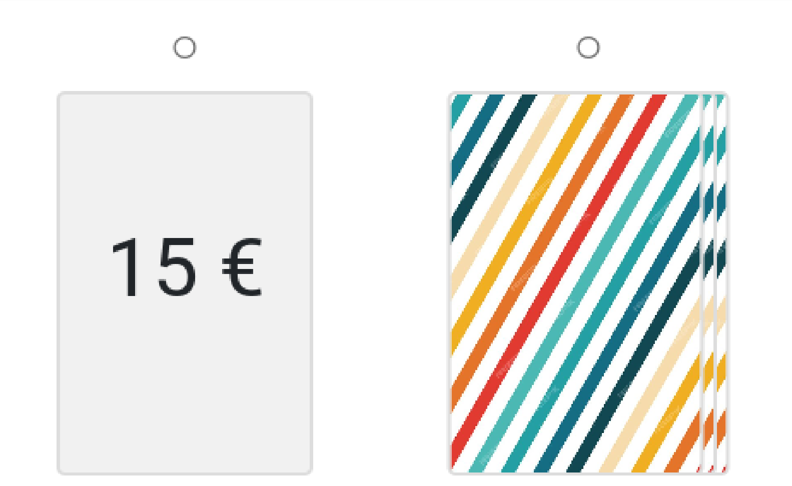
- known probabilities
- known set of outcomes
- no surprises
- learn by description
- small stakes
- no losses

Out of the lab: ”risk”

- fuzzy probabilities
- fuzzy set of outcomes
- surprises
- learn by experience
- high stakes
- losses

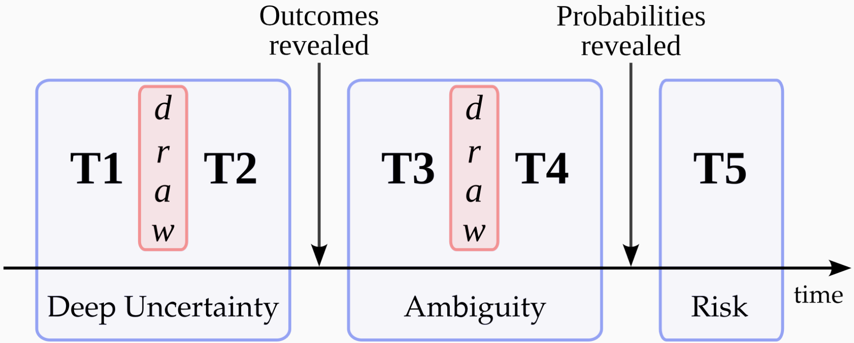
Quite the gap to mind – and bridge

The simplest possible task: binary choice, safe vs risky



"deck contains up to 6 different positive or negative values"

Repeated choices: more information (sampling + description)



Deep U: probabilities & outcomes *unknown*

Ambiguity: probabilities *unkown*, outcomes *known*

Risk: probabilities & outcomes *known*

Our external validity setup

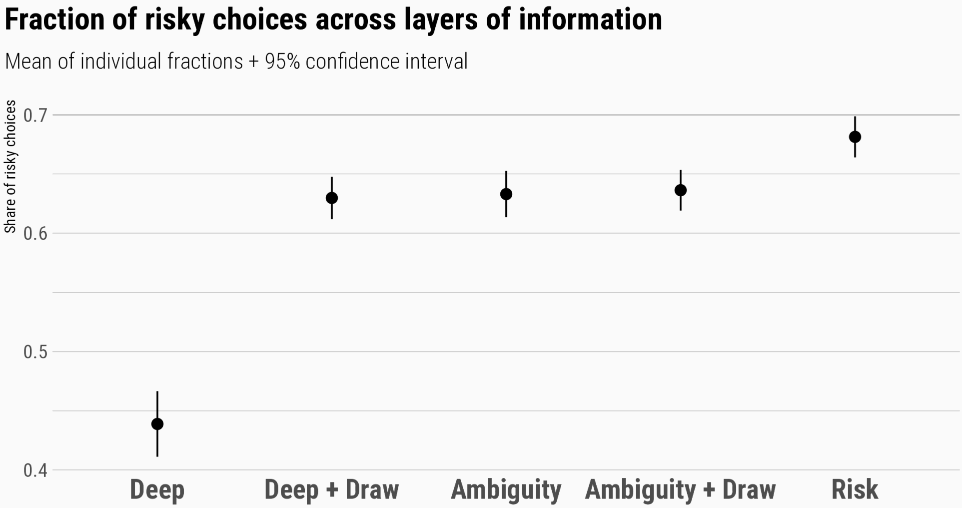
The Daily Reconstruction Method

Anonymized, self-reported list of daily active decisions under risk, irrespective if the risk was taken or avoided, filled at home every evening over **14** *days.*

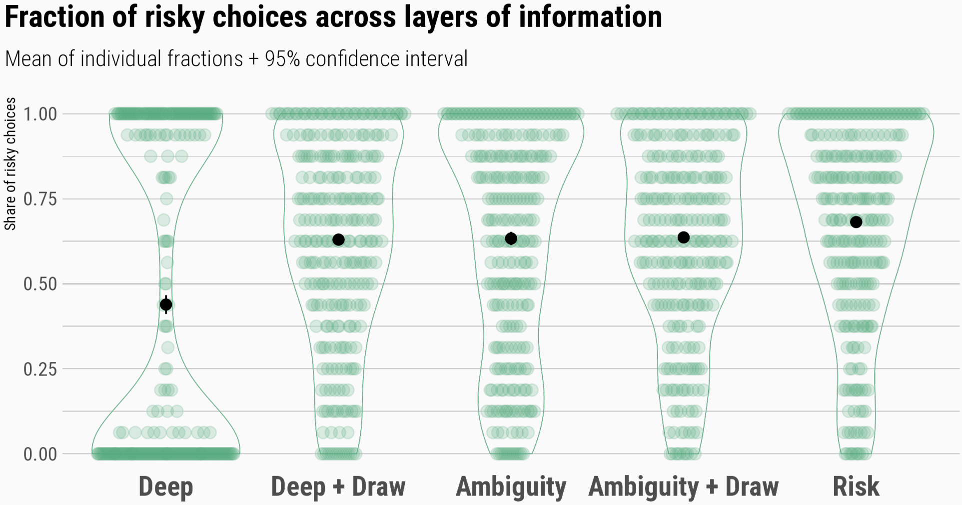
For each activity:

- **Domain**: health, safety, recreation, drive, financial, ethics, social
- **Perception**: of the risk avoided or taken (-10..0..10)
- **Outcomes**: positive (0..10) and negative (0..-10) consequences
- **Probabilities**: positive and negative consequences (0..100%)

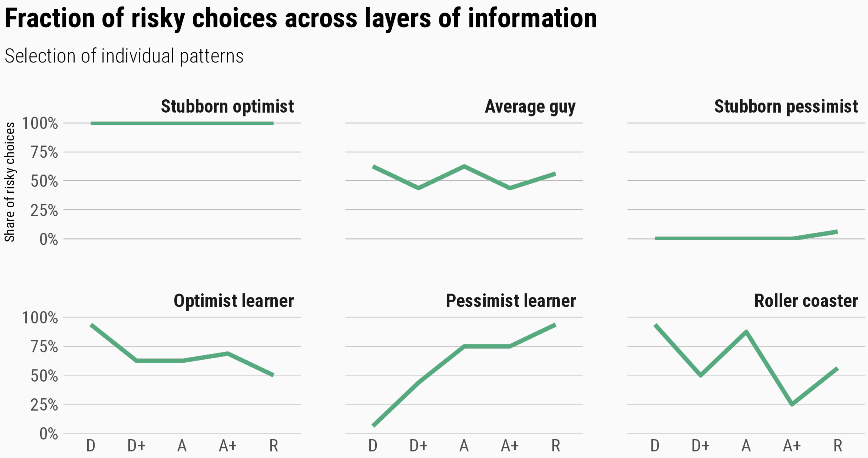
Revealing information increases risk taking



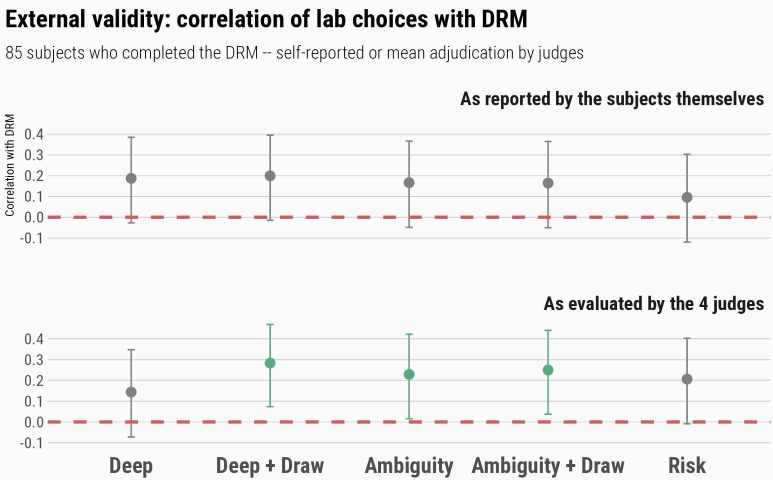
This hides significant heterogeneity across subjects



This hides significant heterogeneity across subjects



Correlations task $\Longleftrightarrow$ external measure of risk attitudes



Risk perception

Do subjects find our tasks risky? Don't ask, don't tell

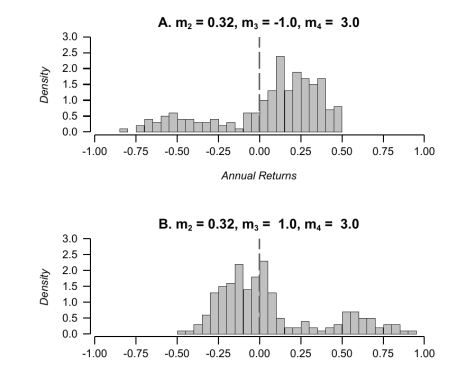
We **don't know** because we just plainly assume they do!

- 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

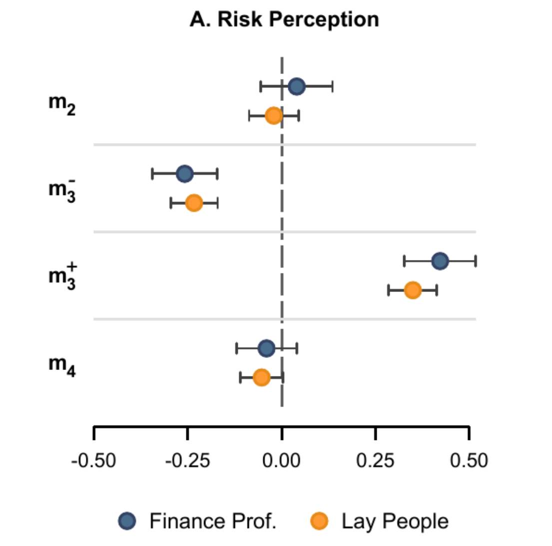
Experimenting on risk perception

Holzmeister et al (Man Sci 2021)

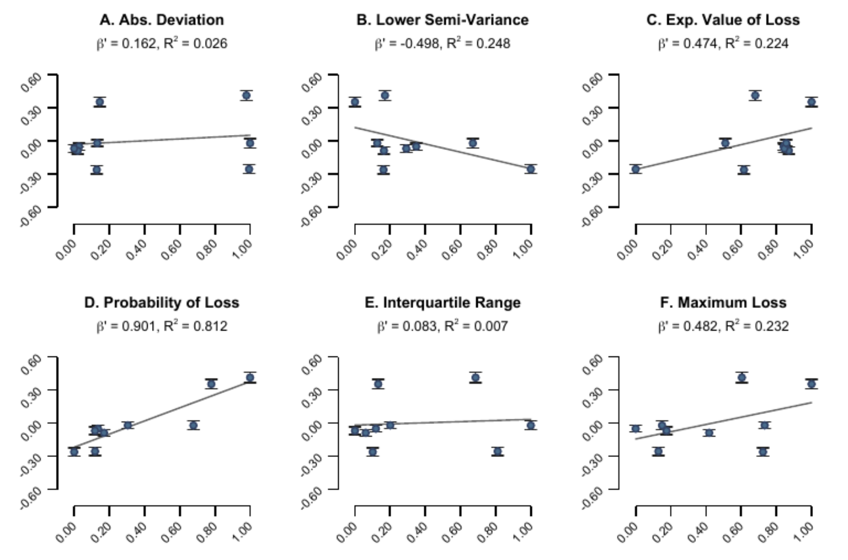
- Rate descriptions of asset returns
- i.e., perceived risk
- ∼7000 subjects
- including ∼2500 **traders**



Risk perception not driven by variance (but skewness)



Best-fitting definition of risk: probability of a loss



Competing theories: noisy coding & cognitive approach

What if risk aversion is not?

We have so far **assumed EUT**. But it's no more the only game in town

- **noisy coding**: risk aversion $\sim$ risk neutraility $+$ the **way we see the world**
- **Logarihtmic** number perception $\sim$ risk aversion (Khaw et al. 2021)
- Loss aversion and probability weighting $\sim$ cognitive **artifacts** (Vieider 2024)
- Choice under risk $\sim$ choice under complexity and **confusion** (Oprea 2024)

Summing up. . .

Summing up

The quest for a good risk measure

- there is no *one good way* of eliciting risk
- the field doe snot produce reliable estimates
- in practical applications, lots of trade-offs and less-worse dilemmas

Summing up

The quest for a good risk measure

- there is no *one good way* of eliciting risk
- the field doe snot produce reliable estimates
- in practical applications, lots of trade-offs and less-worse dilemmas

But!

- Research is ongoing!
- If you just need a control – probably just *ask*
- If you need a parameter: use a low-bias task (as HL, or BRET)
- If you need external validity: beware of risk perception issues!
- ...and maybe Risk Aversion is simply *not*!