Eliciting Risk Attitudes Experimentally

AN ASSESSMENT OF DIFFERENT TASKS

Montpellier, LAMETA, December 3rd, 2015

Paolo Crosetto and Antonio Filippin

Usual boring plan of the talk

1. **Risk elicitation tasks**: introduce and discuss different methods
2. **Within-subject consistency**: wildly unstable
3. **Why?** we'll look at the shape and mechanics of the tasks
4. **Digging deeper**: in-depth task comparison. Noise, gender, safe options...

Experiments: underlying hypotheses

- risk preferences are a **stable** feature of preferences
- i.e., they are not context-dependent and do not vary.
- they can be represented by different concept in different utility functions
  - in EUT: curvature of the utility function
  - in PT: one of three parameters determining utility function
  - ...
- in general, though, they are primitives of models in most of economics...
- ...hence the desire to capture them.

Experiments: different approaches

Risk elicitation tasks: two approaches

1. When aiming at testing competing theories, estimate functional forms:
   - give the subjects many (lottery) choices;
   - use the data to estimate all parameters of some utility function;
   - e.g.: Hey and Orme (1994): testing predictive power of theories.
   - Time consuming, many choices, but very precise.
   - Allows for structural modeling (...)
2. When running other experiments/studies in which uncertainty plays a role:
   - use a custom, unrelated control task;
   - use data as a control.
   - Fast, easier to understand, (possibly) less precise.
   - Possibly the researcher cares about a ranking rather than an accurate point prediction

In this talk we focus on simple, tractable control tasks

How do we evaluate a task?

We compare theoretically and experimentally four risk elicitation tasks

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.

Within-subject consistency?

Problem: different task repeated yield wildly different estimates

- Lots of evidence of inter-subject (between) but also intra-subject (within) inconsistencies
- Deck et al. JEBO 2013 (within: Balloon, Deal-No-Deal, HL, Binswanger)
- Dave et al. JRU 2010 (HL and EG)
- Isaac and James JRU 2000 /BDM and auctions)
- ... (plug in any other comparison of risk tasks)

Within-subject consistency?

Problem: different task repeated yield wildly different estimates

- Lots of evidence of inter-subject (between) but also intra-subject (within) inconsistencies
- Deck et al. JEBO 2013 (within: Balloon, Deal-No-Deal, HL, Binswanger)
- Dave et al. JRU 2010 (HL and EG)
- Isaac and James JRU 2000 /BDM and auctions)
- ... (plug in any other comparison of risk tasks)

Not just scaled up or down: different ranking, risk lovers become risk averse...

Isaac and James - JRU 2000

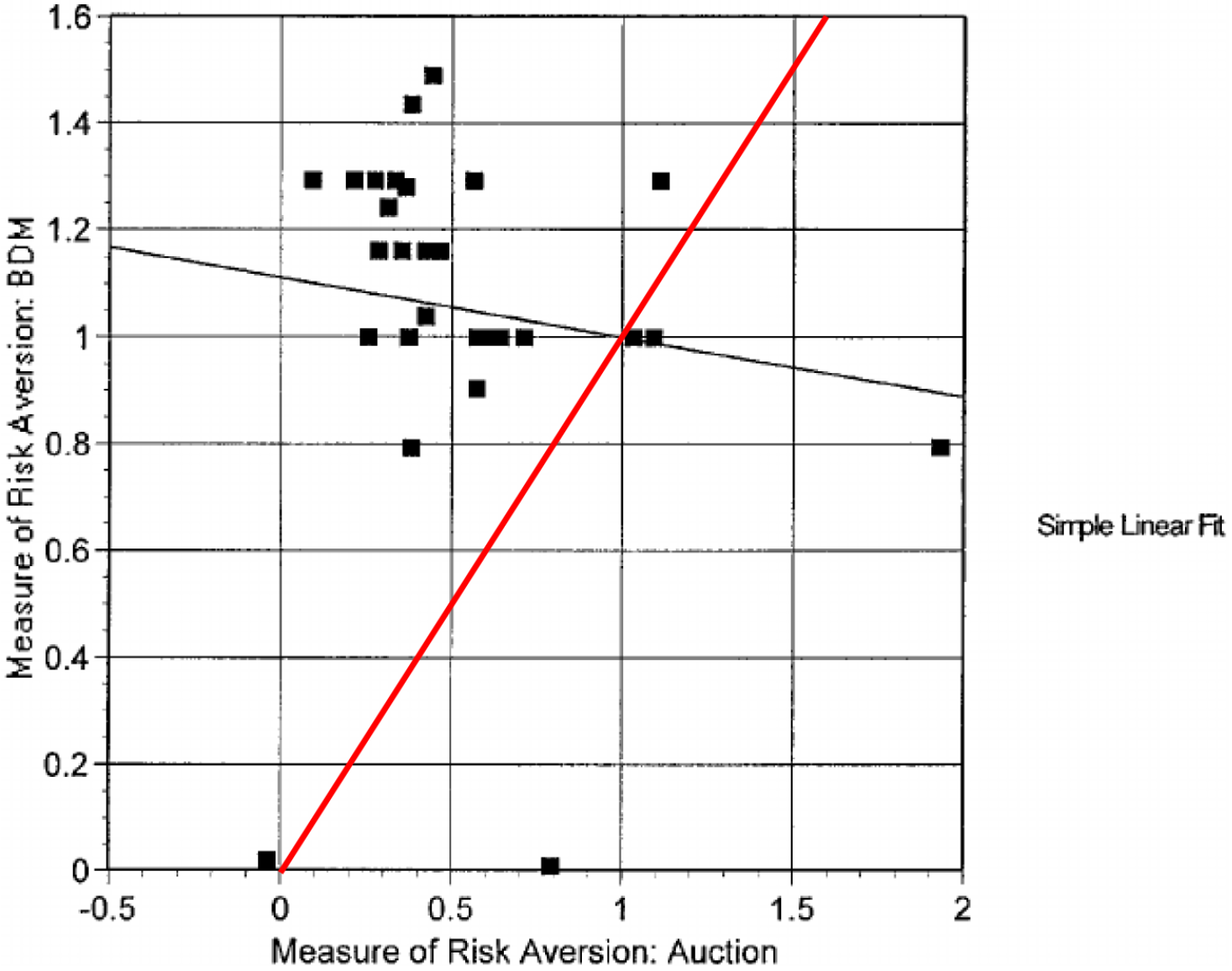
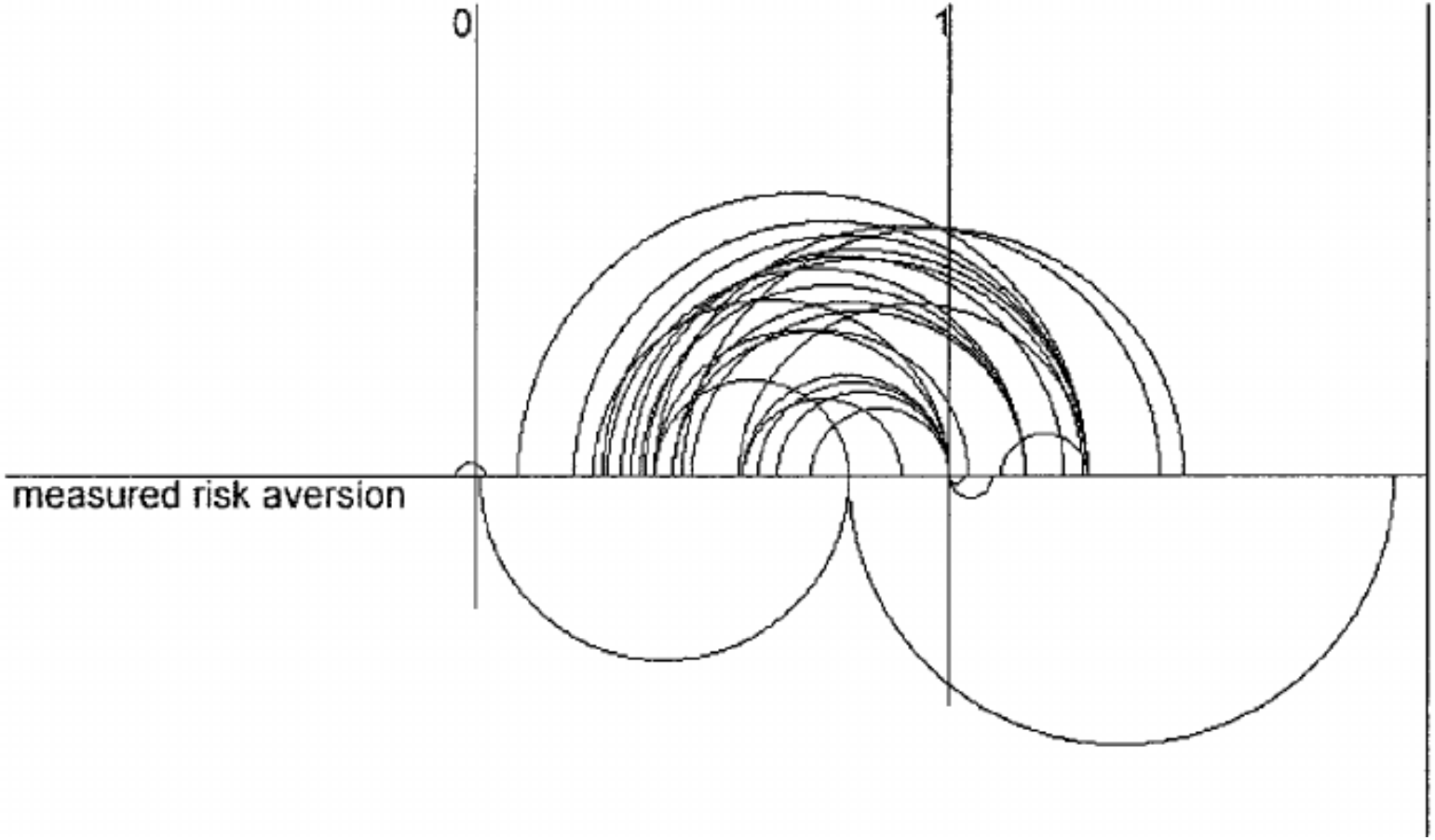
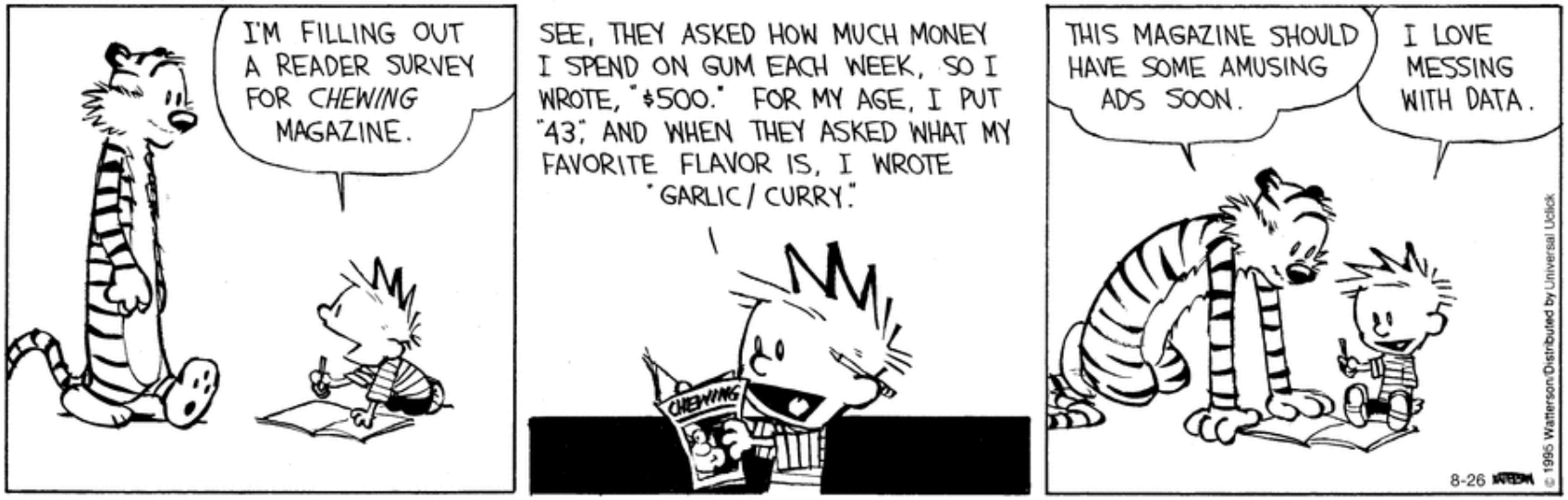


Figure 1

Isaac and James - JRU 2000



Why?


I'M FILLING OUT A READER SURVEY FOR CHEWING MAGAZINE.
SEE, THEY ASKED HOW MUCH MONEY I SPEND ON GUM EACH WEEK, SO I WROTE, "$500." FOR MY AGE, I PUT "43," AND WHEN THEY ASKED WHAT MY FAVORITE FLAVOR IS, I WROTE "GARLIC / CURRY."
CHEWING
THIS MAGAZINE SHOULD HAVE SOME AMUSING ADS SOON.
I LOVE MESSING WITH DATA.
8-26
© 1995 Watterson/Distributed by Universal Uclick

Why?

- Inherent in the within structure: hedging...
- Characteristics of the task
- Wild instability of preferences
- Does it exist such thing as a risk aversion personal trait?

Experimental replication

Between-subjects comparison of tasks

We ran a **between-subjects** comparison within a homogeneous subject pool (Jena)

We included 4 tasks:

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

Risk Elicitation Tasks, I: Holt and Laury

Ten binary lottery choices – risk attitude as switching point

1/10 prob. of 4.0 Euro	9/10 prob. of 3.2 Euro	A	○	○	B	1/10 prob. of 7.7 Euro	9/10 prob. of 0.2 Euro
2/10 prob. of 4.0 Euro	8/10 prob. of 3.2 Euro	A	○	○	B	2/10 prob. of 7.7 Euro	8/10 prob. of 0.2 Euro
3/10 prob. of 4.0 Euro	7/10 prob. of 3.2 Euro	A	○	○	B	3/10 prob. of 7.7 Euro	7/10 prob. of 0.2 Euro
4/10 prob. of 4.0 Euro	6/10 prob. of 3.2 Euro	A	○	○	B	4/10 prob. of 7.7 Euro	6/10 prob. of 0.2 Euro
5/10 prob. of 4.0 Euro	5/10 prob. of 3.2 Euro	A	○	○	B	5/10 prob. of 7.7 Euro	5/10 prob. of 0.2 Euro
6/10 prob. of 4.0 Euro	4/10 prob. of 3.2 Euro	A	○	○	B	6/10 prob. of 7.7 Euro	4/10 prob. of 0.2 Euro
7/10 prob. of 4.0 Euro	3/10 prob. of 3.2 Euro	A	○	○	B	7/10 prob. of 7.7 Euro	3/10 prob. of 0.2 Euro
8/10 prob. of 4.0 Euro	2/10 prob. of 3.2 Euro	A	○	○	B	8/10 prob. of 7.7 Euro	2/10 prob. of 0.2 Euro
9/10 prob. of 4.0 Euro	1/10 prob. of 3.2 Euro	A	○	○	B	9/10 prob. of 7.7 Euro	1/10 prob. of 0.2 Euro
10/10 prob. of 4.0 Euro	0/10 prob. of 3.2 Euro	A	○	○	B	10/10 prob. of 7.7 Euro	0/10 prob. of 0.2 Euro

Risk Elicitation Tasks, I: Holt and Laury

Ten binary lottery choices – risk attitude as switching point

1/10 prob. of 4.0 Euro	9/10 prob. of 3.2 Euro	A	○	○	B	1/10 prob. of 7.7 Euro	9/10 prob. of 0.2 Euro
2/10 prob. of 4.0 Euro	8/10 prob. of 3.2 Euro	A	○	○	B	2/10 prob. of 7.7 Euro	8/10 prob. of 0.2 Euro
3/10 prob. of 4.0 Euro	7/10 prob. of 3.2 Euro	A	○	○	B	3/10 prob. of 7.7 Euro	7/10 prob. of 0.2 Euro
4/10 prob. of 4.0 Euro	6/10 prob. of 3.2 Euro	A	○	○	B	4/10 prob. of 7.7 Euro	6/10 prob. of 0.2 Euro
5/10 prob. of 4.0 Euro	5/10 prob. of 3.2 Euro	A	○	○	B	5/10 prob. of 7.7 Euro	5/10 prob. of 0.2 Euro
6/10 prob. of 4.0 Euro	4/10 prob. of 3.2 Euro	A	○	○	B	6/10 prob. of 7.7 Euro	4/10 prob. of 0.2 Euro
7/10 prob. of 4.0 Euro	3/10 prob. of 3.2 Euro	A	○	○	B	7/10 prob. of 7.7 Euro	3/10 prob. of 0.2 Euro
8/10 prob. of 4.0 Euro	2/10 prob. of 3.2 Euro	A	○	○	B	8/10 prob. of 7.7 Euro	2/10 prob. of 0.2 Euro
9/10 prob. of 4.0 Euro	1/10 prob. of 3.2 Euro	A	○	○	B	9/10 prob. of 7.7 Euro	1/10 prob. of 0.2 Euro
10/10 prob. of 4.0 Euro	0/10 prob. of 3.2 Euro	A	○	○	B	10/10 prob. of 7.7 Euro	0/10 prob. of 0.2 Euro

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

Risk Elicitation Tasks, II: Eckel and Grossman

A single choice among five (six) 50-50 lotteries – chosen lottery is played.

Lottery 1	50%	A	4 Euro
	50%	B	4 Euro
Lottery 2	50%	A	6 Euro
	50%	B	3 Euro
Lottery 3	50%	A	8 Euro
	50%	B	2 Euro
Lottery 4	50%	A	10 Euro
	50%	B	1 Euro
Lottery 5	50%	A	12 Euro
	50%	B	0 Euro

Risk Elicitation Tasks, II: Eckel and Grossman

A single choice among five (six) 50-50 lotteries – chosen lottery is played.

Lottery 1	50%	A	4 Euro
	50%	B	4 Euro
Lottery 2	50%	A	6 Euro
	50%	B	3 Euro
Lottery 3	50%	A	8 Euro
	50%	B	2 Euro
Lottery 4	50%	A	10 Euro
	50%	B	1 Euro
Lottery 5	50%	A	12 Euro
	50%	B	0 Euro

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

Risk Elicitation Tasks,III: 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 Elicitation Tasks,III: 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 $E(\text{risky}) = 1.25 > 1$.

Presenting the BRET

BRET: rules

We developed the 'Bomb' Risk Elicitation Task (BRET)

BRET: rules

We developed the 'Bomb' Risk Elicitation Task (BRET)

- Subjects are shown a field with 100 boxes.
- Are told that under one of the boxes lies a *time bomb*
- Their task is to collect boxes.
- When they hit the *Start* button, the computer starts collecting...
- ...one box per second, automatically, in numerical order.
- The subjects must only *stop* the collection process.

- Once the task is over, the position of the bomb is determined (hence the *time bomb*).
- If bomb collected $\rightarrow$ earnings equal zero.
- If bomb not collected $\rightarrow$ earnings equal to number of boxes collected.

BRET: interface, I

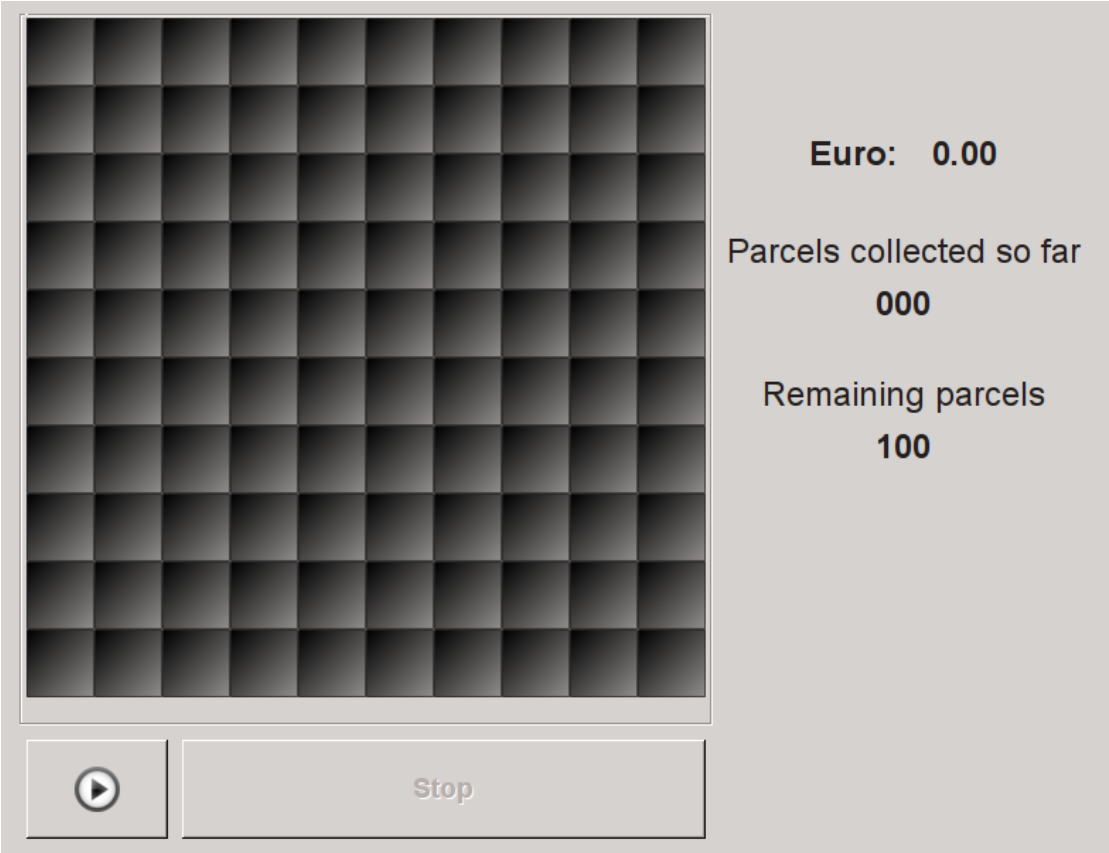


Figure : The BRET interface at the start of the experiment

BRET: interface, I

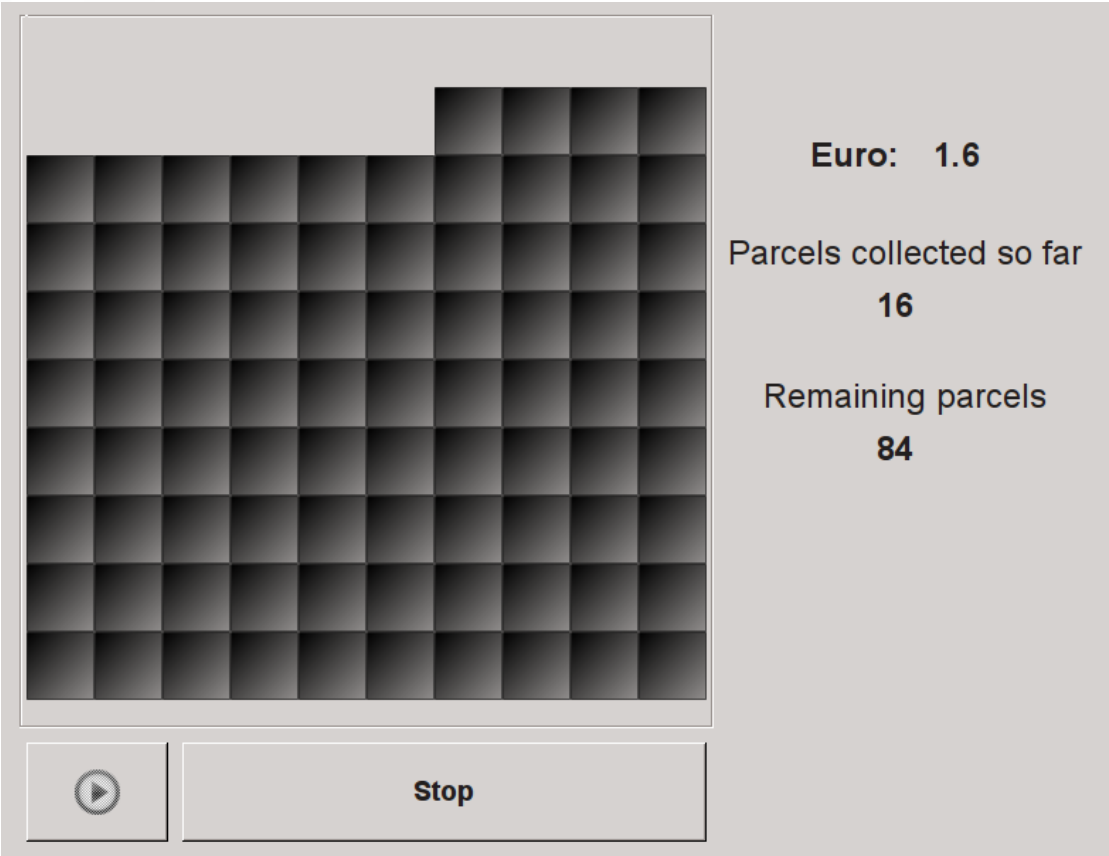


Figure : The BRET interface after 16 seconds

BRET: under the hood

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

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

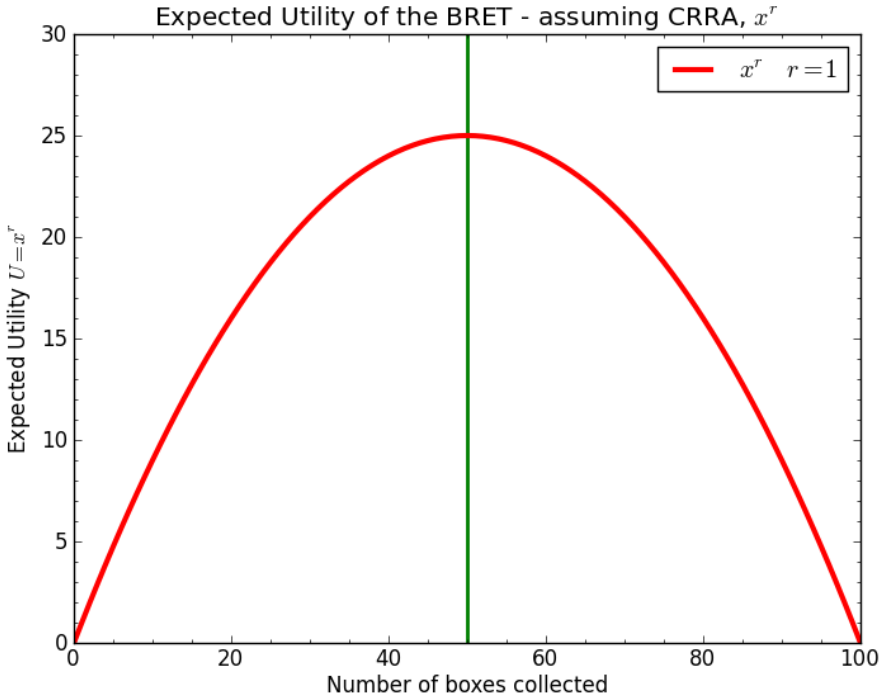
BRET: under the hood

- 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 $k = 20$, $L = \{20\% : 0\,;\, 80\% : 20\}$

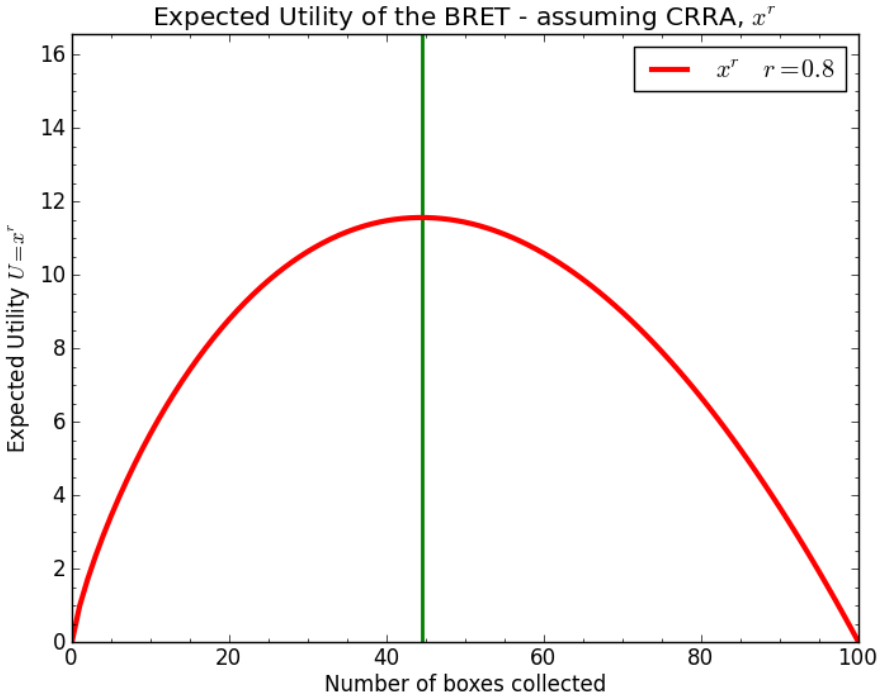
BRET: solution for the expected value maximizer



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

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

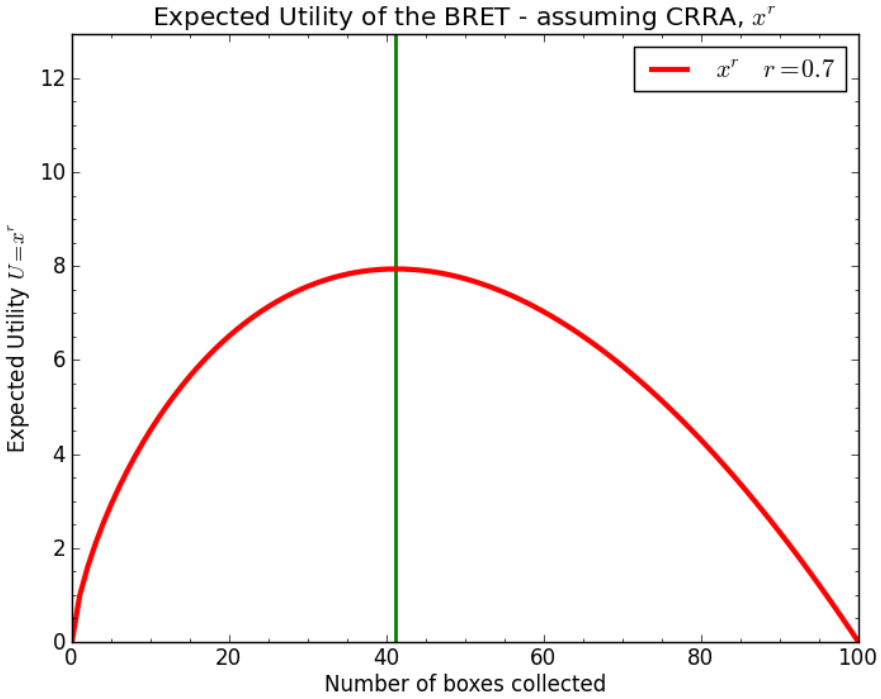
BRET: Risk averse subject



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

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

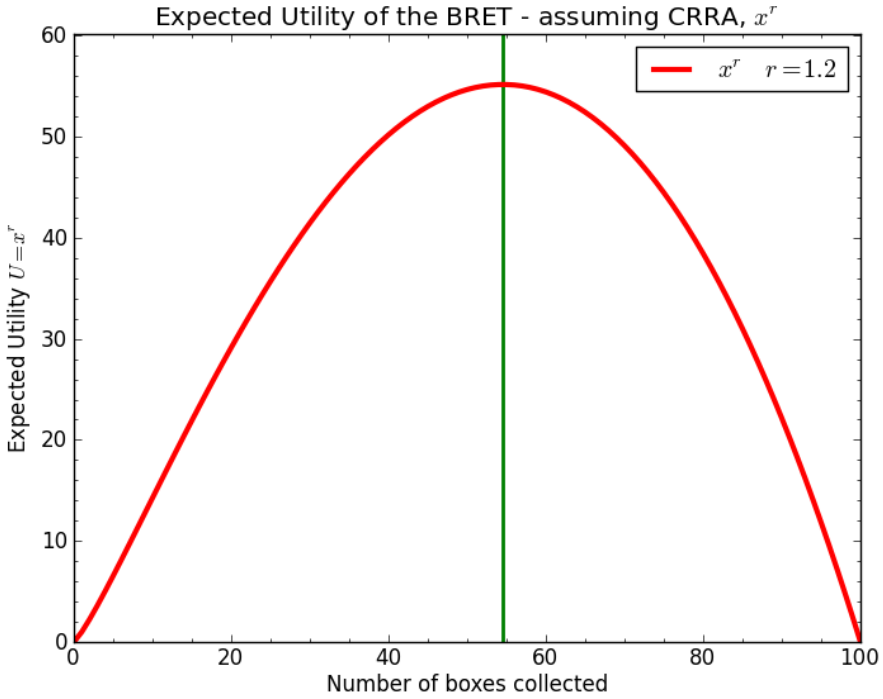
BRET: Risk averse subject



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

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

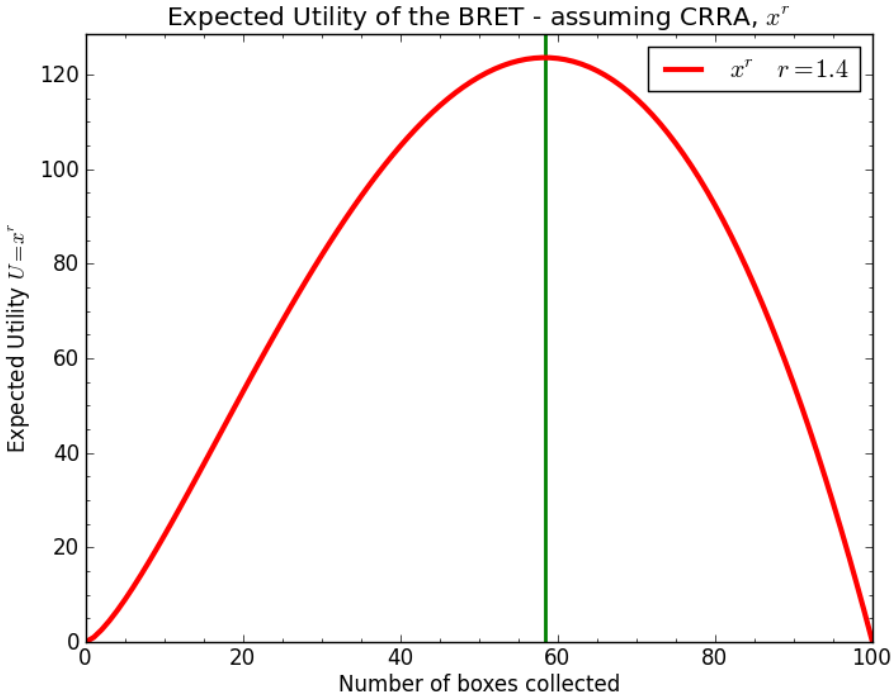
BRET: Risk lover subject



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:

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

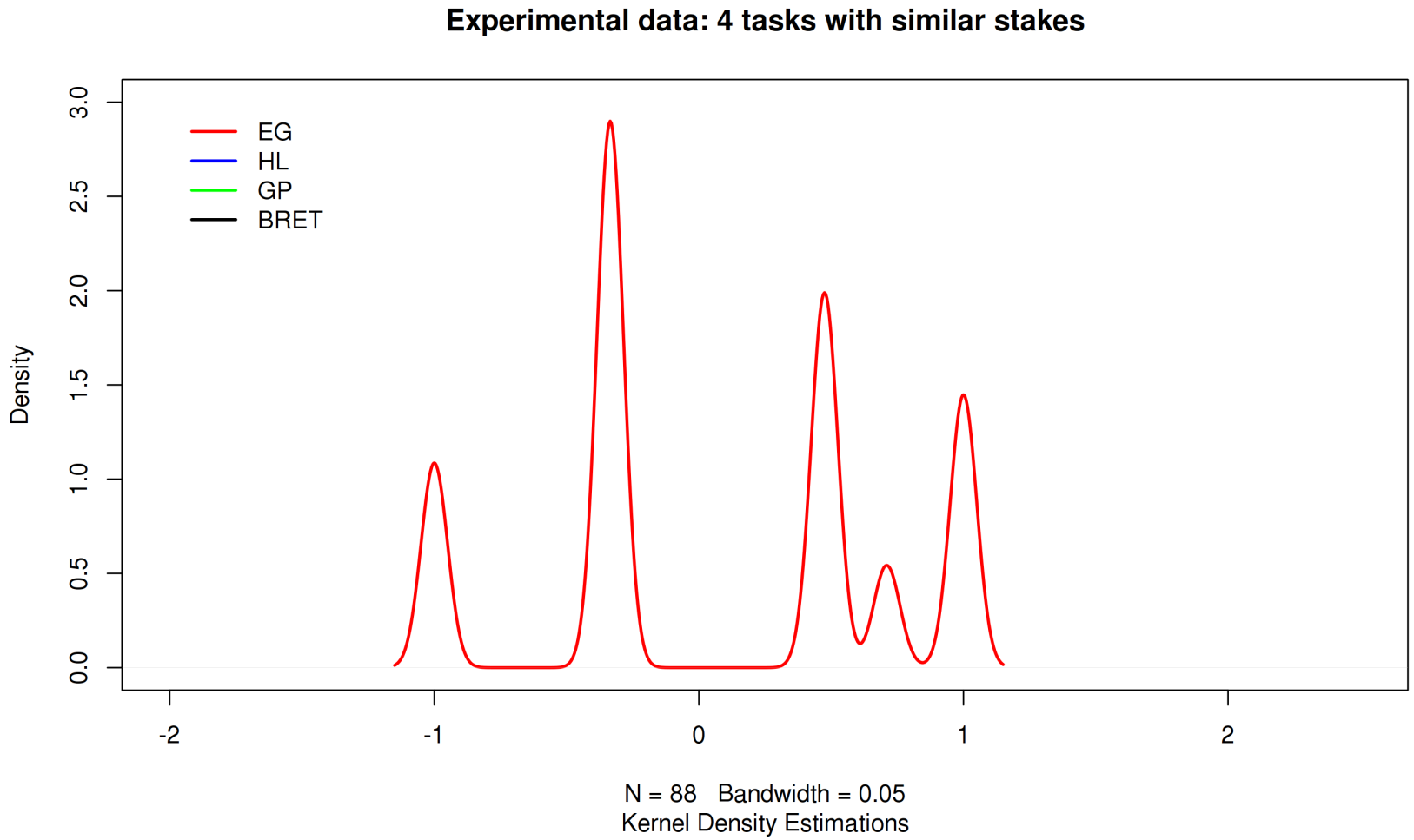
Experimental design

We ran between-subject comparisons within a homogeneous subject pool.

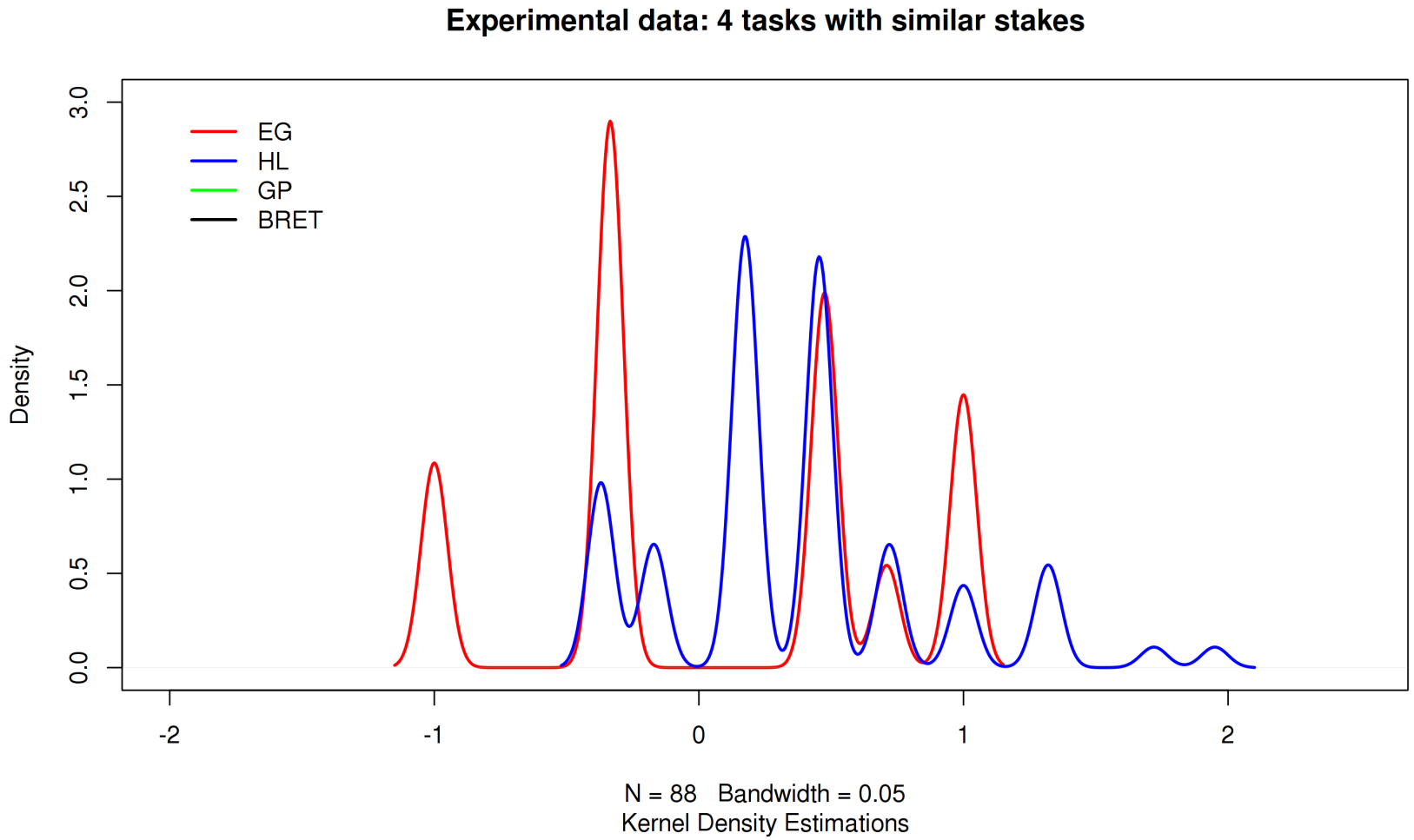
- Max Planck Institute for Economics Lab, Jena, March-May 2012.
- A total of 350 subjects (151 M, 199 F).
- Pure between subjects design, one shot.
- Identical procedures, 'live' resolution of uncertainty at the session level.
- In each treatment we also elicited SOEP self-reported risk-attitude question...
- ...and the DOSPERT risk questionnaire.

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

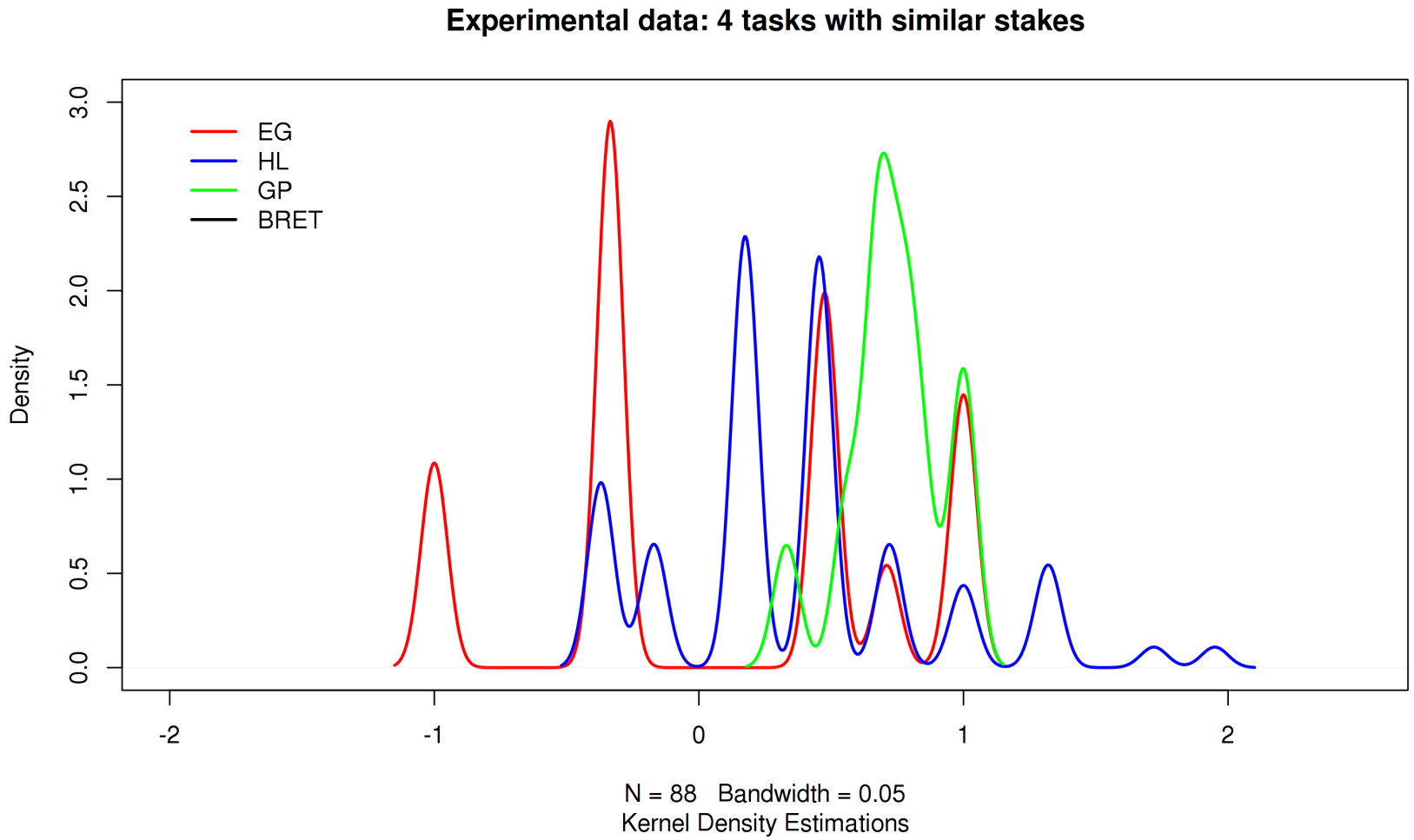
Distribution of r



Distribution of r

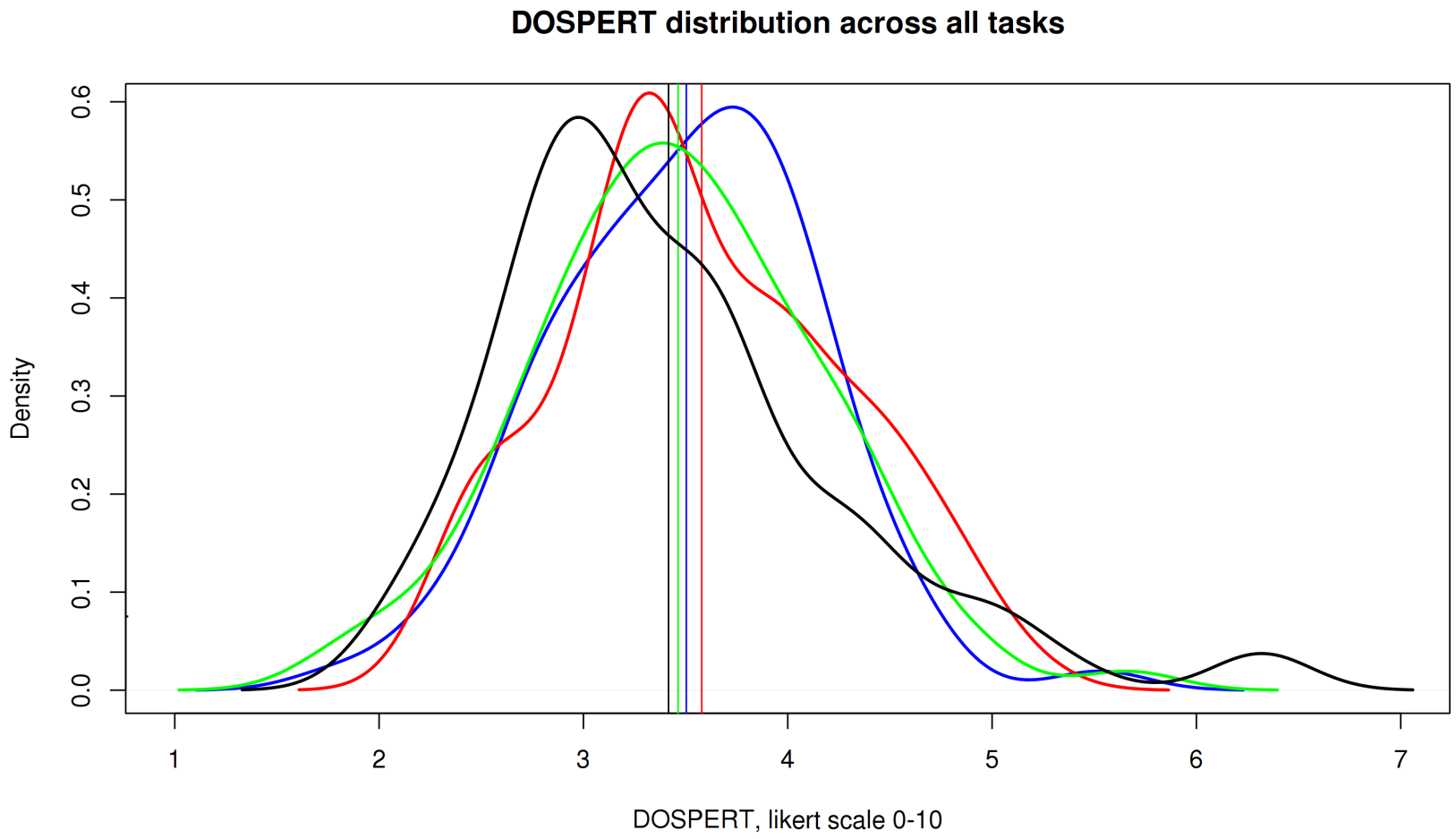


Distribution of r



Distribution of r

Just because we are between-subjects? DOSPERT



Characteristics of the tasks

Theoretical comparison

	Precision (categories of r)	Parsimony (no. of choices)	Completeness (r range)	Ambiguity	Reduction	Intuitiveness
HL	8	10	yes	no	suffers	maths
EG	5	1	no	no	no	maths
CG	100+	1	no	no	no	intuitive
BRET	100	1	yes	no	no	visual

Precision: HL and EG coarser.

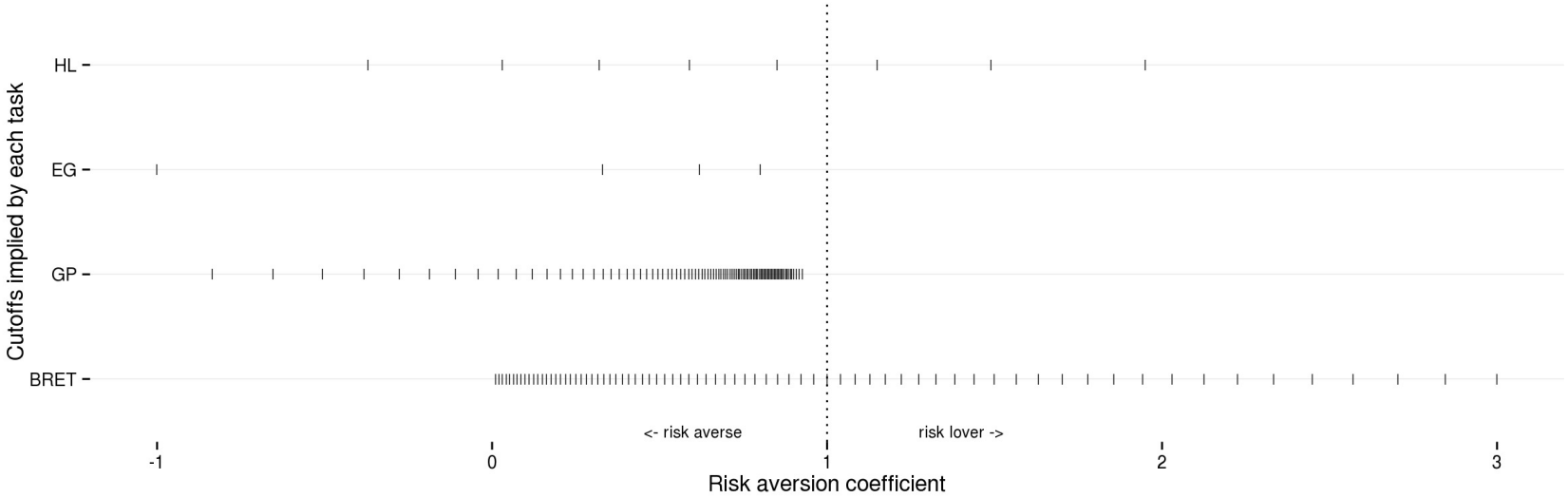
Parsimony: HL has ten choices, all the others 1.

Complete: EG and CGP *cannot* tell apart risk lovers from risk neutrals.

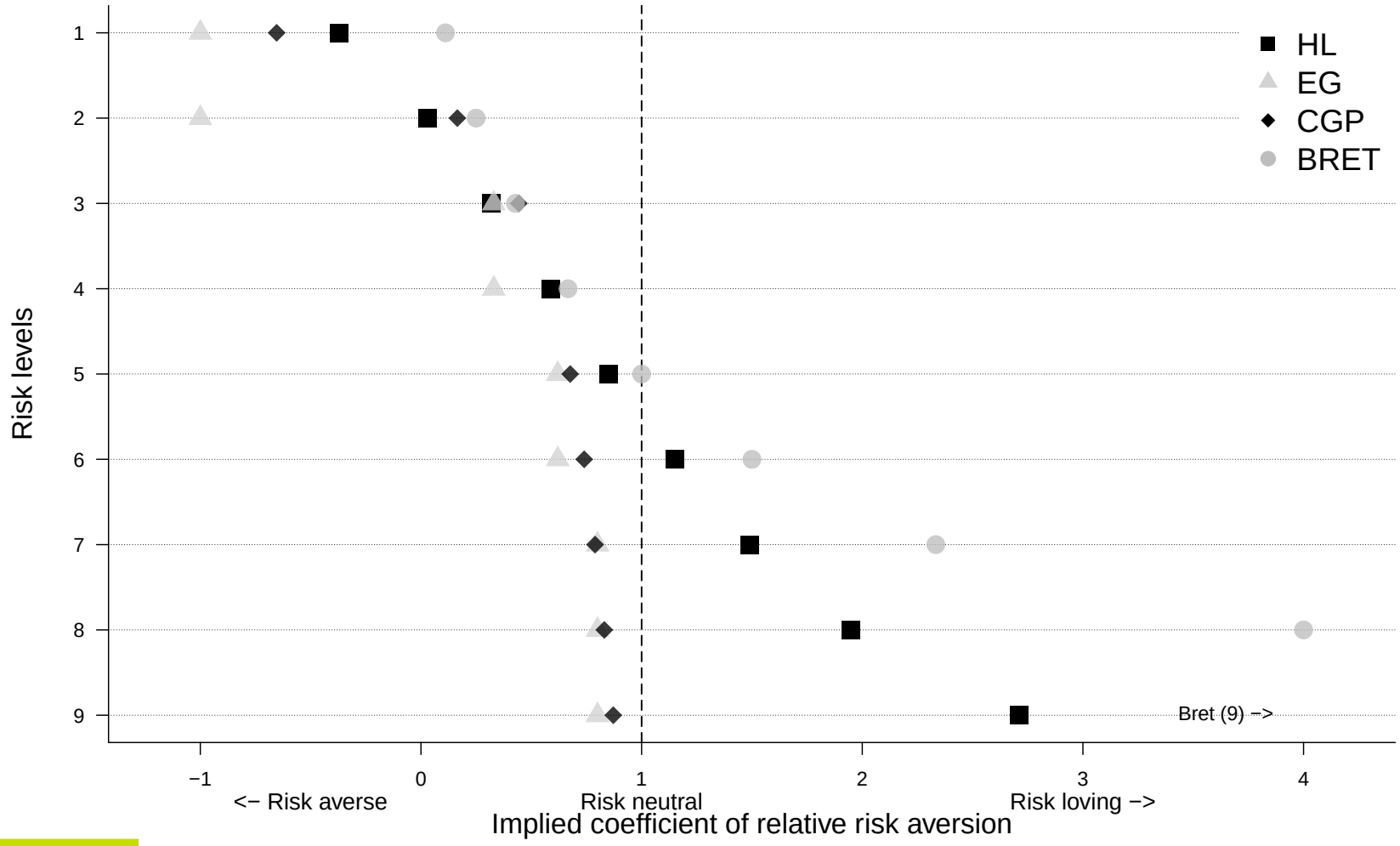
Sound: HL ($>$ 1 choice) not robust to violations of RA.

Intuitive: EG and HL use abstract lotteries (math); others intuitive or visual. Empirical question.

Mapping choices to r: cutoffs



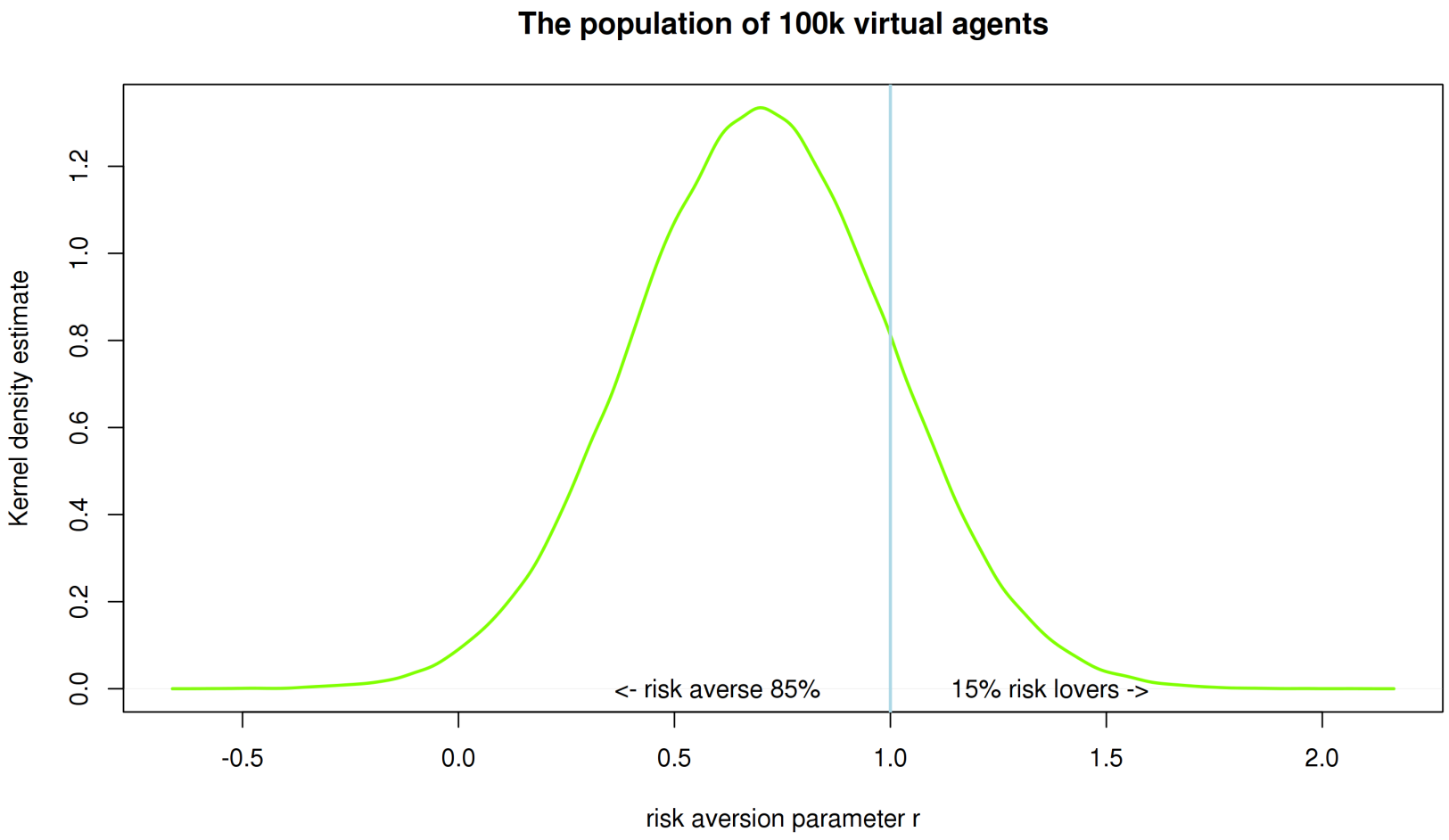
Mapping choices to r: risk levels



Test-drive in a safe environment: 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:

Simulations, II

Three types of simulations:

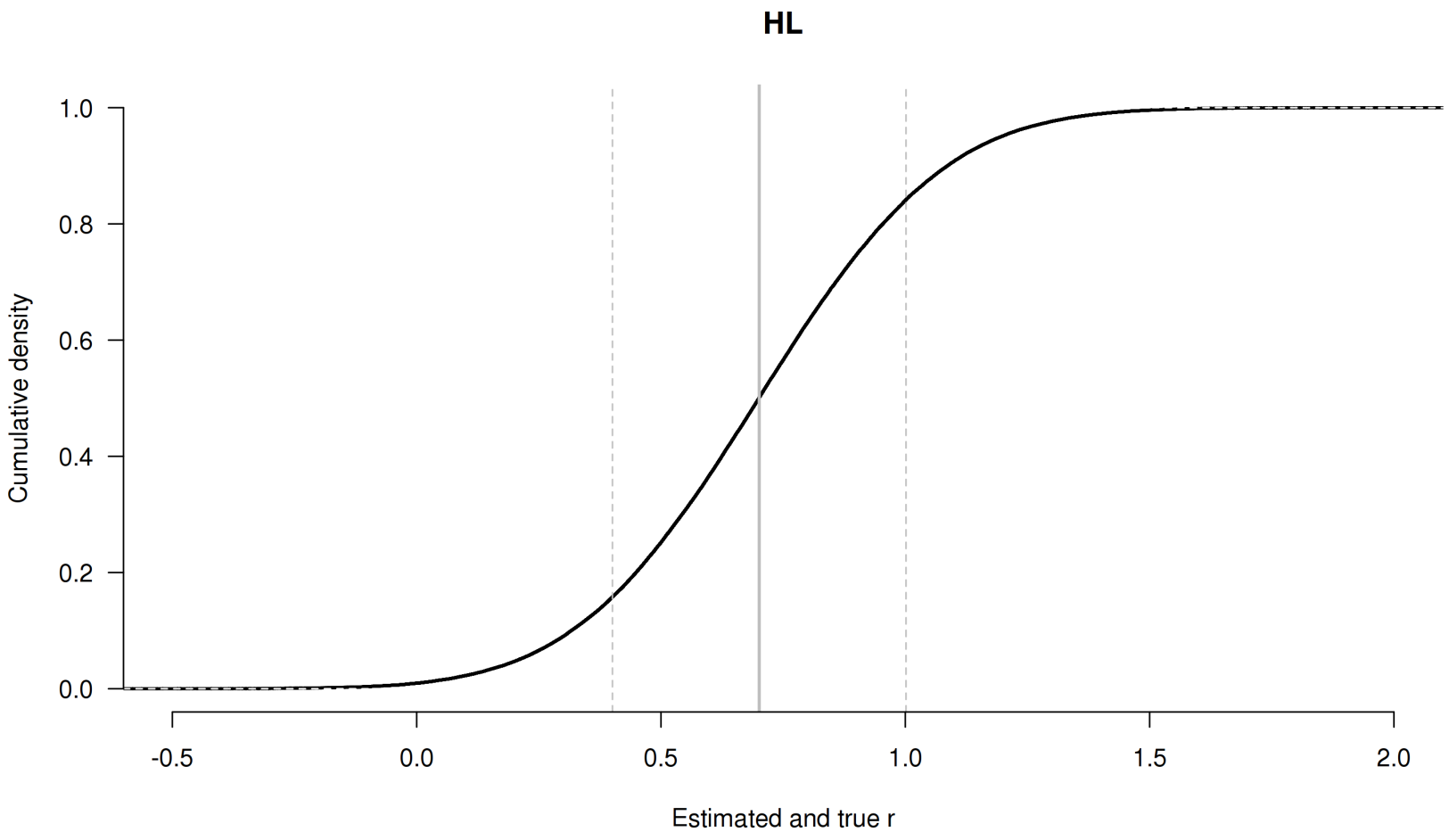
1. Deterministic: virtual subjects play according to their **true r**
2. Stochastic:
   - for each agent, $r_n = r + \varepsilon, \varepsilon \sim N(0, \mu)$
   - that is, the agent deviates from her true preferences with a white noise
   - $\mu = 0.3$ or 0.6
3. Behaviorally random:
   - a 10% share 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
4. NEW! Fechner error
   - (computation this very morning, take them with a grain of salt)
   - People make a $\sim N(0, 0.1)$ error when comparing lotteries
   - assumptions on lotteries comparisons needed (!)

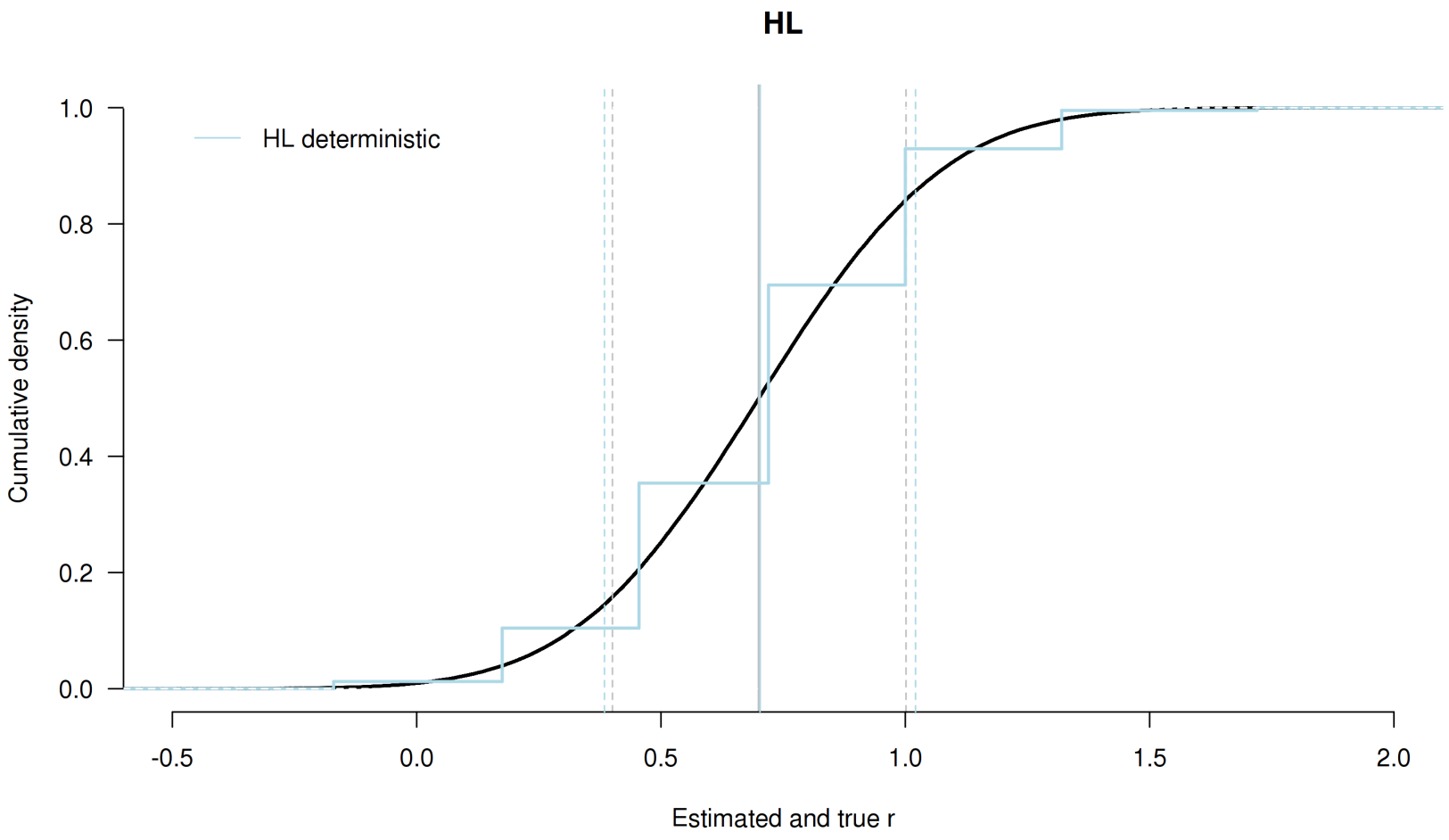
Simulations, II

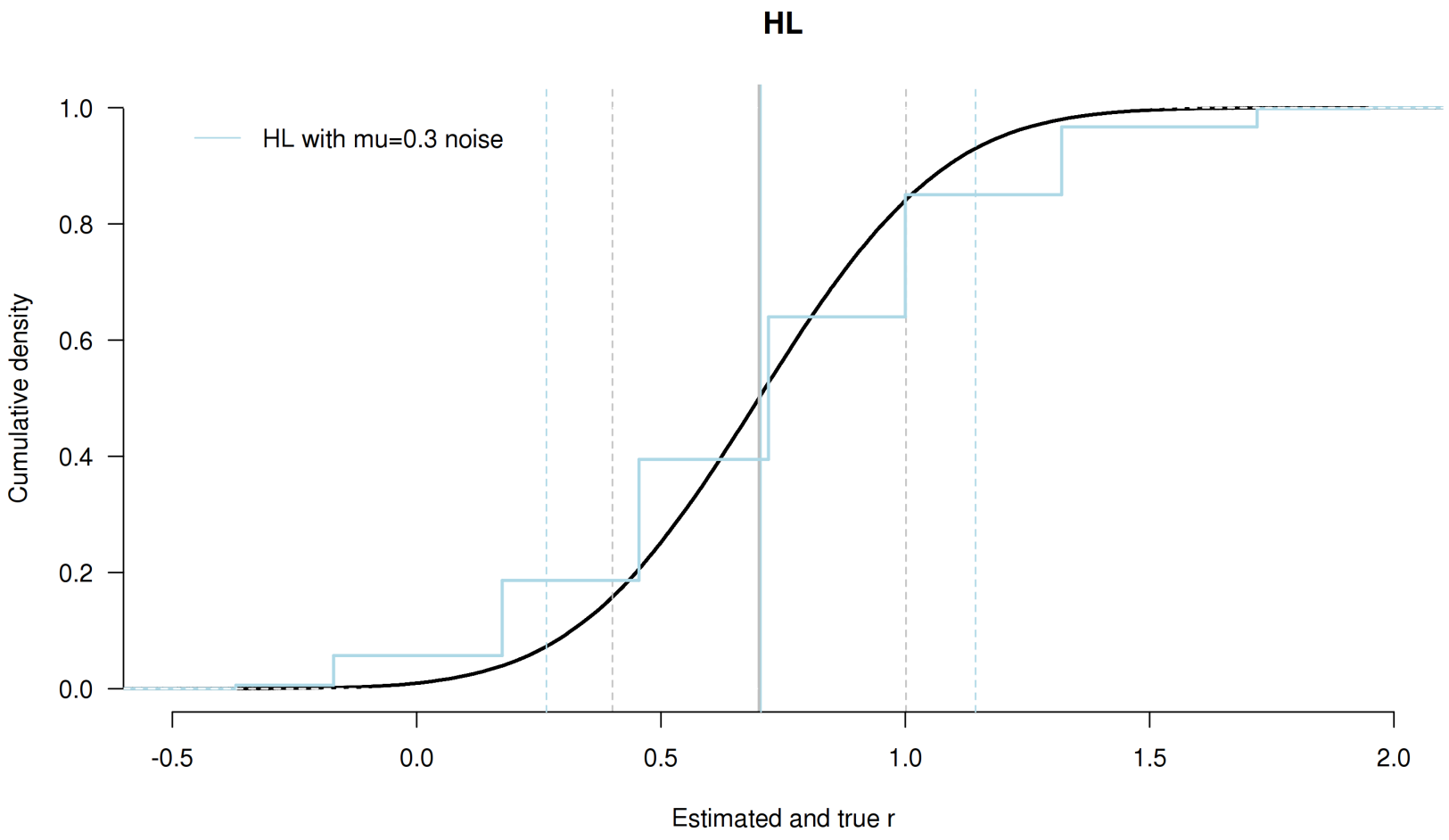
Three types of simulations:

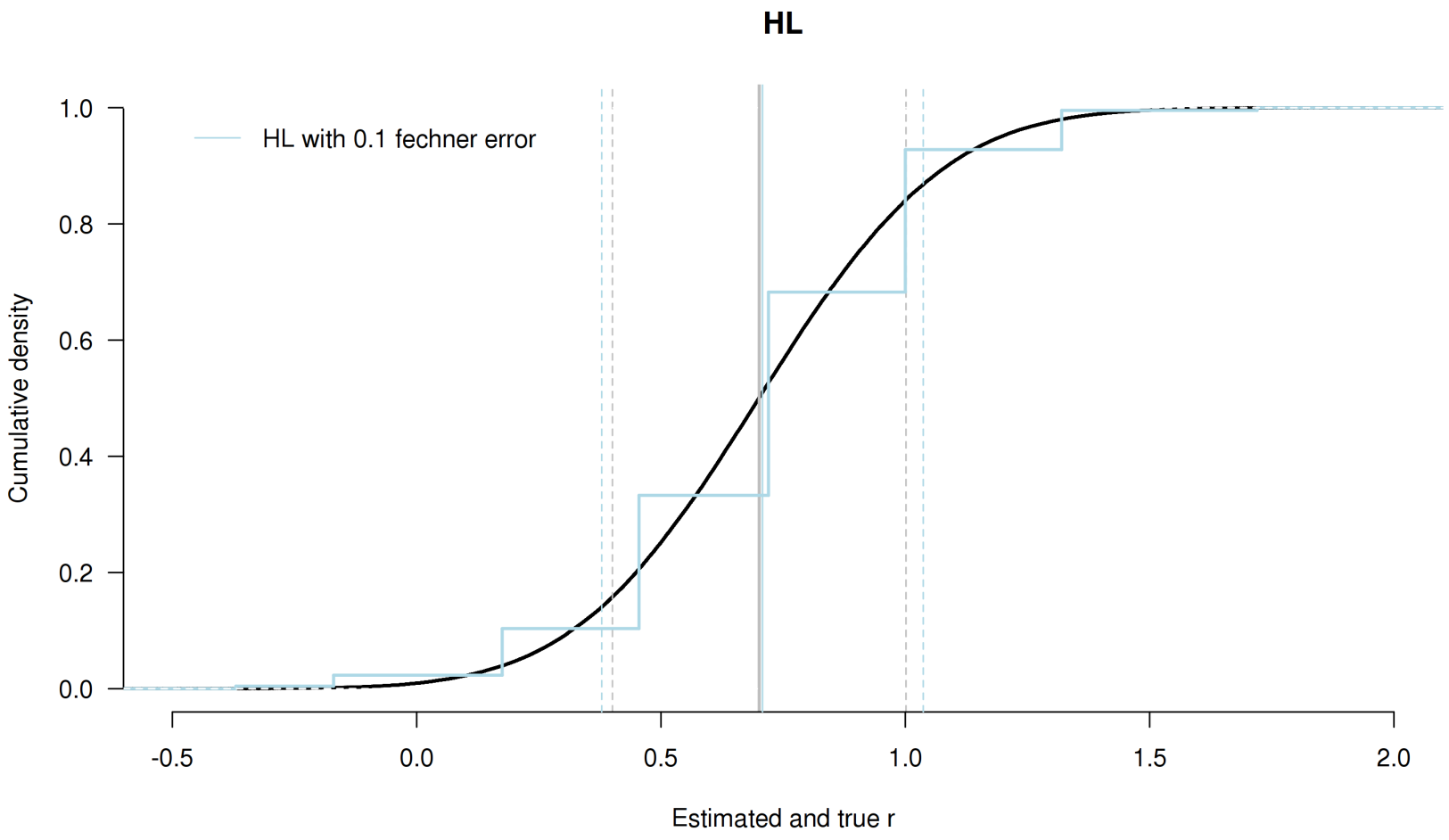
1. Deterministic: virtual subjects play according to their **true r**
2. Stochastic:
   - for each agent, $r_n = r + \varepsilon, \varepsilon \sim N(0, \mu)$
   - that is, the agent deviates from her true preferences with a white noise
   - $\mu = 0.3$ or 0.6
3. Behaviorally random:
   - a 10% share 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
4. NEW! Fechner error
   - (computation this very morning, take them with a grain of salt)
   - People make a $\sim N(0, 0.1)$ error when comparing lotteries
   - assumptions on lotteries comparisons needed (!)

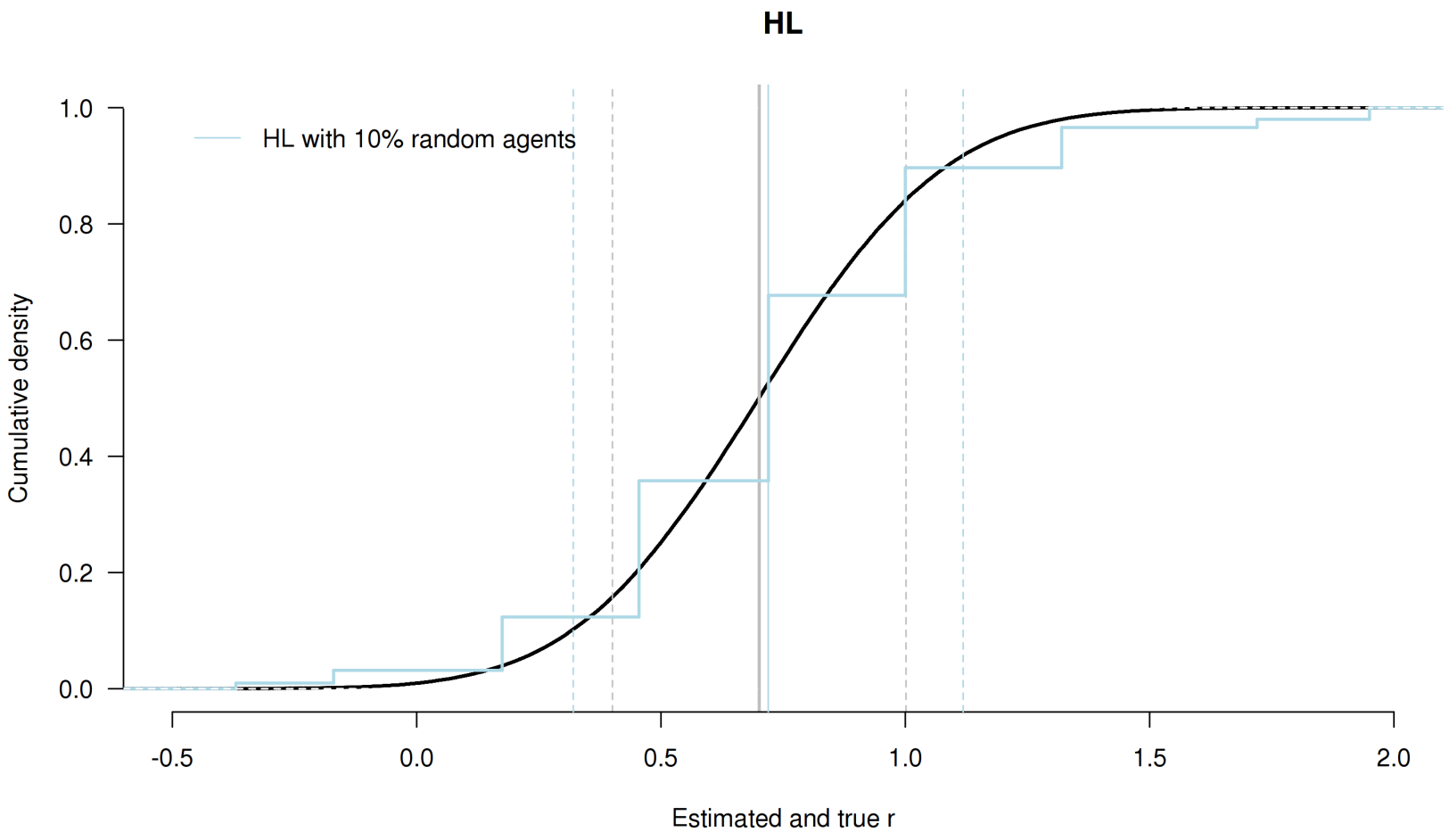
Task by task...

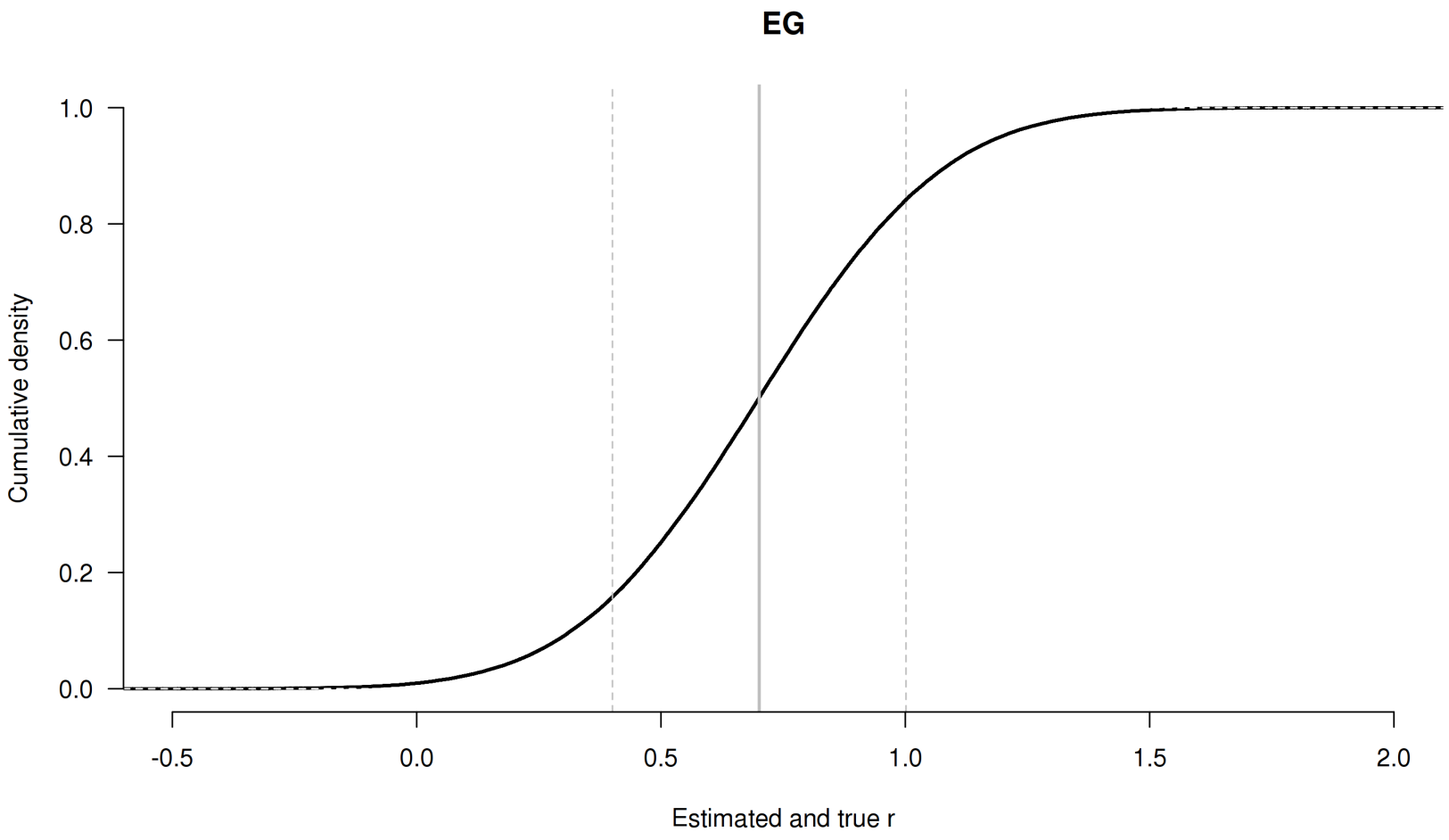

HL
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

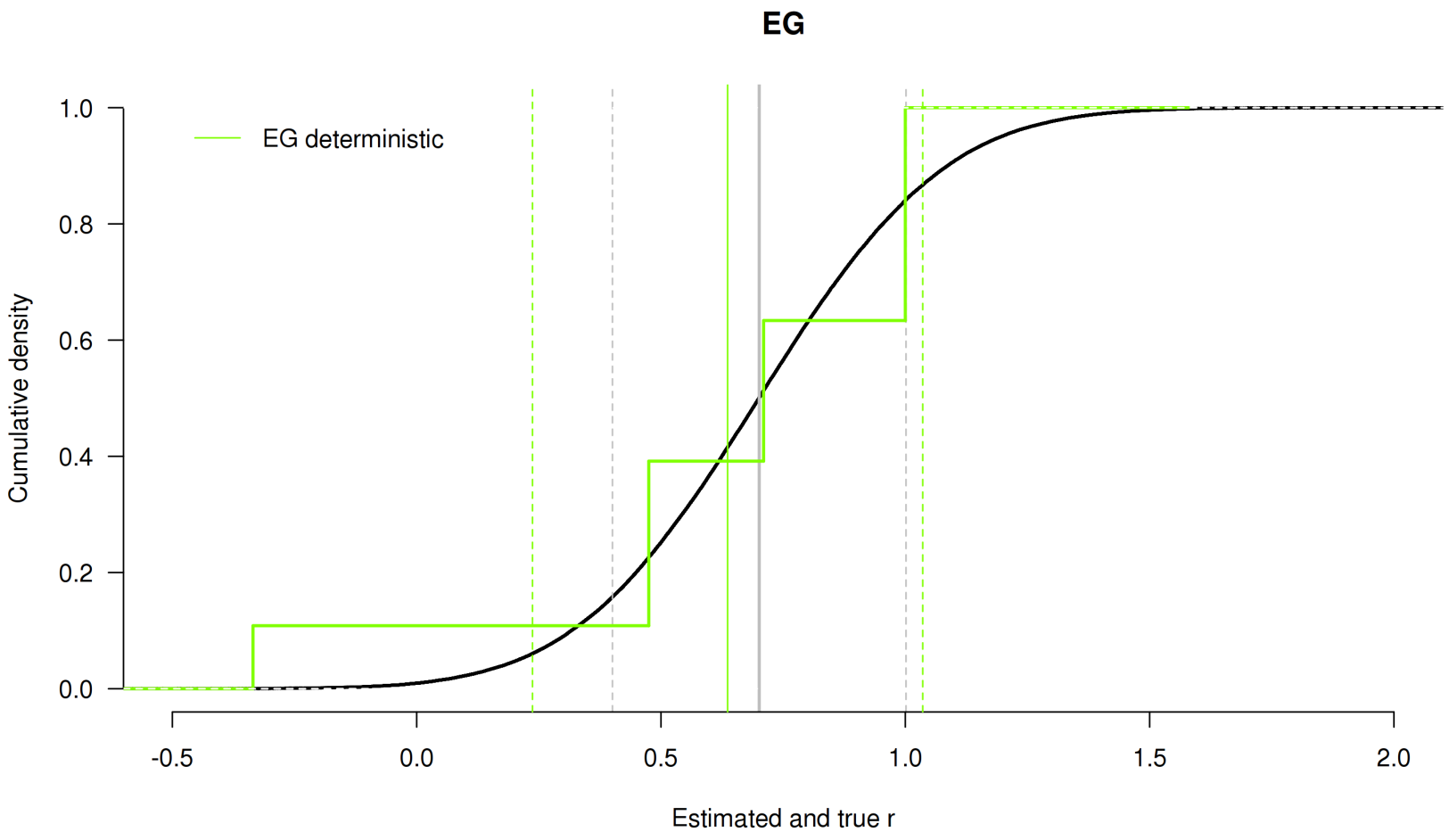

HL
HL deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

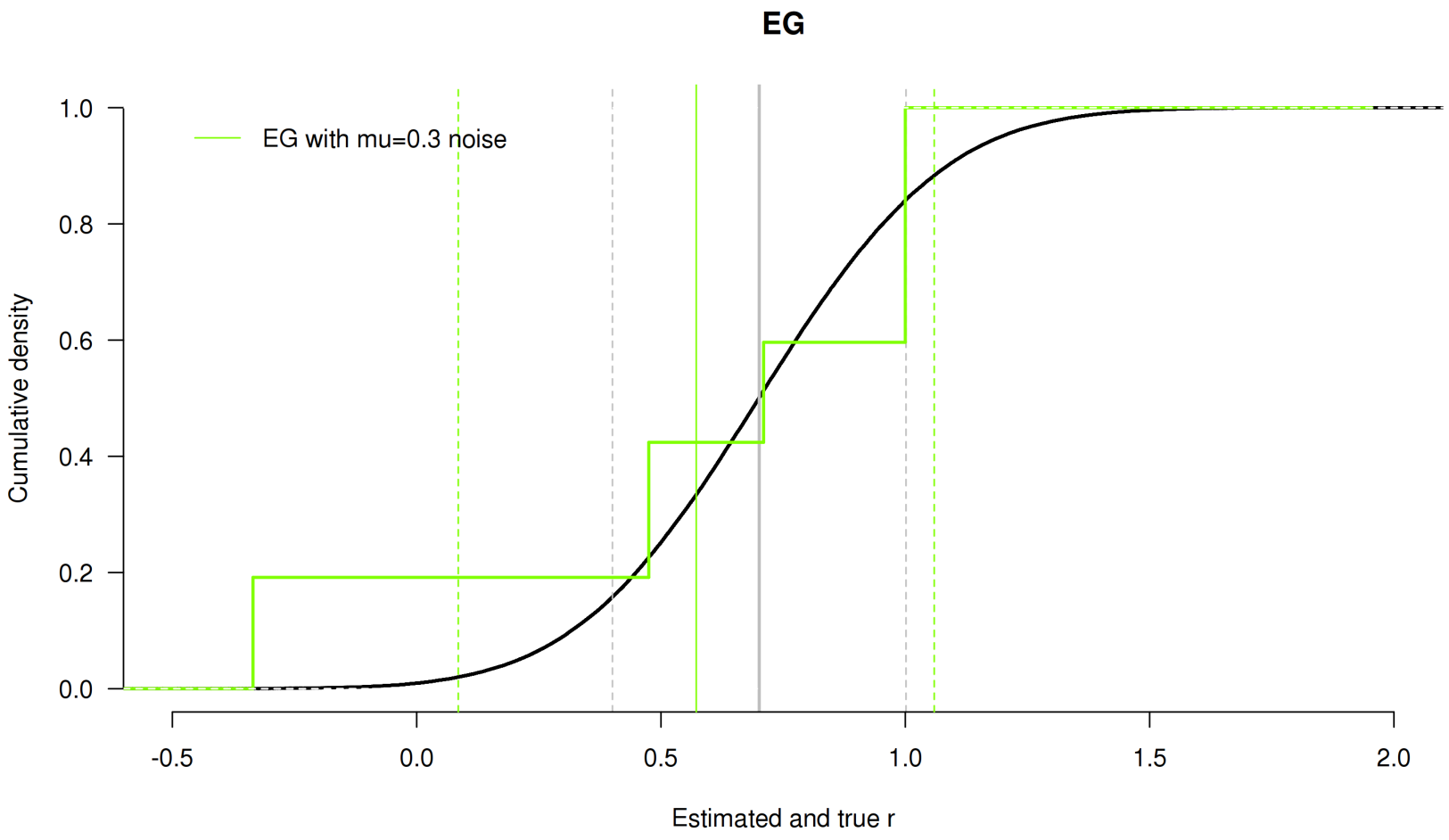

HL
HL with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

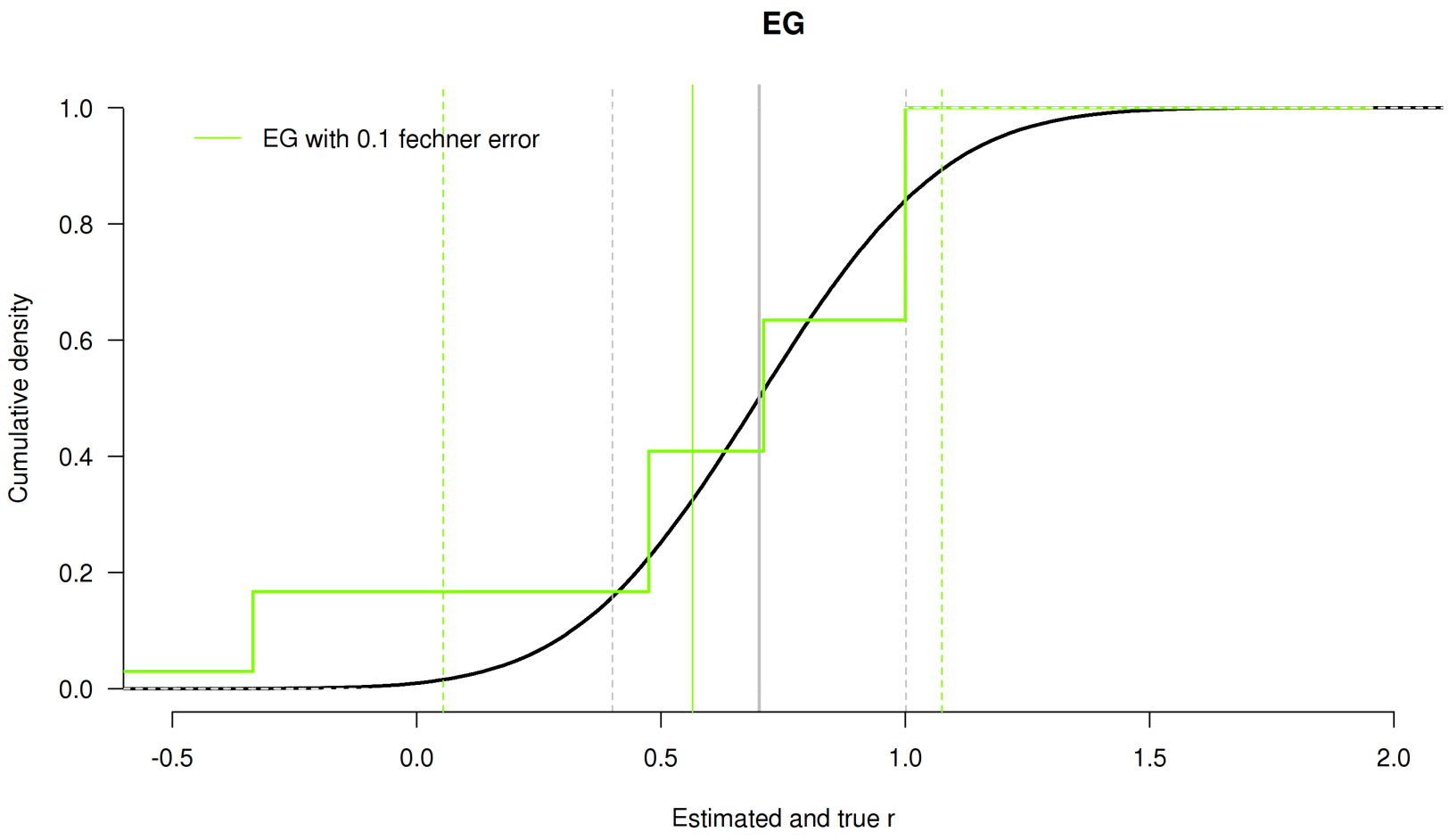

HL
HL with 0.1 fechner error
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

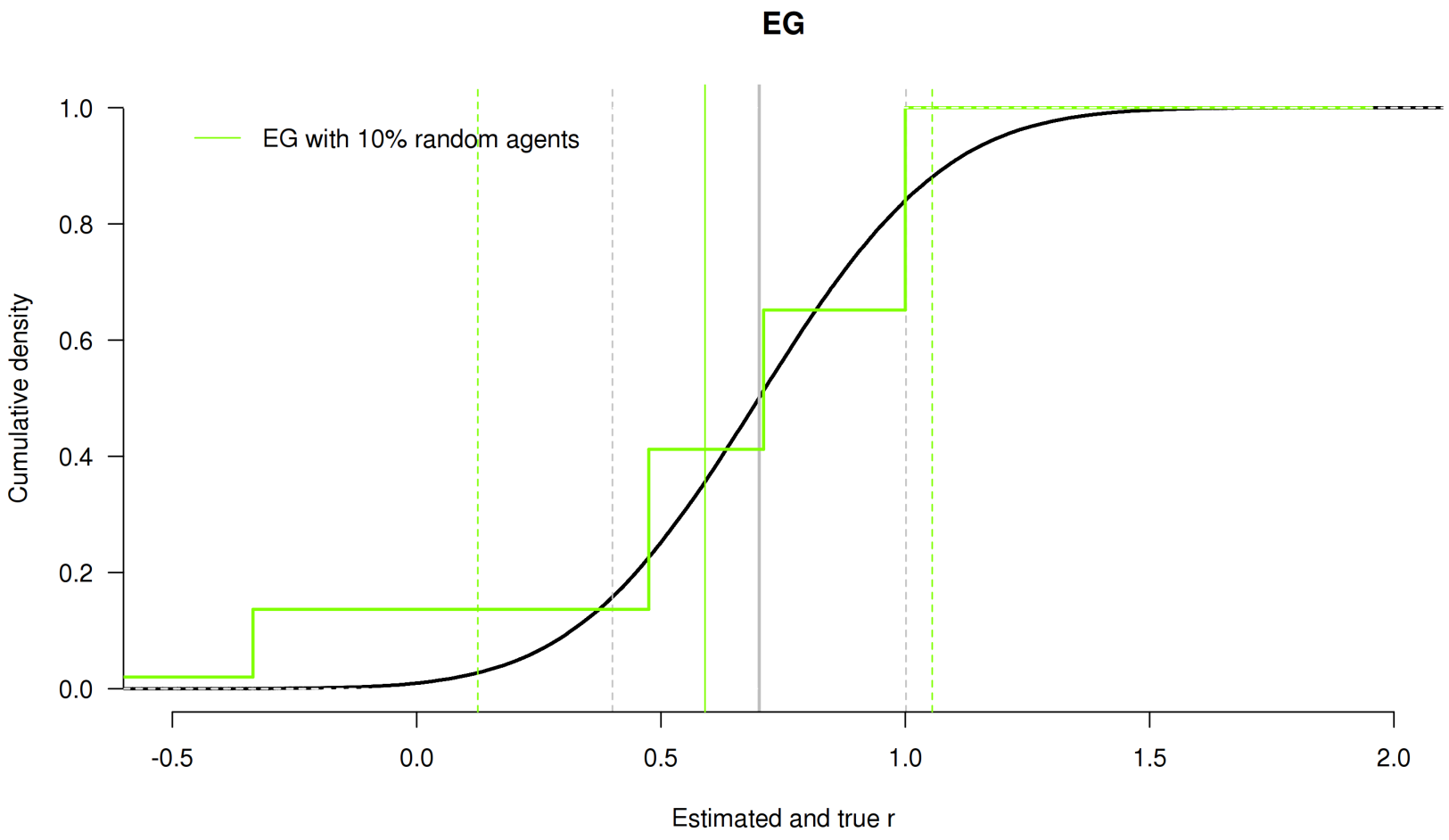

HL
HL with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

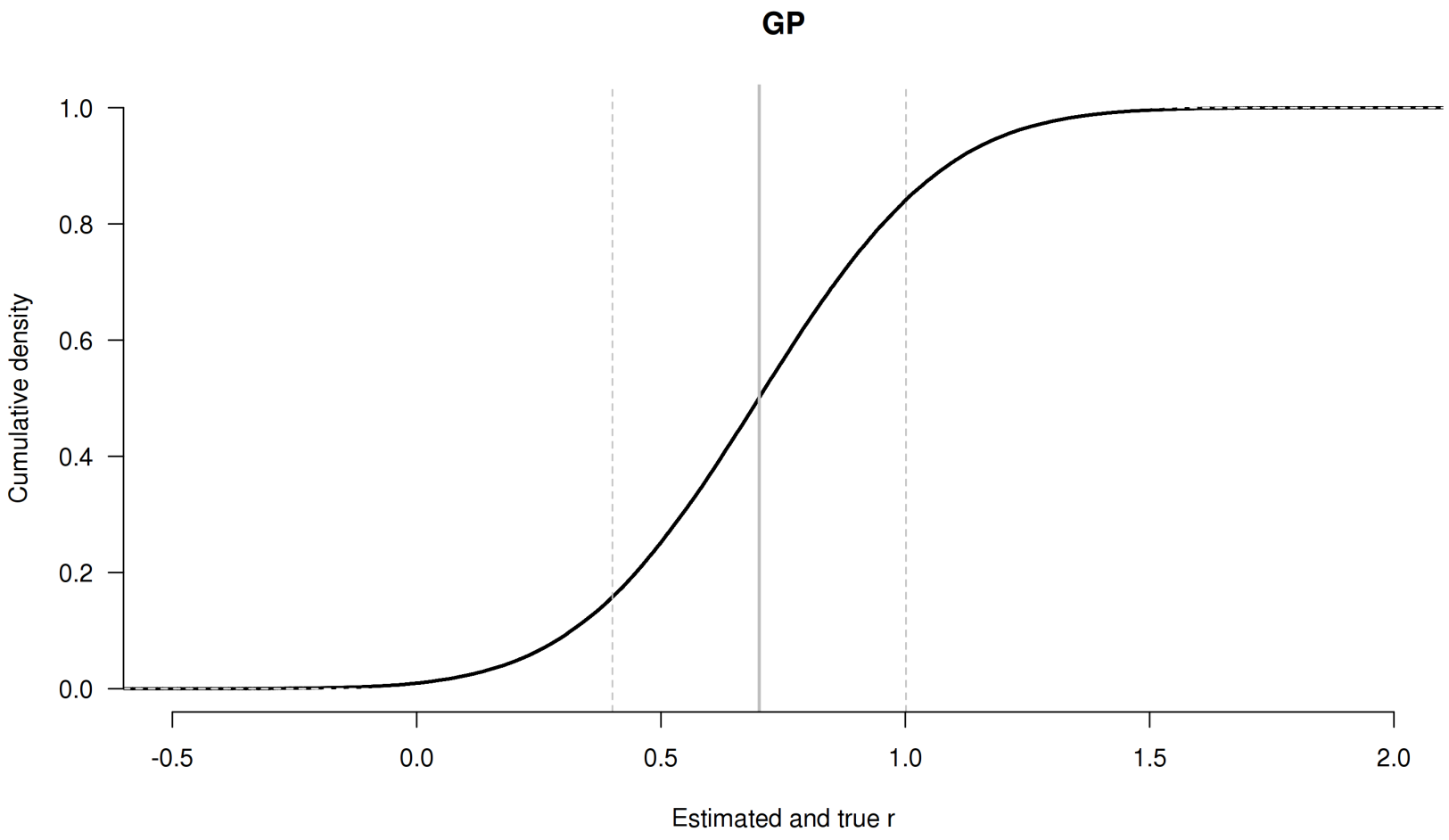

EG
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

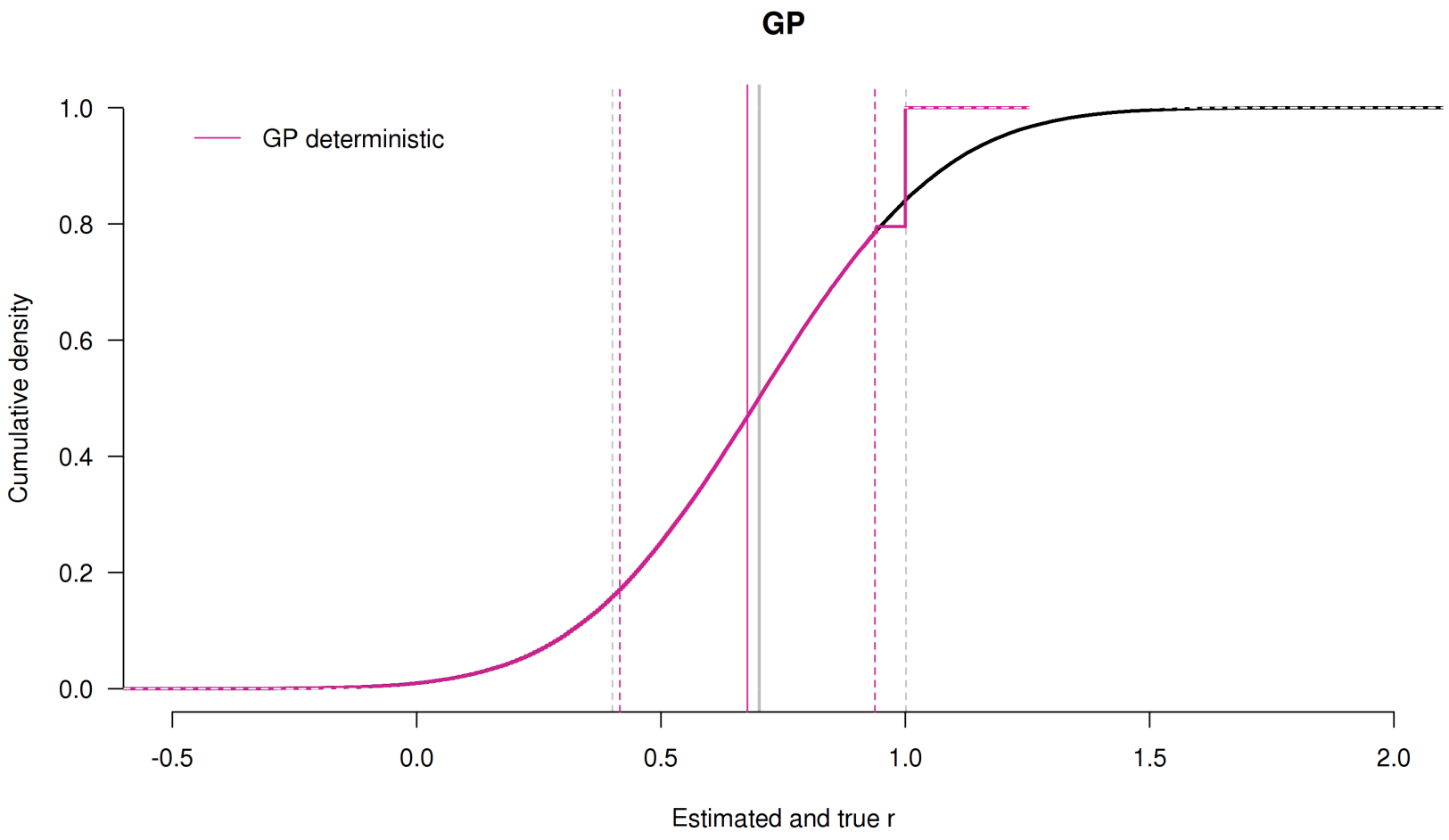

EG
EG deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

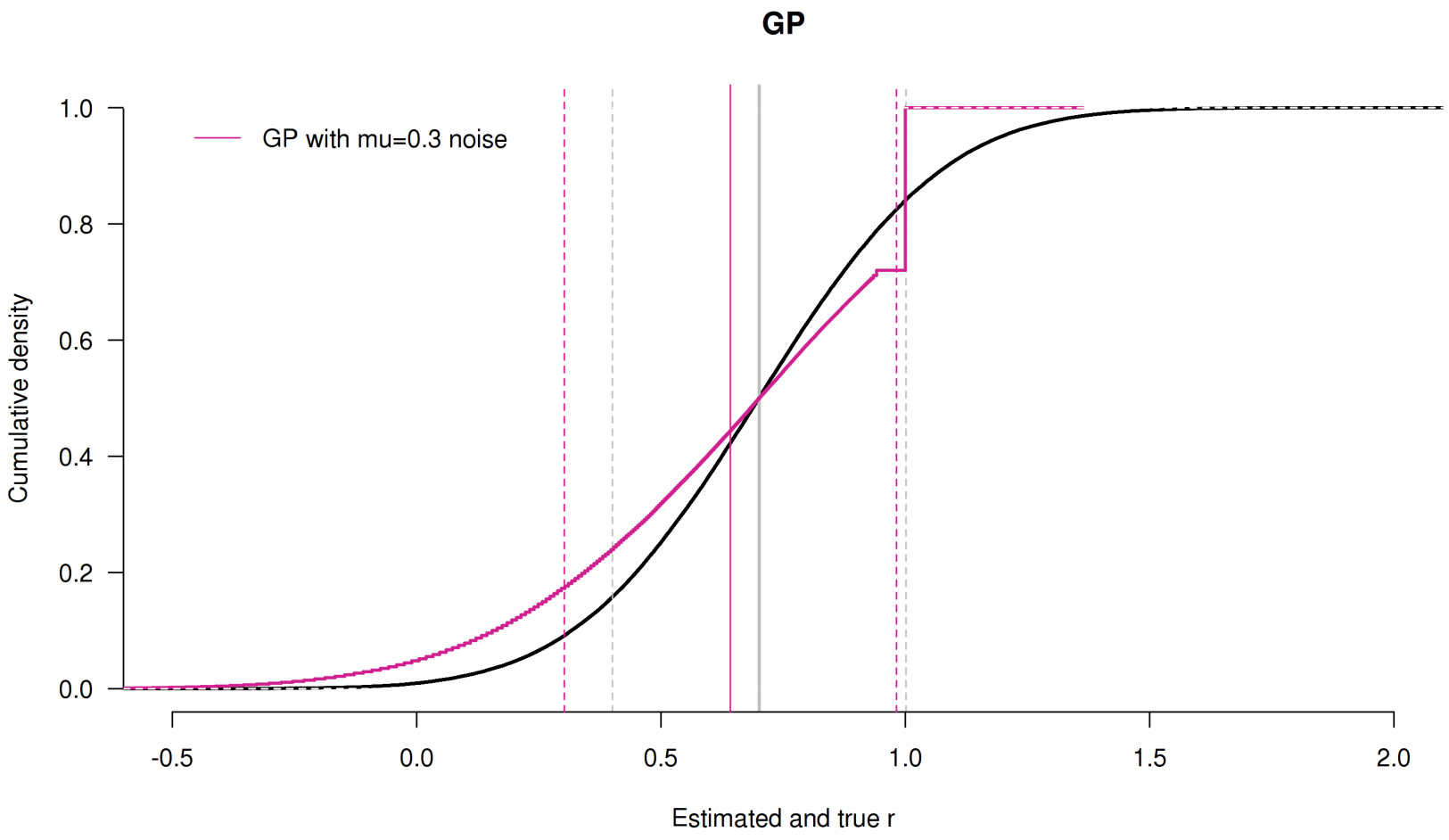

EG
EG with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

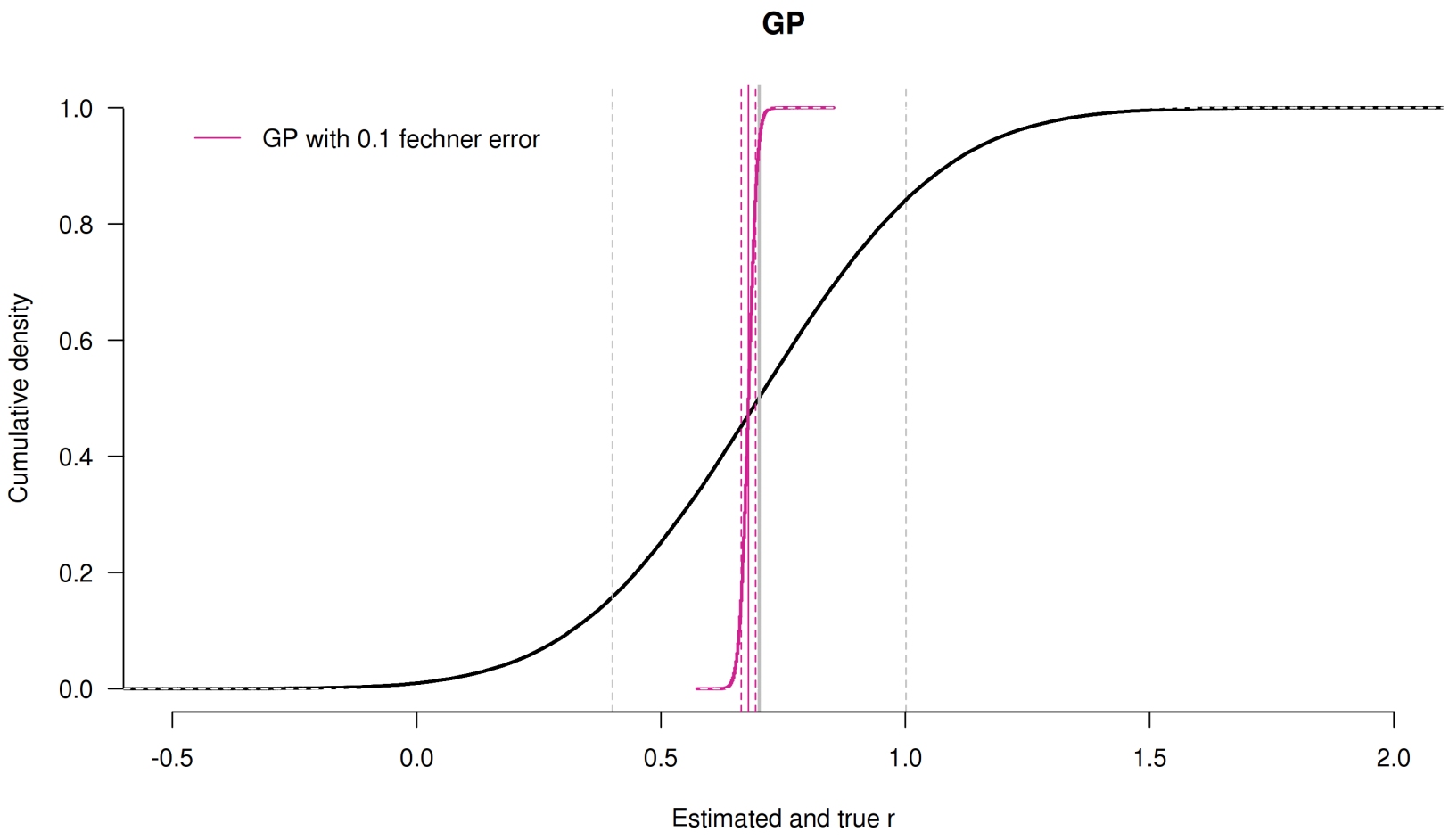

EG
EG with 0.1 fechner error
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

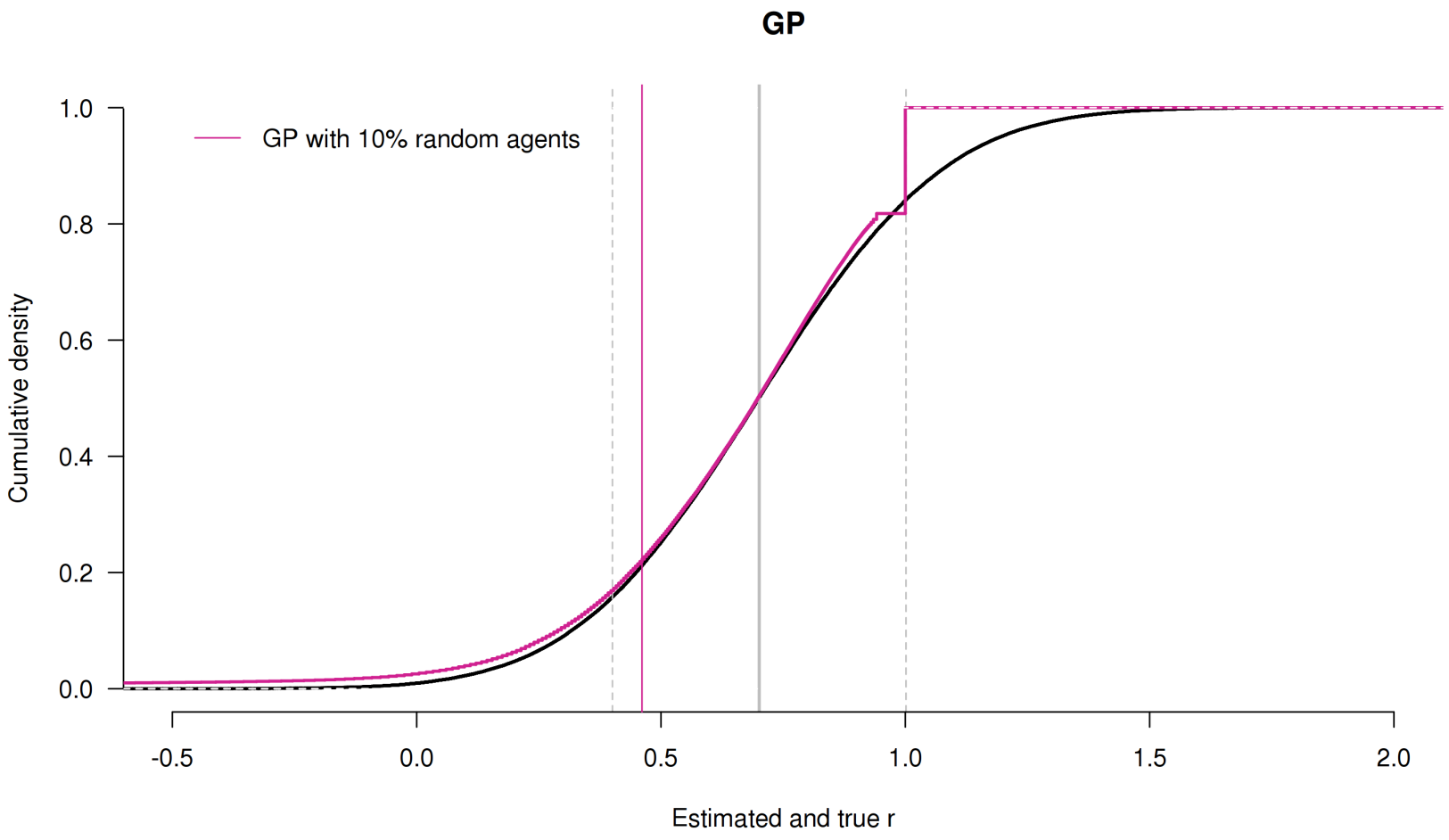

EG
EG with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

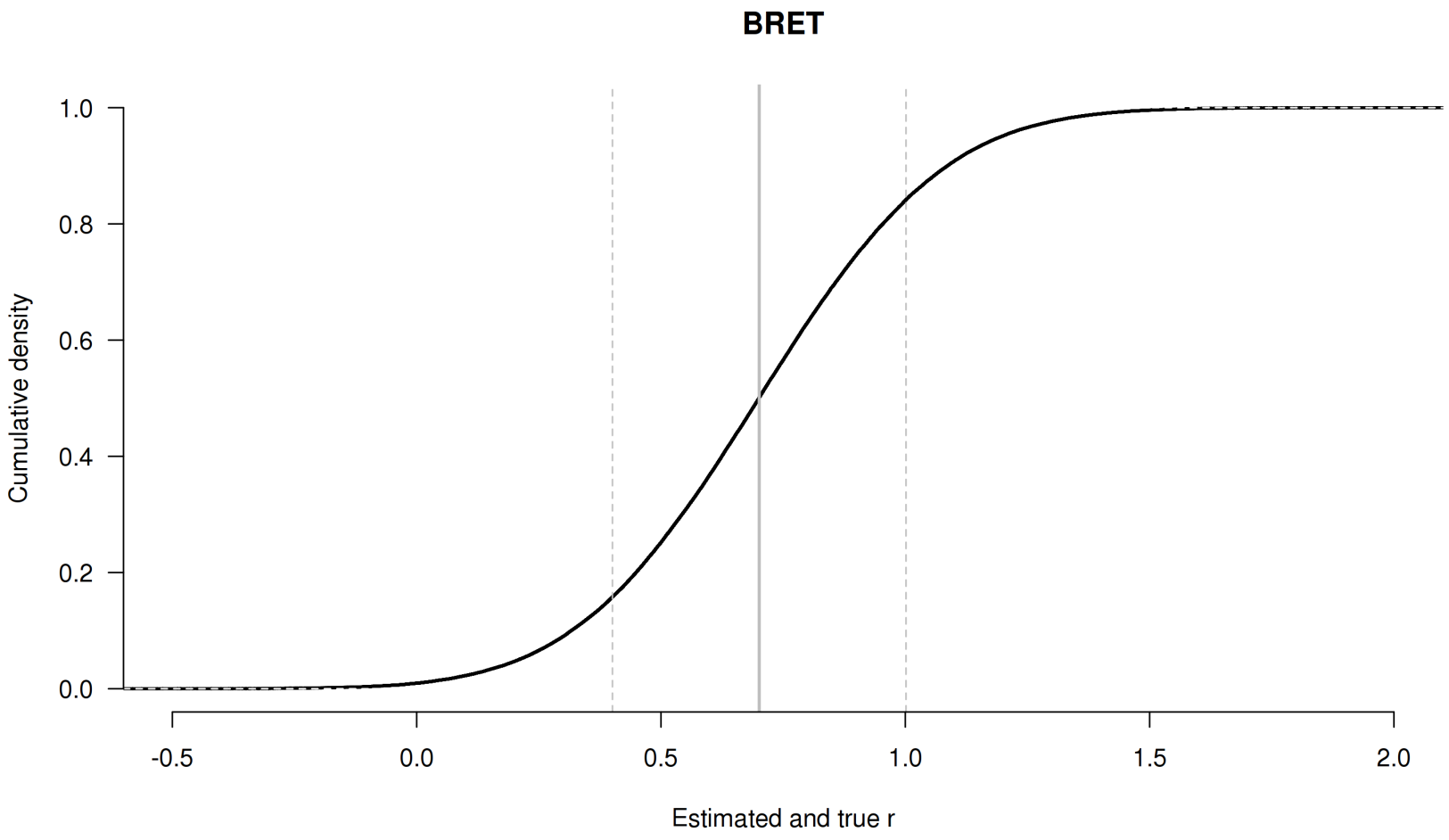

GP
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

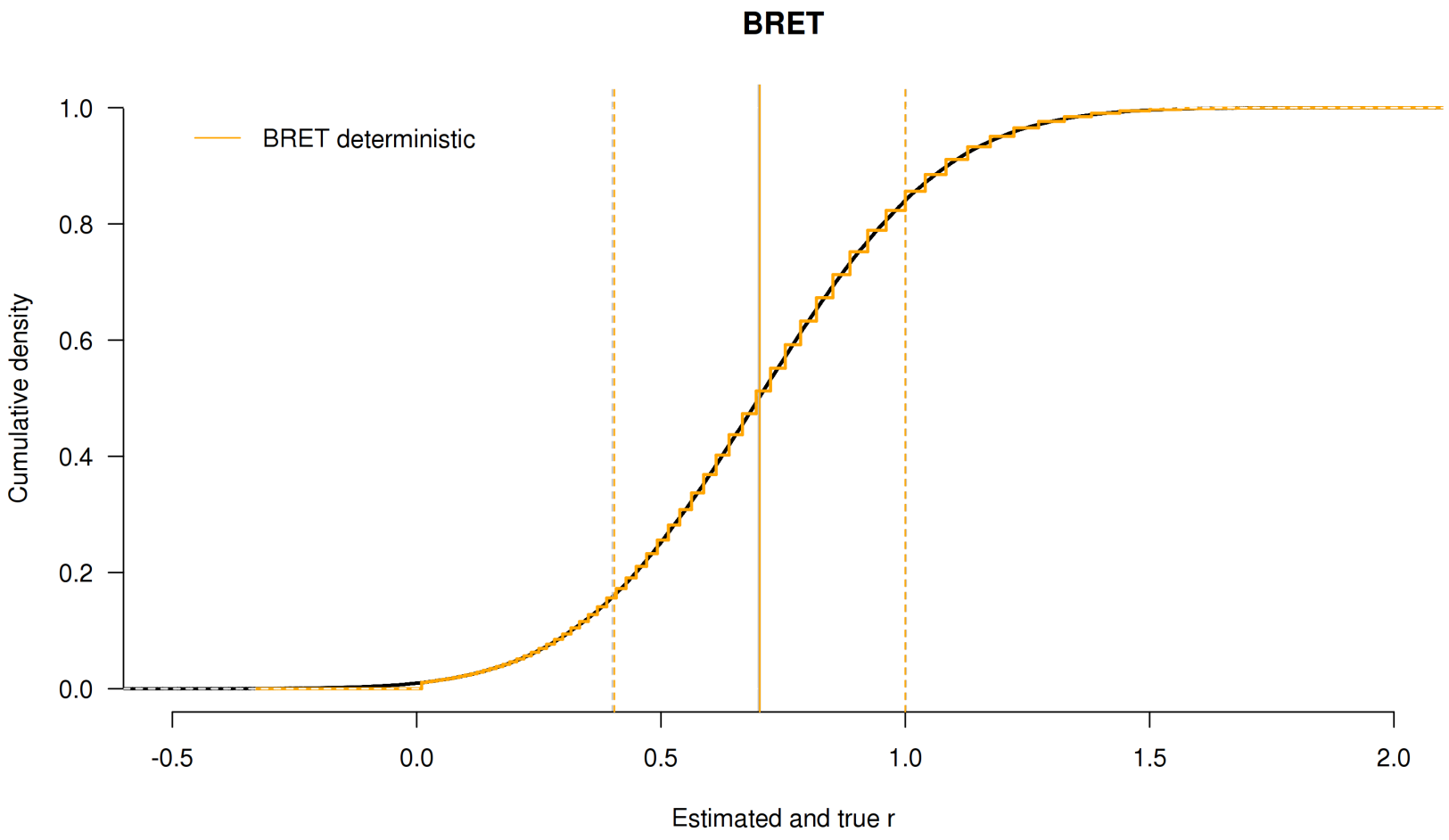

GP
GP deterministic
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

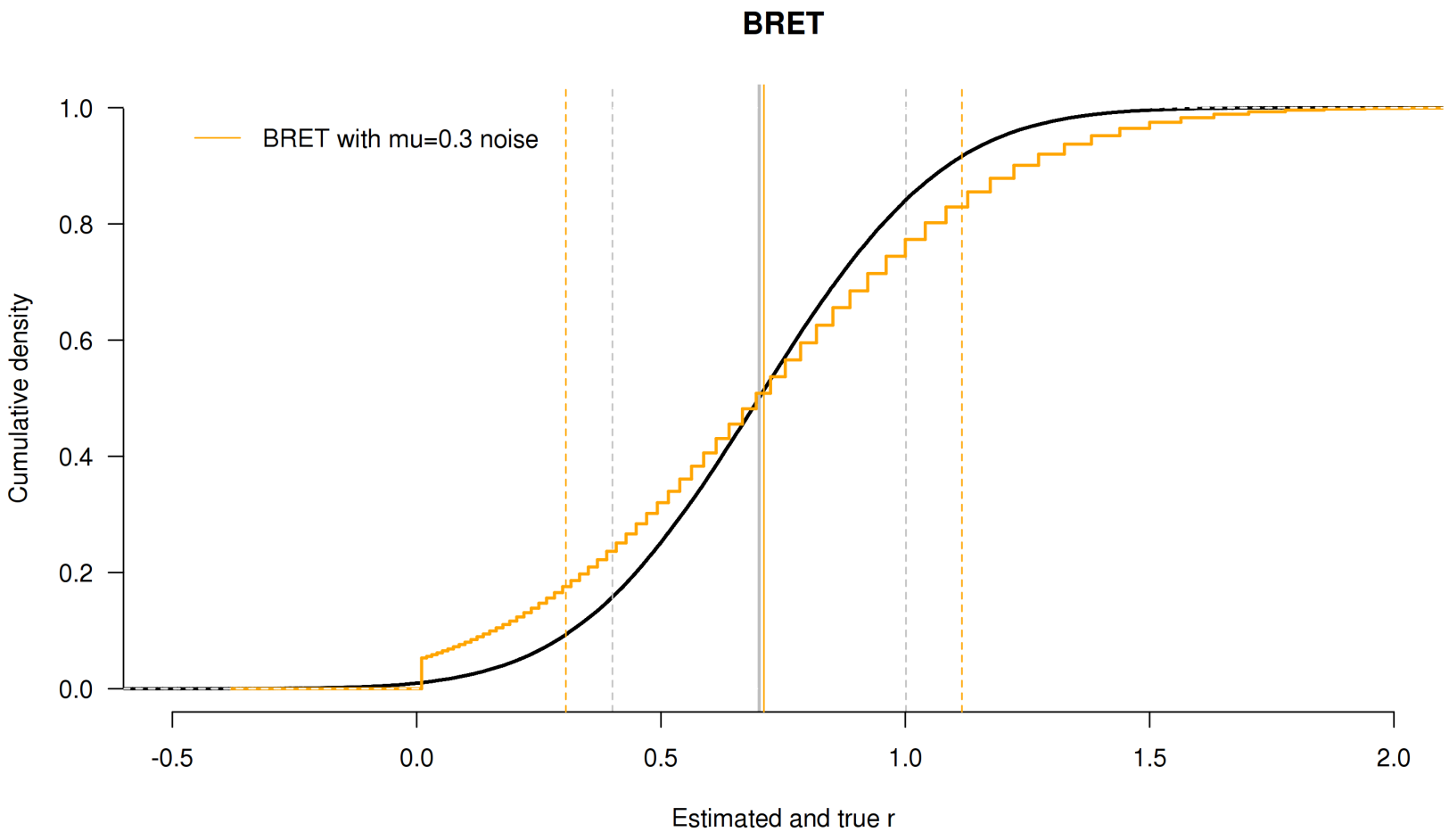

GP
GP with mu=0.3 noise
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

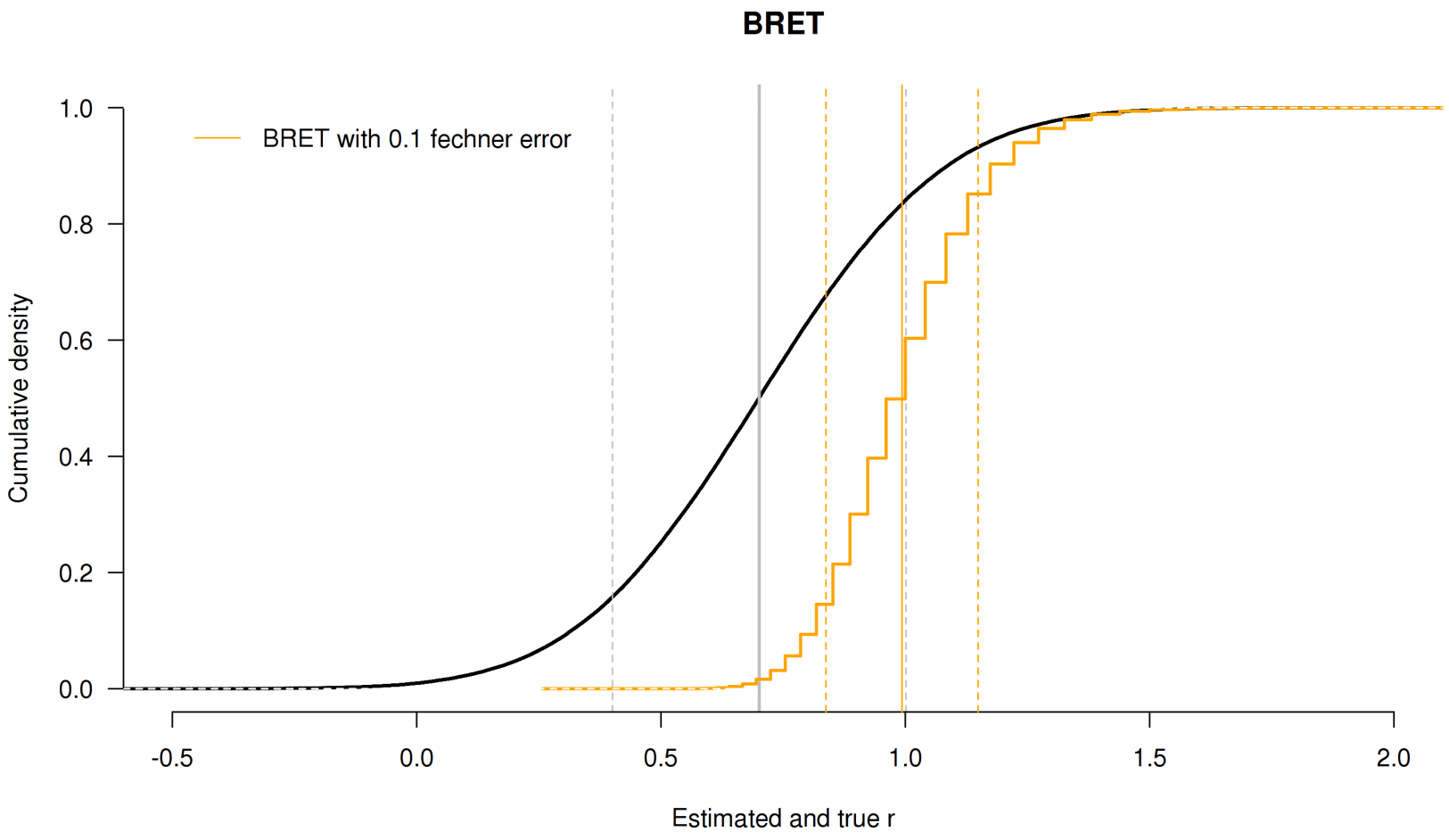

GP
GP with 0.1 fechner error
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

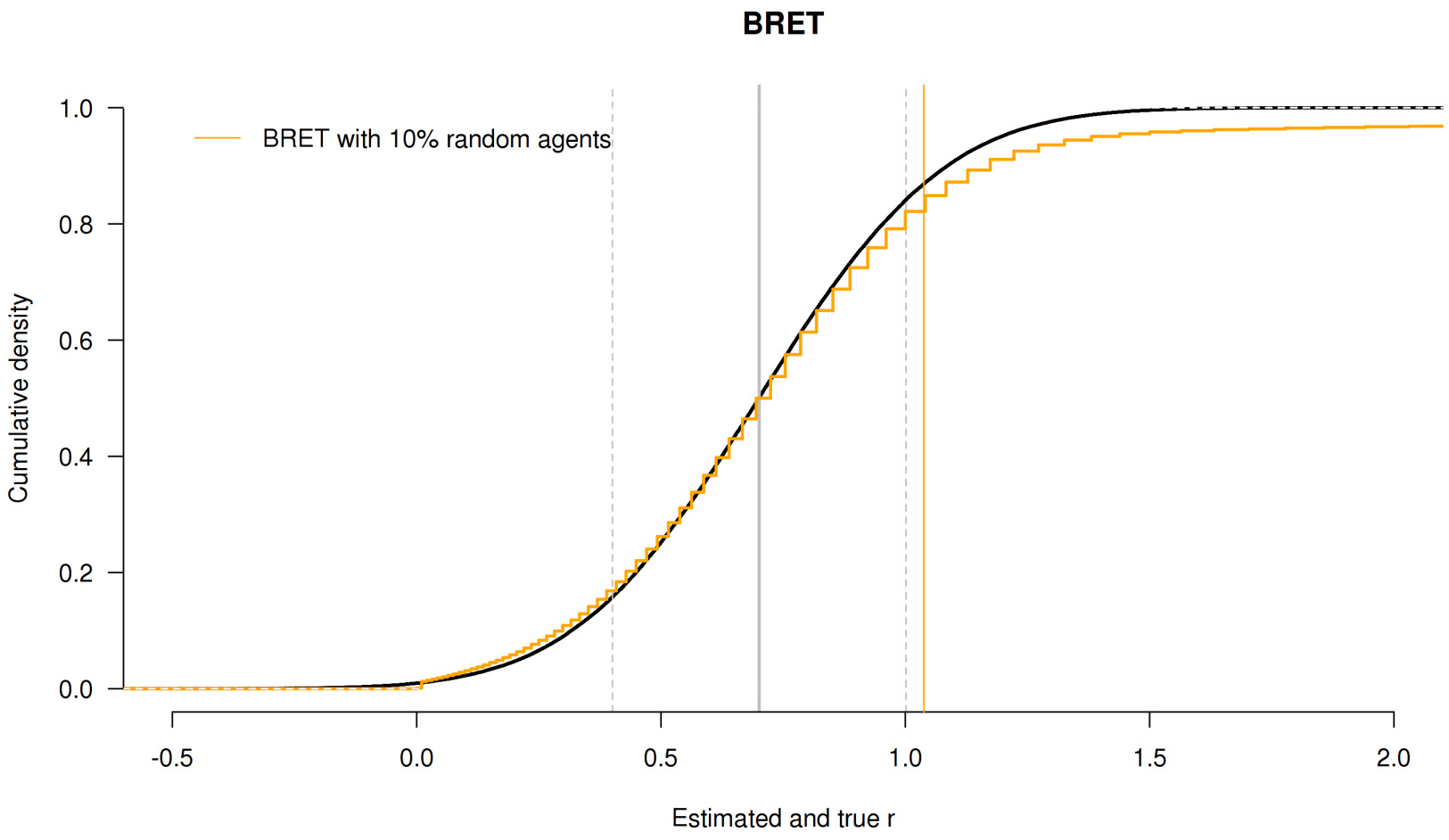

GP
GP with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r


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 0.1 fechner error
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

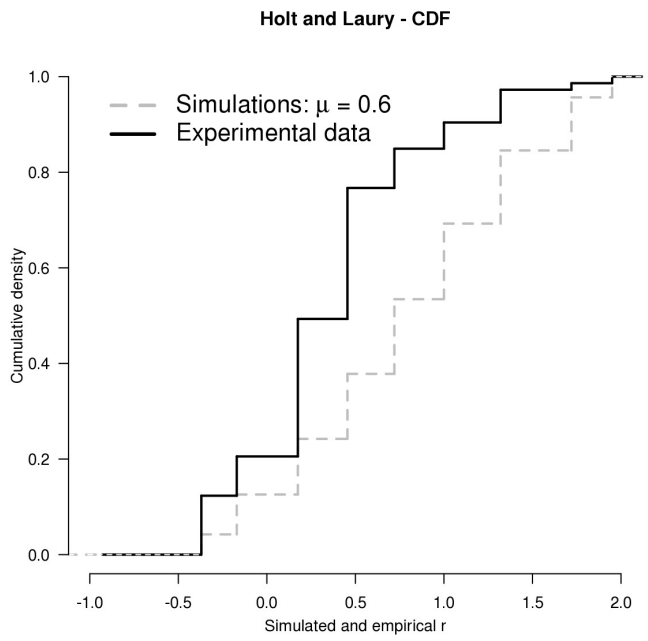

BRET
BRET with 10% random agents
Cumulative density
1.0
0.8
0.6
0.4
0.2
0.0
-0.5
0.0
0.5
1.0
1.5
2.0
Estimated and true r

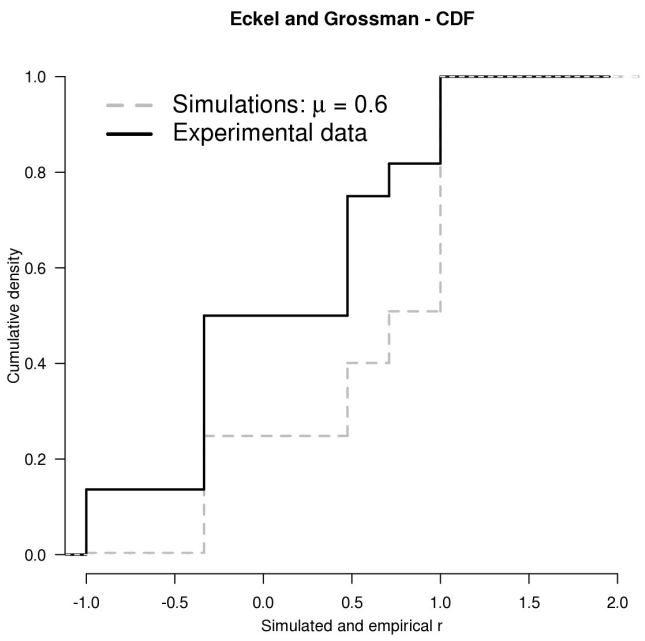
Simulation results: table

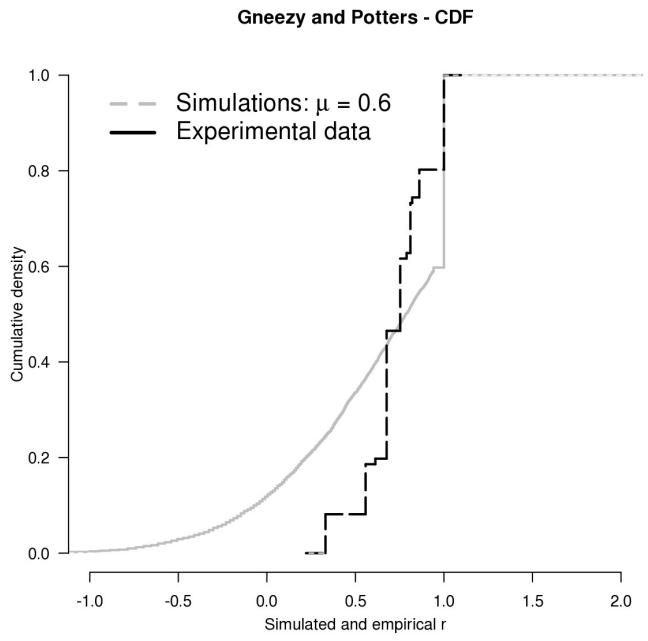
	$\mu = 0$			$\mu = 0.3$			$\mu = 0.6$		
	median	*mean*	*stdev*	*median*	*mean*	*stdev*	*median*	*mean*	*stdev*
True r	**.700**	**.701**	**.300**	**.700**	**.701**	**.300**	**.700**	**.701**	**.300**
HL	.72	.7026	.318	.72	**.704**	.439	.72	**.712**	.627
EG	.71	.636	.399	.71	.572	.487	.71	.490	.572
CGP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

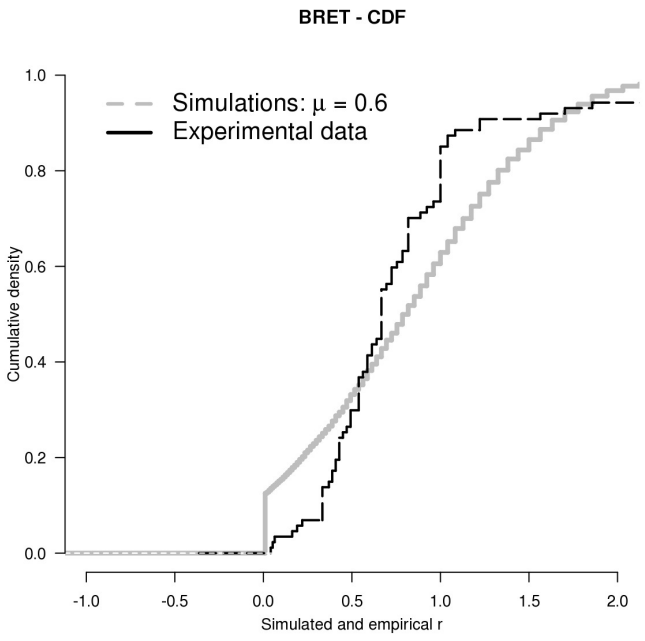
Does all this explain task differences?

- Not completely. 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?


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

Complexity

- We asked to self-report complexity based on a 0-10 complexity scale.
- HL perceived as more difficult
- BRET, CGP and EG lowest point estimate, not significantly different from each other.
- Once we exclude inconsistent subjects from HL, all four tasks undistinguishable

MLE, again

	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.
- Gender difference only in EG, CGP [more later].

Focal points, reference dependence

Conjecture: Although all the tasks are completely framed in the **gain** domain they can entail **reference points** around which subjects could evaluate the other outcomes, perceiving the worse ones as losses.

Reference Point: highest amount of money that a subject can secure with prob. 1.

	Highest sure amount	Focality	Gender gap expected
HL	Low outcome safe lottery	Low	Maybe
EG	Outcome degenerate lottery	High	Yes
CGP	Endowment	High	Yes
BRET	None	None	No

The role of gender

Gender differences: a stylized 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)

My main concern is that your paper gives the impression as if there were no gender differences in risk attitudes. I think you must make very clear in the paper that this is not true.
(An anonymous referee for another paper, 2014)

Where does this evidence come from?

- surveys by Eckel and Grossman (2008) and Croson and Gneezy (2009). But coverage is limited (16 and 10 papers).
- **Eckel Grossman:** Sizeable differences emerge both in the experiment presenting the task (Eckel and Grossman, 2002, 2008a) as well as in almost all the later replications.
- **Investment Game:** A survey by Charness and Gneezy (2012) reports that males invest sizeably, systematically, and significantly more than females.
- self-reported SOEP scale (Dohmen et al. 2011)

But...

- psychology survey: only somewhat more than half of >150 studies reported by Birnes et al. (1999)
- **BRET:** In the BRET task the absence of gender differences is a robust result (Crosetto and Filippin, 2013b). New task but $N > 1000$.
- **Holt and Laury:** usually not found. But hold on for more on this...

Article	N	significant gender difference	lab?	subjects	country	type of evidence	p-value
Holt and Laury (2002)	212	mixed	lab	students	U.S.	text	-
Agnew et al. (2008)	845	yes	lab	non-student	U.S.	t-test	< 0.05
Brañas-Garza and Rustichini (2011)	188	yes	lab	-	Spain	Mann-Whitney	0.0027
Dave et al. (2010)	801	yes	lab	labor force	Canada	coefficient	0.001
Chen et al. (2013)	80	mixed	lab	students	U.S.	various	-
Menon and Perali (2009)	6496	mixed	field	high school	Italy	various	-
Andersen et al. (2006)	90	no	lab	students	Denmark	coefficient	0.38
Anderson and Freeborn (2010)	140	no	lab	students	U.S.	coefficient	0.54
Baker et al. (2008)	120	no	lab	students	U.S.	coefficient	0.891
Carlsson et al. (2012)	213	no	field	rural	China	Wilcoxon	0.14
Chakravarty et al. (2011)	74	no	lab	students	India	coefficient	0.644
Drichoutis and Koundouri (2012)	57	no	lab	general	Greece	coefficient	> 0.05
Eckel and Wilson (2004)	232	no	lab	students	U.S.	coefficient	0.586
Ehmke et al. (2010)	345	no	lab	students	CN, F, Niger, U.S.	text	-
Harrison et al. (2005)	178	no	lab	students	U.S.	text	-
Harrison et al. (2013)	108	no	lab	students	Colombia	coefficient	0.78
Houser et al. (2010)	204	no	lab	students	Germany	text	-
Masclet et al. (2009)	144	no	lab	students+	France	coefficient	0.19
Mueller and Schwieren (2012)	127	no	lab	students	Germany	text	-
Ponti and Carbone (2009)	48	no	lab	students	Spain	correlation	-
Viscusi et al. (2011)	144	no	lab	students	U.S.	text	-

Table : Gender results as reported in the HL literature

Why a comprehensive analysis of HL task

It is difficult to draw a conclusion about the HL task relying upon the evidence available in the literature.

1. Low number of contributions reporting about gender differences
2. Results not easy to compare, even when reported (non parametric or parametric tests, multivariate regressions...)
3. Tasks not always homogeneous
4. Treatment of inconsistencies not homogeneous.

We decided to go beyond a meta-analysis and collect the data of the replications

Replications of HL

Published HL replications as of Jan 31st, 2013:	**118**	
of which:		
Not recording gender or single gender	16	
Duplicate dataset	8	
Universe of reference	**94**	100%
of which:		
No response or not sharing the data	40	42.5%
Final dataset	**54**	57.5%
of which:		
Microdata (shared or available online)	48	
Summary statistics	6	

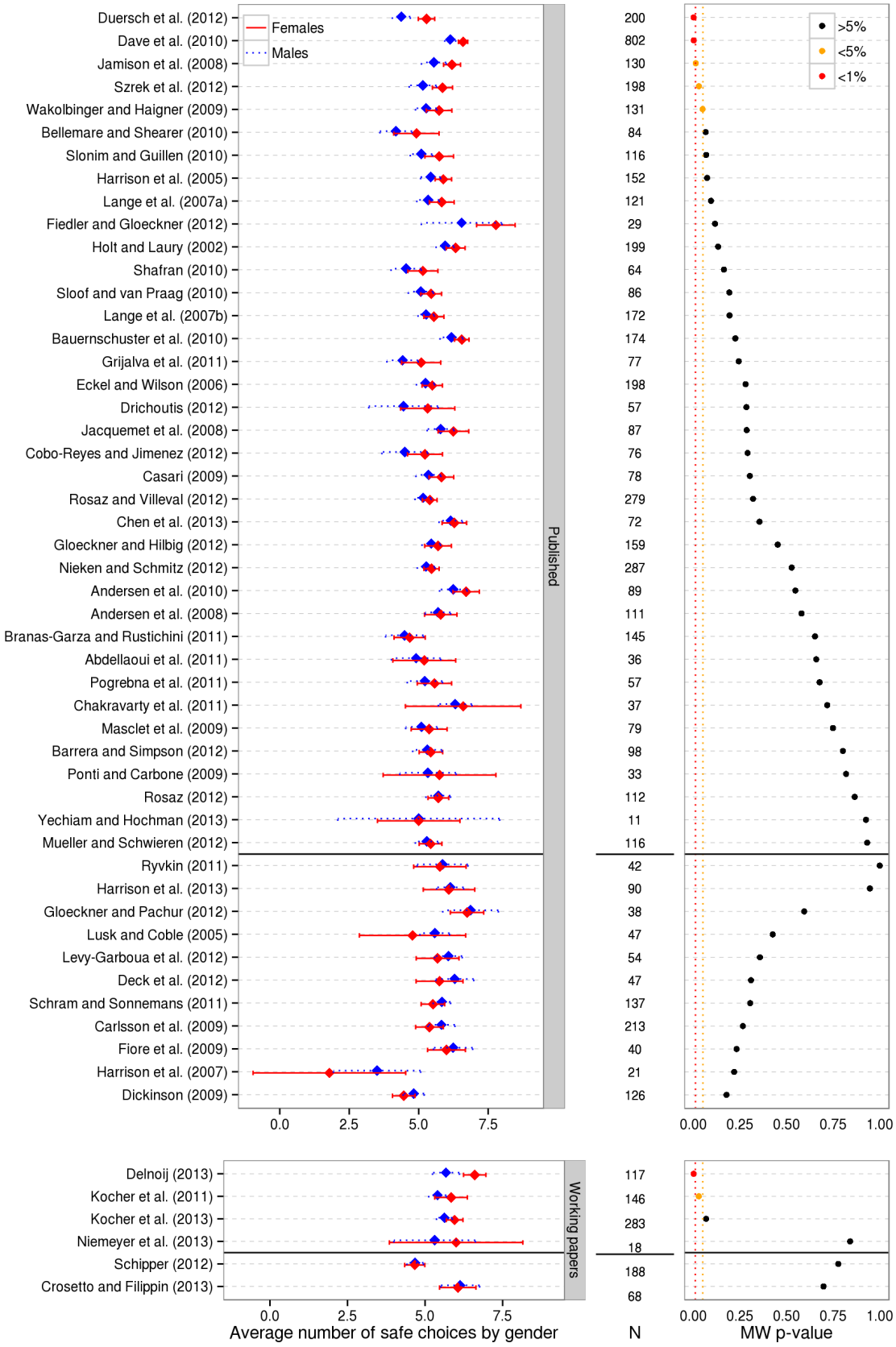
The original Holt and Laury (2002) is included in the 118 replications.

Table : Dataset of HL replications

Subjects in the sample

	Detail	*Consistent subjects*			*Inconsistent subjects*		
		Males	Females	Total	Males	Females	Total
Microdata	*full*	2119	2205	4324	411	502	913
# safe + consistency	*partial*	504	408	912	64	98	162
# safe only		375	324	699	3	1	4
Summary statistics	*summary*	413	359	772	-	-	
Total		3411	3296	**6707**	478	601	**1079**

Table : Subjects in the sample by consistency and type of data. Published papers only.


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)
Andersen et al. (2008)
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. (2013)
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)
Females
Males
Published
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
111
145
36
57
37
79
98
33
112
11
116
42
90
38
47
54
47
137
213
40
21
126
>5%
<5%
<1%
0.0
2.5
5.0
7.5
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

Maximum likelihood model

We assume that subjects are EU maximizers with CRRA preferences $U(x) = x^r$, and that they can make an evaluation error μ when comparing the utility of the two lotteries A (safe) and B (risky). The probability of choosing the safe lottery is

$$Prob(S) = \frac{EU_A^{\frac{1}{\mu}}}{EU_B^{\frac{1}{\mu}} + EU_A^{\frac{1}{\mu}}}, \quad \text{and } EU_i = \sum_j p_j (x_j)^r,$$

The probability converges to $\frac{1}{2}$ as $\mu \to \infty$, and, as $\mu \to 0$, to 1 if $EU_A > EU_B$ and to 0 if $EU_A < EU_B$.
We estimate over the whole dataset a structural model of choice using maximum likelihood and clustering standard errors by subject.

Maximum likelihood results

		Coeff.		*St.Err.*
r	constant	0.640	***	(0.0179)
	female	-0.0633	**	(0.0203)
	realmoney/100	-0.457	***	(0.1028)
	realmoney^2/100	0.00158	***	(0.0003)
	randomorder	-0.0950	*	(0.0392)
	exchange/1000	0.00348		(0.0313)
μ	constant	0.229	***	(0.0073)
	female	-0.0135		(0.0085)
	realmoney/100	-0.19	***	(0.0247)
	realmoney^2/100	0.000658	***	(0.0000)
	randomorder	0.012		(0.0160)
	exchange/1000	0.00861		(0.1160)
N decisions				52735
n subjects				5237
Log-likelihood				-23494.025
Wald χ^2 p-value				0.000

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. t statistics in parentheses.

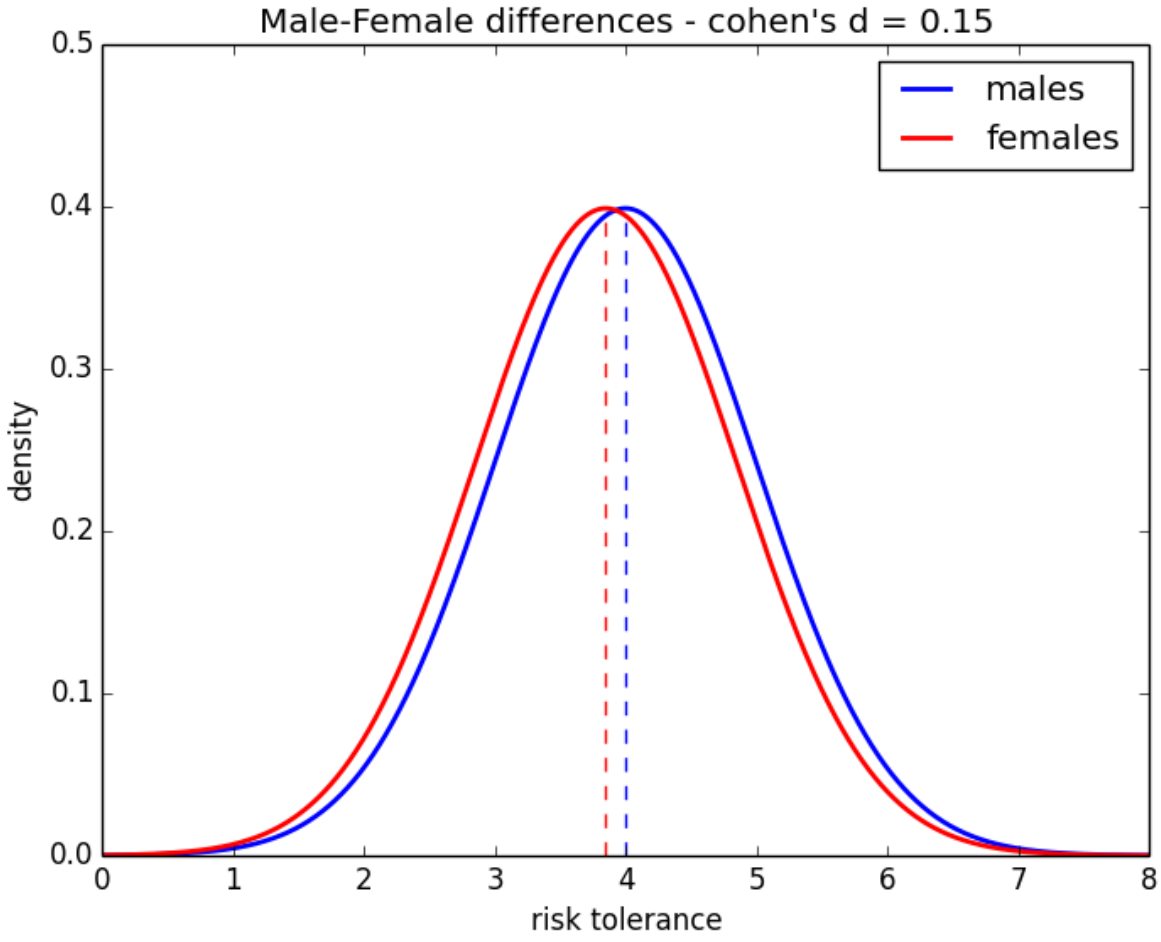
Merging the dataset 2

Focus on gender differences:

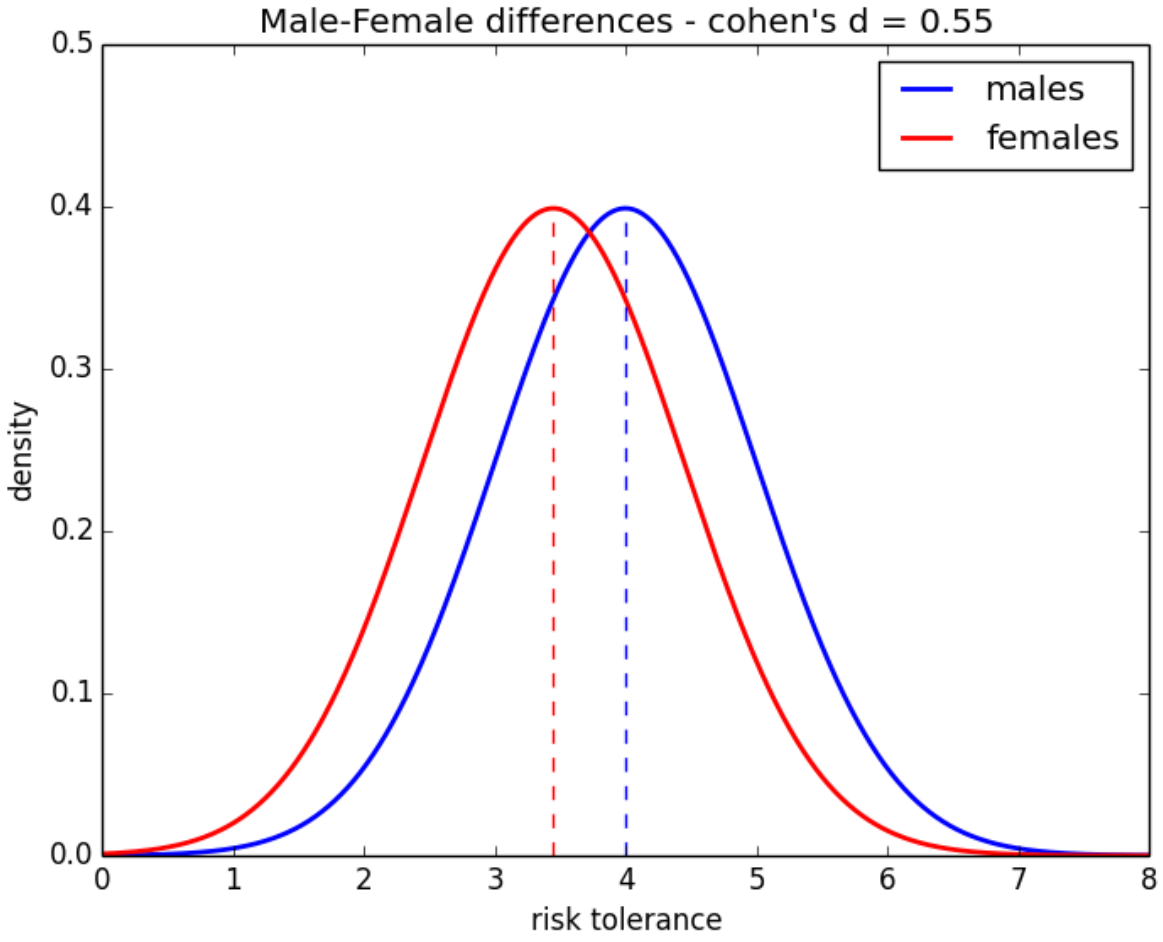
- Pooling all the microdata boosts the statistical power and gender differences become significant ($p < 0.001$)
- Results are confirmed in a Maximum Likelihood specification allowing for decisions with errors.
- The magnitude is very small, however: Cohen's $d = 0.15$ (about 15% of a standard deviation), below the threshold commonly used to define a small effect (0.2)
- The same statistic computed for the Investment Game and for the EG gives a Cohen's $d = 0.55$, above the threshold commonly used to define a medium effect (0.5)

Punch line: the likelihood of observing gender differences as well as their magnitude show clearly different patterns across elicitation methods. This seems to reflect some measurement error induced by the (features of the) tasks rather than genuine differences in risk aversion.
These findings are confirmed by our own experimental replications of the aforementioned tasks.

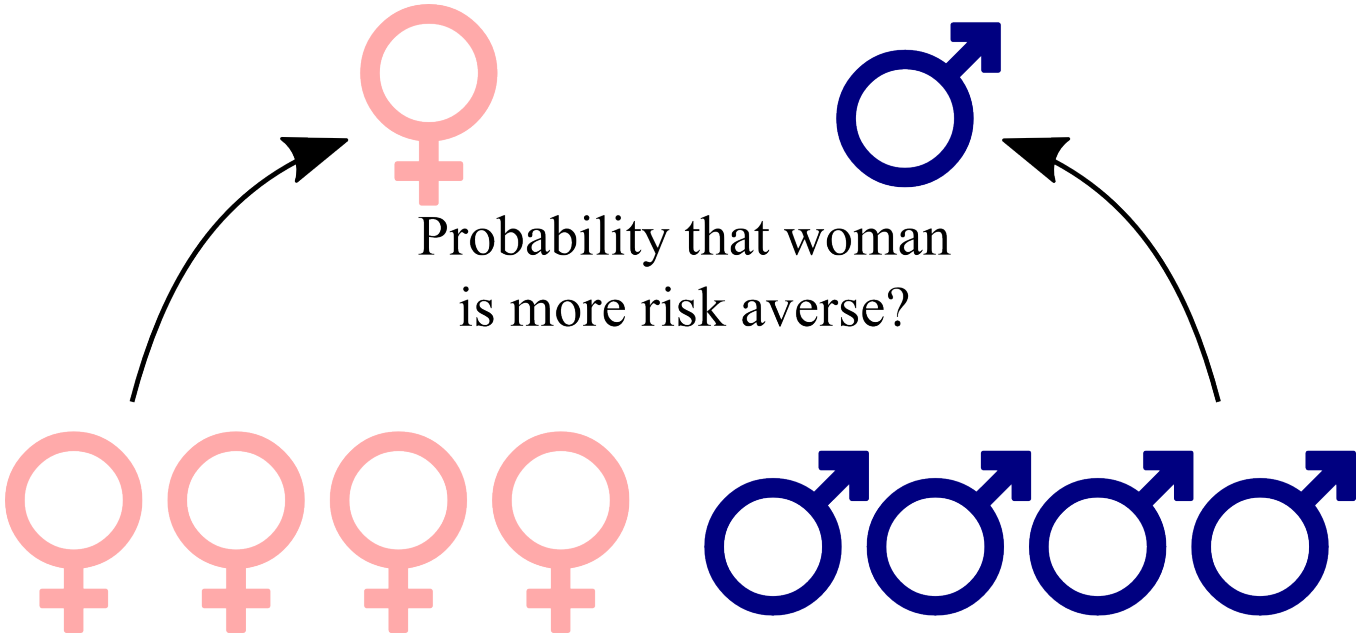
Cohen's D - I


Male-Female differences - cohen's d = 0.15
males
females
density
risk tolerance
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
6
7
8

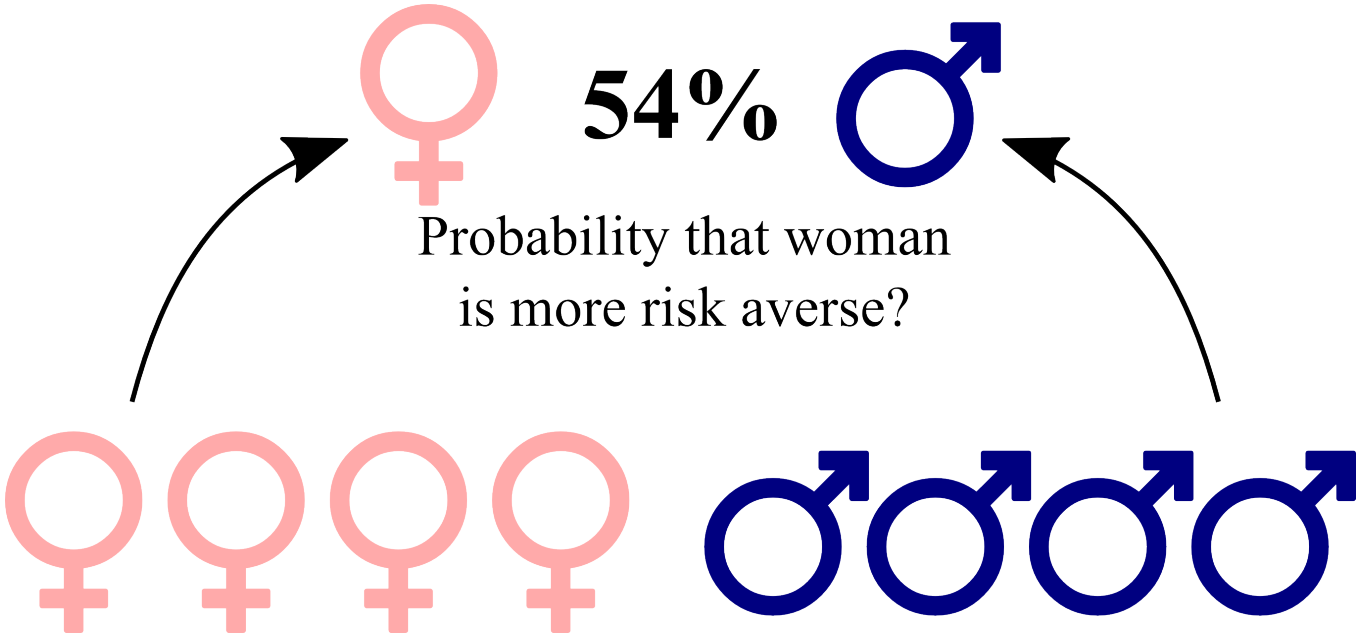
Cohen's D - II


Male-Female differences - cohen's d = 0.55
males
females
density
risk tolerance
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
6
7
8

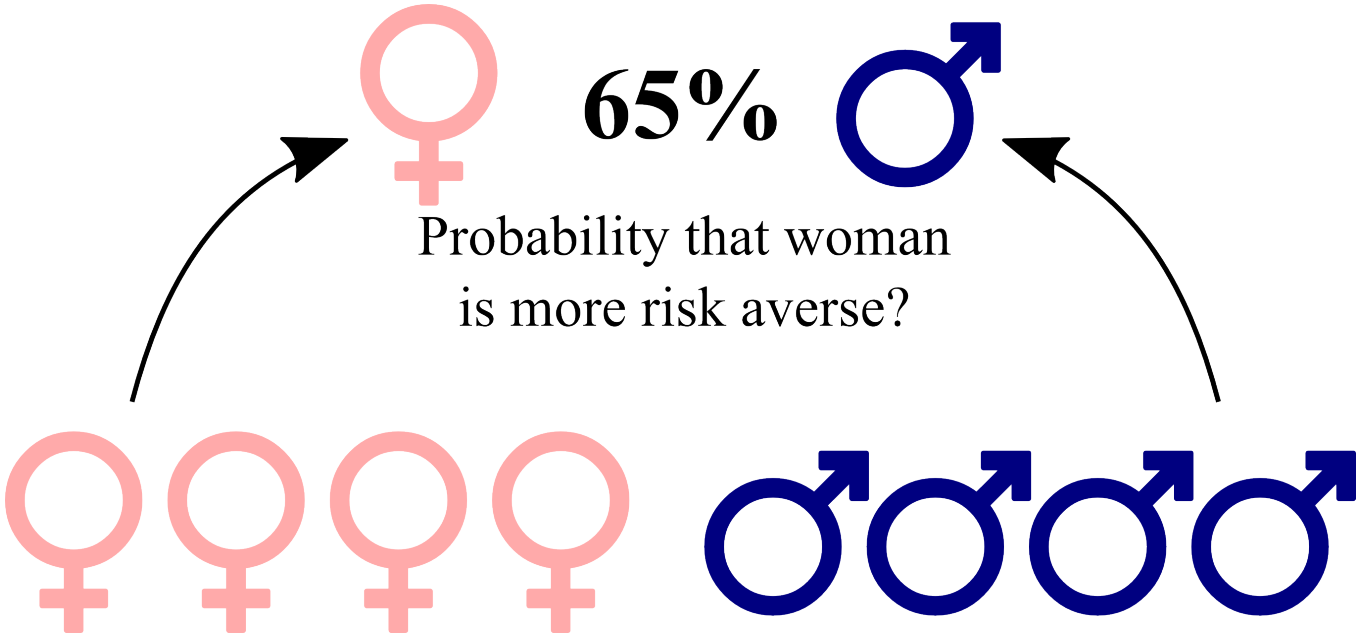
Implication



Implication - cohen's D = 0.15



Implication - cohen's D = 0.55



I. Replication of the tasks: results by gender

	Males		Females		Mann
	N	Mean choice	N	Mean choice	Whitney
HL	31	3.74	42	3.57	$\|p\|$=0.9090
EG	45	3.22	43	2.34	$\|p\|$=0.0050
CGP	37	2.93	49	2.25	$\|p\|$=0.0021
BRET	32	39.72	55	40.25	$\|p\|$=0.7913

Table : Choice by gender

Gender difference appears in tasks that provide a clear reference point.

Conclusion

Conclusion

We compared four tasks, including the new 'Bomb' task.

1. The elicited risk preferences crucially depend on the task used.
2. Some tasks cannot identify risk lovers $\Rightarrow$ demand effect? Bias?
3. Focal points might trigger certainty effects or loss aversion even in the gain domain.
4. In some tasks there is a gender difference, in others not.

One single **best** task does not exist. Trade-offs do. Be aware of biases.

Thanks!

Agnew, Julie R., Lisa R. Anderson, Jeffrey R. Gerlach, and Lisa R. Szykman, "Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults," *The American Economic Review*, 2008, *98* (2), pp. 418–422.

Andersen, Steffen, Glenn Harrison, Morten Lau, and E. Rutström, "Elicitation using multiple price list formats," *Experimental Economics*, 2006, *9*, 383–405.

Anderson, LisaR. and BethA. Freeborn, "Varying the intensity of competition in a multiple prize rent seeking experiment," *Public Choice*, 2010, *143*, 237–254.

Baker, Ronald J., Susan K. Laury, and Arlington Walton Williams, "Comparing Small-Group and Individual Behavior in Lottery-Choice Experiments," *Southern Economic Journal*, 2008, *75* (2), 367–382.

Brañas-Garza, Pablo and Aldo Rustichini, "Organizing Effects of Testosterone and Economic Behavior: Not Just Risk Taking," *PLoS ONE*, 2011, *6* (12), e29842.

Carlsson, Fredrik, Haoran He, Peter Martinsson, Ping Qin, and Matthias Sutter, "Household decision making in rural China: Using experiments to estimate the influences of spouses," *Journal of Economic Behavior & Organization*, 2012, *84* (2), 525–536.

Chakravarty, Sujoy, Glenn W. Harrison, Ernan E. Haruvy, and Elisabet E. Rutström, "Are You Risk Averse Over Other People's Money?," *Southern Economic Journal*, 2011, *77* (4), 901 – 913.

Chen, Yan, Peter Katuščák, and Emre Ozdenoren, "Why Can't a Woman Bid More Like a Man?," *Games and Economic Behaviour*, 2013, *77* (1), 181–213.

Dave, Chetan, Catherine Eckel, Cathleen Johnson, and Christian Rojas, "Eliciting risk preferences: When is simple better?," *Journal of Risk and Uncertainty*, 2010, *41* (3), 219–243.

Drichoutis, Andreas C. and Phoebe Koundouri, "Estimating risk attitudes in conventional and artefactual lab experiments: The importance of the underlying assumptions," *Economics - The Open-Access, Open-Assessment E-Journal*, 2012, *6* (38), 1–15.

Eckel, Catherine C. and Rick K. Wilson, "Is trust a risky decision?," *Journal of Economic Behavior & Organization*, 2004, *55* (4), 447 – 465.

Ehmke, Mariah, Jayson Lusk, and Wallace Tyner, "Multidimensional tests for economic behavior differences across cultures," *The Journal of Socio-Economics*, 2010, *39* (1), 37 – 45.

Harrison, Glenn W., Eric Johnson, Melayne M. McInnes, and E. Elisabet Rutström, "Risk Aversion and Incentive Effects: Comment," *American Economic Review*, June 2005, *95* (3), 897–901.

_ , Morten I. Lau, E. Elisabet Rutstrm, and Marcela Tarazona-Gmez, "Preferences over social risk," *Oxford Economic Papers*, 2013, *65* (1), 25–46.

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

Houser, Daniel, Daniel Schunk, and Joachim Winter, "Distinguishing trust from risk: An anatomy of the investment game," *Journal of Economic Behavior & Organization*, 2010, *74* (1-2), 72 – 81.

Masclet, David, Nathalie Colombier, Laurent Denant-Boemont, and Youenn Lohéac, "Group and individual risk preferences: A lottery-choice experiment with self-employed and salaried workers," *Journal of Economic Behavior & Organization*, 2009, *70* (3), 470 – 484.

Menon, Martina and Federico Perali, "Eliciting Risk and Time Preferences in Field Experiments: Are They Related to Cognitive and Non-Cognitive Outcomes? Are

Circumstances Important?," *Rivista Internazionale di Scienze Sociali*, 2009, *117* (3), 593–630.

Mueller, Julia and Christiane Schwieren, "Can personality explain what is underlying women's unwillingness to compete?," *Journal of Economic Psychology*, 2012, *33* (3), 448 – 460.

Ponti, Giovanni and Enrica Carbone, "Positional learning with noise," *Research in Economics*, 2009, *63* (4), 225 – 241.

Viscusi, W., Owen Phillips, and Stephan Kroll, "Risky investment decisions: How are individuals influenced by their groups?," *Journal of Risk and Uncertainty*, 2011, *43* (2), 81–106.