The Bomb Risk Elicitation Task

Paolo Crosetto *and* Antonio Filippin

Max Planck Institute for Economics
and
Università degli Studi di Milano

Nordic Conference,
Bergen, September 28^{th}*, 2012*

Risk Aversion Tasks

Risk elicitation tasks: two approaches

1. When running experiment testing Expected Utility *per se*:
   - give the subjects many (lottery) choices;
   - use the data to estimate all parameters of a utility function (e.g.: $U = x^r$);
   - e.g.: Hey and Orme (1994).
   - Time consuming, many choices, but very precise.
2. When running other experiments (games, auctions...)
   - Use a custom, unrelated control task...
   - ...usually at the end of the experiment;
   - use data as a control.
   - Fast, easier to understand, less precise

Which tasks have been used?

- Most of the task involve lottery choices, as Holt and Laury (2002)
- Some use WTP-WTA or certainty equivalent elicitations, as Becker-DeGroot-Marschak (1964)
- Some use intuitive visual representations, as Lejuez et al (2002)
- In some instances, not incentivised questionnaires are used (Dohmen et al 2011)

In this paper we propose a new incentivised risk elicitation task

How do we evaluate a task?

Should elicit a *complete range* of risk attitudes (risk-averse *and* risk-loving)

Should be *precise*: yielding a fine estimate of risk attitudes

Should be *parsimonious*: should need a low number of choices from subjects

Should be *intuitive*: easy to understand, easy to implement (lab, field)

Should be easy to *manipulate* to allow researchers to tackle several issues

Risk Elicitation Tasks, I: Holt and Laury

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

Figure: An implementation of the Holt and Laury battery of lotteries

Risk Elicitation Tasks, II: Eckel Grossman

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

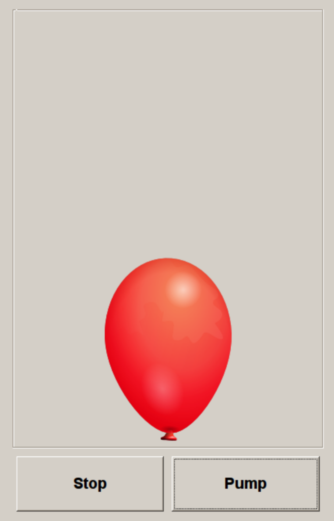
Figure: An implementation of the Eckel and Grossman lottery choice task

Risk Elicitation Tasks,III: Charness and Gneezy

- Subjects are given X tokens as endowment;
- They must allocate it over two accounts (shares a and $X - a$):
- a safe account that has fixed return,
- and a risky account that has returns $\{\frac{1}{2} : 0; \frac{1}{2} : 2.5a\}$.

Solution: risk neutral should invest all as expected value is higher than 1.

Risk Elicitation Tasks, IV: Balloon



Summing up

	Precision (categories of r)	Parsimony (no. of choices)	Intuitiveness	Completeness (r range)	Ambiguity
HL	8	10	low	yes	no
EG	5	1	ok	no	no
CG	100+	1	high	no	no
Balloon	$0 \leq n \leq 128$	$0 \leq n \leq 128$	high	yes (trunc)	yes

Can one do better?

Presenting the BRET

BRET: rules

We developed the Bomb Risk Aversion 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 → earnings equal zero.
- If bomb not collected → earnings equal to number of boxes collected.

BRET: interface, I

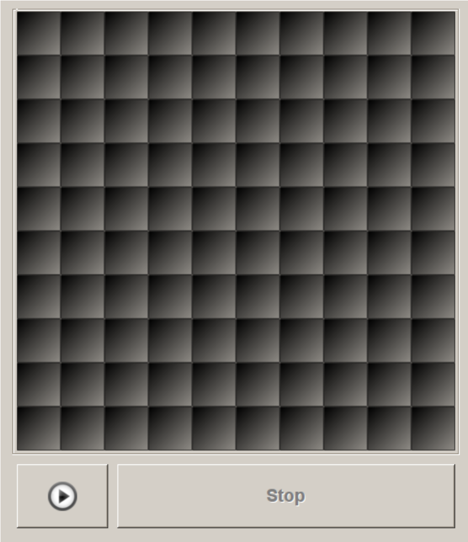


Figure: The BRET interface at the start of the experiment

BRET: under the hood

- Theoretically, the task amounts to choosing the preferred of 100 lotteries.
- Each lottery is characterised as

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

- The 100 lotteries are all summarised by the parameter k...
- ...that is also governing probabilities.
- Example: at $k = 20$, $L = \{20\% : 0\,; \, 80\% : 20\}$

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

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

which implies risk neutral chooses $k^* = 50$.

BRET: features

Features

Precise : can estimate 100 values for r.

Parsimonious : only 1 choice.

Complete : both risk averse and risk loving allowed.

Easy : intuitive, visual, in continuous time.

	Precision	Parsimony	Intuitiveness	Completeness	Ambiguity
BRET	100	1	high	yes	no

Further features

Probability : clear representation of probabilities (100 boxes)

No truncation : data are not truncated as in Balloon

Loss aversion : as no reference point, very likely loss aversion plays no role.

Reduction Axiom : robust to violations of the reduction axiom.

Continuous time : suitable (also) to dynamic environments (stocks, auctions...).

BRET: results

Experimental details, I: procedures, sample

- Experiment run in Jena, March-May 2012
- Vast scale $\rightarrow$ 38 sessions, 1093 subjects
- Small payment $\rightarrow$ Expected Value = 2.5 € (to be used as a control)
- Efforts to test robustness and validate the task

Validation:

- All subjects also answered the DOSPERT (Blais 2006), and
- The SOEP German Panel question (Dohmen et al. 2011)

	Age bracket			**N**
	18 – 22	23 – 27	28 – 60	
Male	158	226	72	449
Female	285	298	54	634
Total	443	524	126	1093

Table: Demographics of the experimental sample

Results: distribution of choices in the Baseline Dynamic

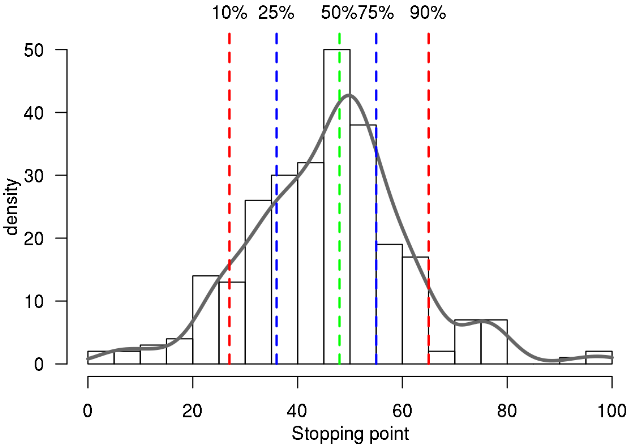


Figure: Distribution of decisions in the dynamic version

Results: details

- Mean k^* is 46.5 – the average subject is slightly risk averse ($r \cong 0.85$)
- 51.3% risk averse ($k \leq 49$), 14.1% risk neutral ($k = 50$), 34.6% risk-seekers ($k \geq 51$)
- 50% between 26 ad 55

In particular

- A higher share of risk lovers w.r.t. other tasks
- possibly consistent with EUT? over small stakes, subjects should be risk neutral
- Large domain of choices generate quite a few extreme outliers

Static paper-and-pencil version

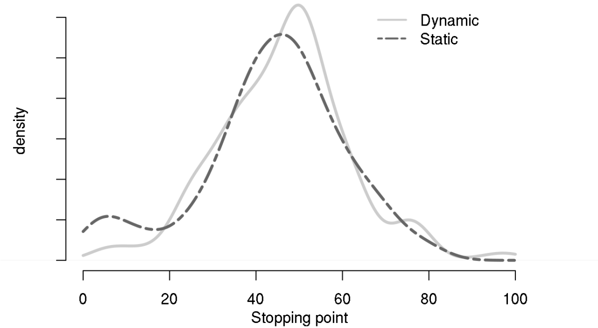
Does the visual representation make a difference?

- We run a *static*, paper-and-pencil treatment
- This amounts to choosing a number $k \in [1, 100]$
- Then, a number $b \sim U[1, 100]$ is drawn
- If $k \geq b \rightarrow \pi = 0$; if $k < b \rightarrow \pi = k$

Results

- There is no difference between the two treatments;
- the paper-and-pencil is possibly *less* intuitive, as subjects make more often extreme choices.
- the Dynamic is more intuitive and allows for more easily implemented manipulation

Results: Baseline Static *vs.* Baseline Dynamic



Treatment	**N**	**Mean**	**Mann-Withney z**	**Complexity**	**Mann-Withney z**
a. full sample					
Static	84	43.7	$Prob > \|z\| = 0.34$	2.50	$Prob > \|z\| = 0.58$
Dynamic	269	46.5		2.44	
b. sensitivity analysis without 2.5% *of obs. in each tail of the distribution in the whole experiment*					
Static	78	45.8	$Prob > \|z\| = 0.91$	2.51	$Prob > \|z\| = 0.64$
Dynamic	259	45.9		2.43	

Results by gender: no gender difference in Risk Aversion

We find NO gender difference in the Baseline (both dynamic and static).

		N	BRET Mean	BRET Mann-Whitney z	SOEP Mean	SOEP Mann-Whitney z
Static	Males	30	44.23	Prob $> \|z\| = 0.66$	4.63	Prob $> \|z\| = 0.97$
	Females	54	43.44		4.65	
Dynamic	Males	105	46.38	Prob $> \|z\| = 0.66$	5.33	Prob $> \|z\| = 0.04$
	Females	164	46.65		4.83	

Table: Stopping time in the baseline treatments, breakdown by gender

- Gender effect often found in risk aversion elicitation (Eckel and Grossman, Charness and Gneezy)
- Always treated as 'a fact', but many studies *do not* find it
- Turns out to be task-specific
- Why is it not there in the BRET?

BRET: loss aversion

No gender gap: the role of loss aversion

There is evidence that gender gap could be due to *loss aversion*
(Booji and Van de Kuilen, 2009)

- Questionnaire study: $N = 1935$, representative panel of Dutch population
- Task: Wakker and Deneffe
- Prospect Theory framework

$$V(x) = \begin{cases} x^r & x > 0 \\ -\lambda(-x)^r & x < 0 \end{cases}$$

- Females have same risk aversion parameter r...
- ...but significantly different loss aversion λ

Differences in *lambda* for a Prospect Theory Value Function

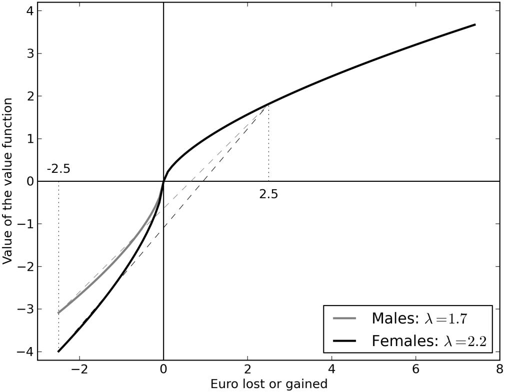


Figure: Two prospect Theory value functions, with the same r, but different λ

Framing the BRET for losses

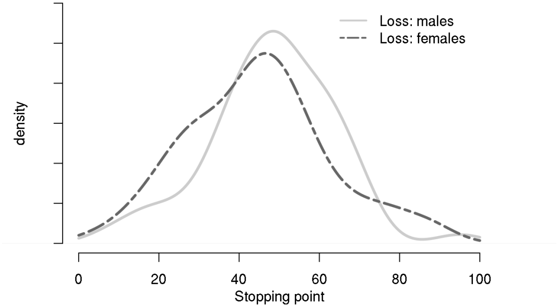
We ran a specific *Loss Aversion* treatment - changing just the **frame**

- Subjects on arrival find 2.5e on their desk
- These are on top of the 2.5e show-up fee
- The 2.5 are at stake in a framed BRET
- in which information is conveyed as losses/gains around a reference point

e.g.: at $k = 16 \rightarrow$ 'you are losing 0.9e w.r.t. the starting endowment'
e.g.: at $k = 37 \rightarrow$ 'you are gaining 1.2e w.r.t. the starting endowment'

Results should not change, except if subjects (especially females) *loss averse*.

Loss Aversion: results by gender



		BRET		SOEP	
	N	Mean	Mann-Whitney z	Mean	Mann-Whitney z
a. full sample					
Males	62	48.26	Prob $> \|z\| = 0.129$	5.69	Prob $> \|z\| < 0.001$
Females	73	44.67		4.62	
b. Sensitivity analysis without 2.5% *of obs. in each tail of the distribution in the whole experiment*					
Males	60	48.16	Prob $> \|z\| = 0.057$	5.61	Prob $> \|z\| = 0.004$
Females	69	43.43		4.56	

BRET: other controls

Experimental details, II: treatments

We ran several treatments to check robustness and validate the task

	Treatment	N	% Outliers	% Extreme
Baseline	Dynamic	269	0.37	3.72
Baseline	Static	84	2.30	7.14
Loss Aversion	Inducing a reference point at 2.5€	135	4.17	4.44
Explosion	Bombs explode upon collect	122	0	0.82
High Stakes	Box value: 0.2 €	87	1.14	3.45
Size	**Big**: 20×20; Deletion time: 0.25 seconds	32	0	6.25
	Small: 5×5; Deletion time: 4 seconds	92	0	0
	Mixed: 10×10; Outcomes updated every 4 sec.	55	0	7.27
Fast	Deletion time: 0.5 sec	92	0	2.17
Random	Collection sequence: random	32	0	6.25
Wealth effects	Task performed after another experiment	93	0	1.08
Repeated	Unannounced repetition of the task for 5 times	(30)	0.66	5.33
Total		1093		

Explosion version

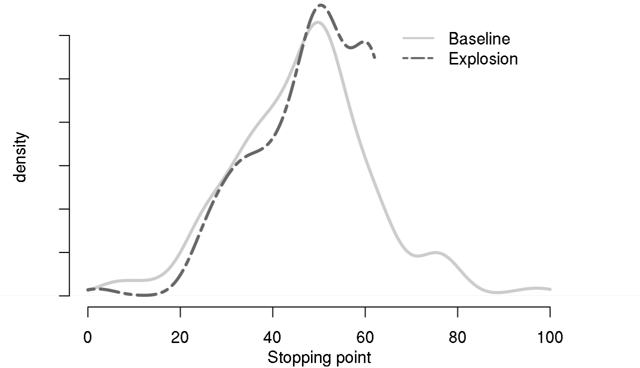
Does the delayed explosion make risk less salient?

- 'Explosion' in the *Baseline* is not live - to avoid truncation
- Subject could perceive the situation less risky than it is
- And hence show lower risk aversion
- We run a treatment with **Live explosion**
- $\rightarrow$ data truncated; earnings are real.
- The exploding box is randomly predetermined at $k = 62$

Results

- There is no significant difference between the two treatments...
- ...possibly due to truncation (at $k = 62$, 20.5% still in *vs.* 11.9% in Baseline)
- ...but subjects tend to risk slightly more in *Explosion*.
- The absence of live explosion *does not* induce more risk taking.

Explosion: results



	N	Mean	Std. Dev.	Mann-Whitney z
Baseline	269	46.546	15.138	Prob $> \|z\| = 0.121$
Explosion	122	47.95	11.87	

Figure: Results and kernel density of decision by treatment, Baseline *vs.* Explosion

High Stakes version

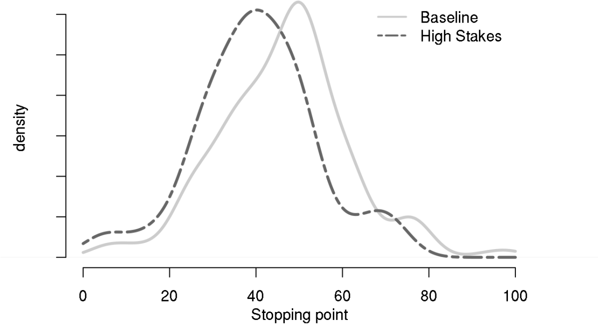
Does increasing the stakes change the results?

- Stakes are very low in the Baseline to provide a benchmark for controls
- Many papers show that increasing stakes changes risk preferences (Holt Laury 2002)
- We run a treatment with **doubled** stakes.

Results

- There is a significant difference: with high stakes, **higher risk aversion**
- $k^*_{high} = 40.1$, while $k^*_{base} = 46.5$
- but still no gender difference
- All as expected.

High Stakes *vs.* Baseline



	N	Mean	Std. Dev.	Mann-Whitney z
Baseline	269	46.546	15.138	Prob $> \lvert z \rvert = 0.0002$
High Stakes	87	40.057	13.980	

Figure: Results and kernel density of decision by treatment, Baseline *vs.* High Stake

Fast version

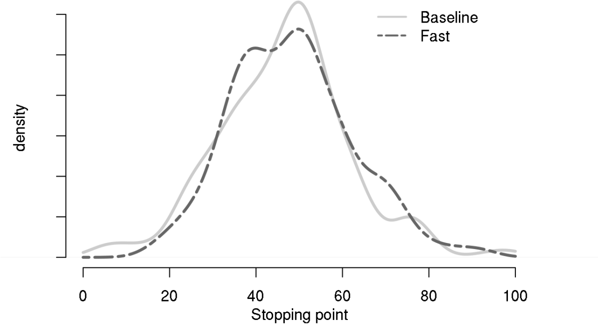
Does increasing the speed of the deletion process change the reuslts?

- The baseline implements a very slow collection of boxes (1 per second)
- Patience, nerves... could make a difference
- We run a treatment with **doubled** speed - a box is deleted every 0.5 secs
- Total time to go through the test → 50 secs
- If patience or nerves played a role we should see different results

Results

- There is NO significant difference w.r.t.Baseline
- Average slightly higher, but not significant.

Speed of deletion: Fast *vs.* Baseline



	N	Mean	Std. Dev.	Mann-Whitney z
Baseline	269	46.546	15.138	Prob $> \lvert z \rvert = 0.483$
Fast	92	48.315	13.985	

Figure: Results and kernel density of decision by treatment, Baseline *vs.* Fast

Changing the size

Does changing the size of the field of boxes change results?

- The baseline implements a very intuitive representation of probabilities
- Choosing to display 100 boxes.
- What happens if we reduce-increase the number of boxes?
- Theoretically, should make no difference up to rounding problems

We implement 3 treatments:

- 5×5 with 25 boxes, each collected every 4 seconds and worth 4 tokens
- 20×20 with 400 boxes, collected every $\frac{1}{4}$ sec and worth 0.25 tokens
- Mixed5×5: visually as baseline, but collected as in 5×5

Changing the size: interface

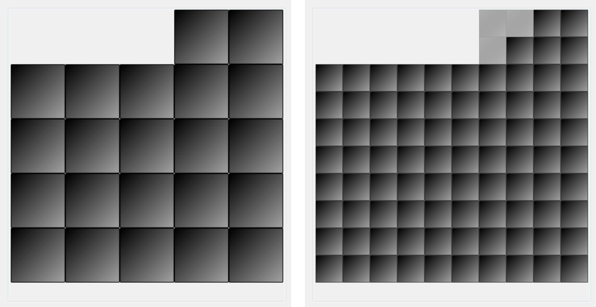
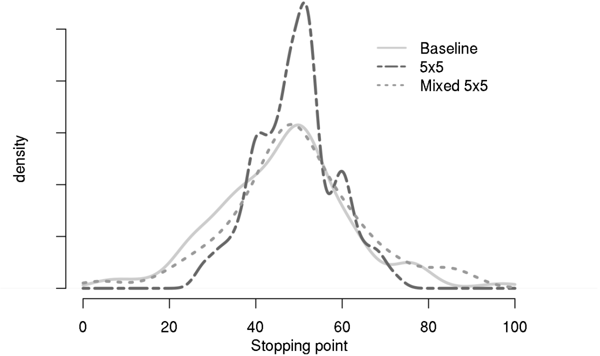


Figure: Screenshot after 15 seconds in the 5×5 (left) and Mixed 5×5 (right)

Changing the size: results



	N	Mean	Std. Dev.	Mann-Whitney z
Baseline	269	46.55	15.14	
20 × 20	32	46.29	21.29	Prob $> \lvert z \rvert = 0.879$
5 × 5	92	49.13	9.11	Prob $> \lvert z \rvert = 0.067$
Mixed 5 × 5	55	49.40	15.63	Prob $> \lvert z \rvert = 0.249$

Figure: Results and kernel density of decision by treatment, Baseline *vs.* Size

Conclusion

Conclusion

We developed a new Risk Elicitation Task in continuous time

- It's more *precise* and *parsimonious*
- It's *intuitively* grasped by subjects (easiest of the lot)
- It's *complete* as it measures both risk loving and risk aversion

Moreover...

- It features *no gender gap* in risk aversion...
- ...and brings evidence that the gender gap might be due to *loss aversion*
- It is robust to several specifications
- It is easy extendable to test other issues (ambiguity, illusion of control)

Thanks!