J Risk Uncertain (2013) 47:31–65
DOI 10.1007/s11166-013-9170-z

The "bomb" risk elicitation task

Paolo Crosetto · Antonio Filippin

Published online: 24 July 2013

Abstract This paper presents the Bomb Risk Elicitation Task (BRET), an intuitive procedure aimed at measuring risk attitudes. Subjects decide how many boxes to collect out of 100, one of which contains a bomb. Earnings increase linearly with the number of boxes accumulated but are zero if the bomb is also collected. The BRET requires minimal numeracy skills, avoids truncation of the data, allows the precise estimation of both risk aversion and risk seeking, and is not affected by the degree of loss aversion or by violations of the Reduction Axiom. We validate the BRET, test its robustness in a large-scale experiment, and compare it to three popular risk elicitation tasks. Choices react significantly only to increased stakes, and are sensible to wealth effects. Our experiment rationalizes the gender gap that often characterizes choices under uncertainty by means of a higher loss rather than risk aversion.

Keywords Risk aversion · Loss aversion · Elicitation method

JEL Classifications C81 · C91 · D81

Uncertainty is a recurrent feature of the decision-making process in several domains such as investment, insurance, education, tax compliance, and labor market choices.

P. Crosetto (✉)
Max Planck Institute of Economics, Kahlaische Straße 10, 07745 Jena, Germany
e-mail: crosetto@econ.mpg.de

A. Filippin
Department of Economics, University of Milan, Via Conservatorio 7, 20122 Milano, Italy
e-mail: antonio.filippin@unimi.it

A. Filippin
Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany

Agents react to uncertainty according to their individual preferences, and measuring their risk attitudes is important both to isolate possible regularities and to control for idiosyncratic characteristics that could otherwise act as a confounding factor when other decision tasks are analyzed.

Not surprisingly, risk attitude has received a great deal of attention on the theoretical level as well as in the laboratory. However, a consensus view is not apparent in either case. From a theoretical point of view, the mainstream paradigm (Expected Utility, von Neumann and Morgenstern 1944) has been criticized for its unsatisfactory predictive power, and other constructs have been proposed, the most famous being Prospect Theory (Kahneman and Tversky 1979). Similarly, many tasks have been developed to elicit risk attitudes by means of incentivized choices. Some of the most popular ones can be found in Holt and Laury (2005), Eckel and Grossman (2002), Gneezy and Potters (1997), Wakker and Deneffe (1996), and Becker et al. (1964). Together with others not quoted here, they all have in common that they either propose a choice between prospects or elicit the certainty equivalent of some lotteries in such a way that the coefficient of risk aversion can be estimated once a parametric form of the utility function is assumed. The coefficient of risk aversion can also be retrieved through auction bids (Cox et al. 1982). A method that shares with this paper the visual representation of probabilities is the Balloon Task (Lejuez et al. 2002). Sometimes risk attitudes are self-reported, answering simple questions such as those asked in the German Socio-Economic Panel or in the Italian Survey of Household Income and Wealth, or validated questionnaires such as the Domain-Specific Risk-Taking (DOSPERT, Blais and Weber 2006). However, as Garcia-Gallego et al. (2012) point out, in the economics literature insufficient attention is often paid to the validation of risk elicitation tasks.

In this paper, we present the Bomb Risk Elicitation Task (BRET), a choice-based elicitation method characterized by several useful features. The BRET asks subjects to decide at which point to stop collecting a series of 100 boxes, one of which contains a time bomb. Earnings increase linearly with the number of boxes collected but are equal to zero if one of them contains the bomb. The task is designed to avoid potential truncation of the data, so that subjects are free to choose any number between 0 and 100.[1]

The task proposes to the subjects 101 lotteries fully described in terms of outcomes and probabilities by a single parameter, the number of boxes collected. We propose two versions of the BRET. The first is a choice in a static framework that can be performed even with paper and pencil. It can be easily administered in any research in the field as well as included in questionnaires and surveys. Relying on a visual representation in continuous time, the dynamic version provides a more intuitive and transparent illustration of probabilities and outcomes that makes the task even easier to understand. It requires an electronic support and it is therefore more suitable for laboratory experiments. Both versions require low numeracy skills. The BRET allows

[1]The task is similar to, but not inspired by, the unpublished Chip Draw task proposed by Eckel et al. (2003). We were pointed to the Chip Draw task by its authors only after this paper had circulated.

the precise estimation of both risk aversion and risk seeking, generating a virtually continuous distribution of outcomes. Therefore, it is well positioned in the trade-off, as explored by Dave et al. (2010), between precision and understandability since it features both a fine-grained measure of risk preferences and a simple and intuitive design.

The BRET also has several theoretical advantages. First, it does not suffer from loss aversion as a potential confounding factor because it is entirely defined in the gain domain and, in contrast with other well-known tasks in the literature, does not even provide endogenous reference points against which some outcomes could be perceived as losses. Second, the possibility of clearly distinguishing risk aversion from loss aversion also allows us to shed some light on gender differences in risk attitudes, showing that females are not characterized by a different *risk* aversion but a stronger *loss* aversion instead. Third, the BRET entails a unique choice, thereby avoiding the possibility that results can be biased by violations of the Reduction Axiom that are commonly observed (see Kaivanto and Kroll 2011 and references therein).

Choices in the task turn out to be sensitive only to wealth effects and to the amount of money at stake. The wide range of choices available to subjects poses a drawback increasing the sensitivity of results to decisions far in the tails of the distribution. We will therefore perform a sensitivity analysis, deleting the extreme 2.5% of the observations in each tail. In a few cases, this slightly affects the results.

We also compare the BRET with three other popular tasks: the Multiple Price List *à la* Holt and Laury (2002), the Ordered Lottery Choice of Eckel and Grossman (2002) and the Investment Game introduced by Gneezy and Potters (1997).

The outline of the paper is as follows. In Section 1, we summarize the Bomb Risk Elicitation Task along with the underlying theoretical framework. The experimental procedure is detailed in Section 2. The evidence obtained from the baseline versions as well as the validation and robustness treatments is presented in Section 3. In Section 4 we report the results of the comparison treatments. Section 5 concludes.

1 The bomb risk elicitation task

1.1 Static version

In the basic version of the task, subjects face a 10×10 square in which each cell represents a box. They are told that 99 boxes are empty, while one contains a time bomb programmed to explode at the end of the task, i.e., *after* choices have been made. Subjects are asked to choose a number $k^* \in [0, 100]$ that corresponds to the number of boxes they want to collect, starting from the upper left corner of the square. The position of the time bomb ($b \in [1, 100]$) is determined after the choice is made by drawing a number from 1 to 100 from an urn. If $k_i^* \geq b$, it means that subject i collected the bomb, which by exploding wipes out the subject's earnings. In contrast, if $k_i^* < b$, subject i leaves the minefield without the bomb and receives γ euro cents for every box collected.[2]

Subjects' decisions can be formalized as the choice of their favorite among the lotteries

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

summarizing the trade-off between the amount of money that can be earned and the likelihood of obtaining it. Note that the task amounts to choosing the preferred among 101 lotteries, fully described both in terms of probabilities and outcomes by a single parameter $k \in [0, 100]$, while $\gamma > 0$ is a scale factor.

The expected value of these lotteries is equal to $\gamma(k - 0.01k^2)$, a bow-shaped function with a maximum at $k = 50$ and trivially equal to zero for $k = 0$ and $k = 100$. Normalizing $u(0) = 0$, an expected utility maximizer should choose:

$$k^* : \qquad \frac{u(k)}{u'(k)} = 100 - k. \tag{1}$$

Assuming the classic (CRRA) power utility function $u(x) = x^r$:

$$k^* = 100\frac{r}{1 + r'} \tag{2}$$

which implies that a risk neutral subject should choose $k^* = 50$. The implied levels of r for every possible choice k can be found in Appendix A: our task allows estimation of 100 intervals for $r \in [0, 68.275]$.

This basic version of the BRET is simple, and it can be run with paper and pencil. However, it requires a certain level of abstraction that, in principle, could hamper its effective comprehension. In order to facilitate subjects' understanding we elaborated an equivalent version represented as a visual task in (almost) continuous time.

1.2 Dynamic version

The dynamic visual version in continuous time of the BRET is represented on the PC screen as a square formed by 10×10 cells, each one representing a box. Below the square is a "Start" and a "Stop" button. From the moment the subject presses "Start" one cell is automatically deleted from the screen at each second, representing a box that is collected. The deletion process follows a predetermined (arabic) sequence – i.e., the first element of the first row is deleted first, the second element of the first row second, and so on. A screenshot of the visual version after 45 seconds (i.e., after 45 boxes have been collected) as shown to the subjects is reported in Fig. 1. The subject

[2]This mechanism appears to have much in common with the Becker-DeGroot-Marshak (BDM) procedure (Becker et al. 1964), but this is not the case. The BDM mechanism induces subjects to truthfully reveal the reservation price for an item. It has been used as a risk elicitation task by eliciting the willingness to pay for, or the willingness to accept, a lottery ticket (see Grether and Plott (1979); Harrison (1990) for examples of its use and Karni and Safra (1987) for a critical assessment of its incentive compatibility). A random device is then used to decide whether the transaction takes place or the lottery is actually played. Instead, the BRET amounts to a choice between different lotteries and the preferred lottery is always played.

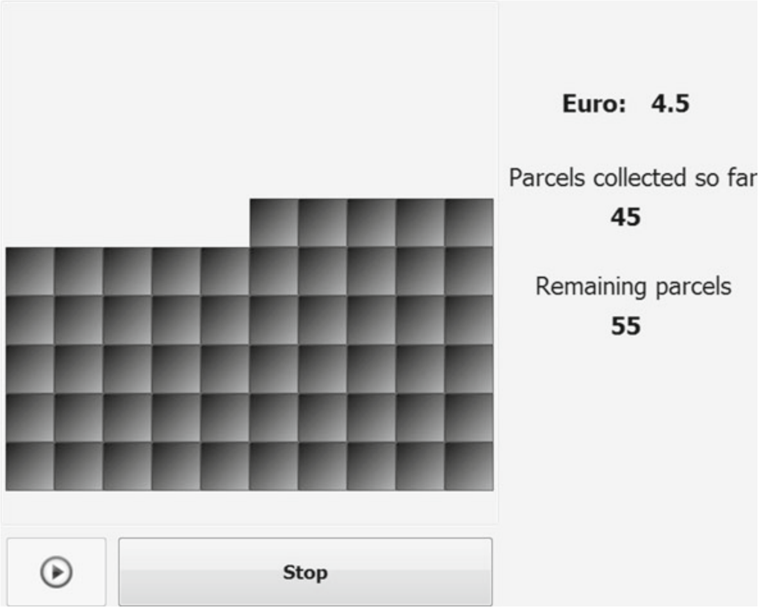


Fig. 1 The BRET interface after 45 seconds - dynamic version

is informed about the number of boxes collected at any point in time. Each time a box is collected, the subject's provisional account is credited with γ additional euro cents. The subject can, at any time, stop the drawing process by hitting the "Stop" button.

Subjects are not informed about the content of the boxes ("empty" or "bomb") since the position $b \in [1, 100]$ of the time bomb, and therefore whether it has been collected or not, is randomly determined only at the end of the experiment. The metaphor of the time bomb has the crucial merit of avoiding the truncation of the data that would otherwise happen in case of a real-time notification. Subjects are explicitly warned that the earnings they see during the experiment are provisional since they would be equal to zero in case the time bomb was in one of the boxes collected.

From a theoretical point of view, the decision can be represented as a sequence of binary choices, governed, as in the static case, by the parameter k. After k boxes have been collected, the choice subjects face is:

$$L^k = \begin{cases} 0 & \frac{k}{100} \\ \gamma k & \frac{100-k}{100} \end{cases} \quad \text{vs.} \quad L^{k+1} = \begin{cases} 0 & \frac{k+1}{100} \\ \gamma (k+1) & \frac{100-k-1}{100} \end{cases} .$$

The solution is equivalent to the static version, provided the subject is characterized by well-behaved preferences.[3] In fact, in both cases the subject has no way of determining the bomb's position during the task and faces the same opportunity set.

[3]A sufficient condition ensuring an identical solution is that the expected utility function is not characterized by multiple and separate local maxima.

The dynamic visual version of the BRET has several advantages. First, it presents the set of lotteries in a sequential manner, inducing subjects to focus on only two lotteries at any moment in time. Such a framework makes evident that the lotteries can be ranked from the safest but less rewarding to the riskiest and more rewarding. Second, the visual representation and information provided allow a clear description of the probabilities involved. These features make the visual version intuitive and understandable even for subjects with low mathematical skills, ensuring that the likelihood of the subject's decision being driven by confusion or an imperfect comprehension of the final outcomes and their probabilities is reduced to a minimum. Moreover, the dynamic version of the task is particularly appropriate to measure risk attitudes in decisions entailing a time dimension, such as trading, Dutch auctions, etc.

The BRET is well positioned from several points of view along which risk elicitation tasks can be evaluated. The task allows precise estimation of the coefficient of risk aversion both in the risk aversion and the risk loving domains. Moreover, the task can provide estimates of the coefficient of risk aversion that are not biased by the degree of loss aversion. This is possible because subjects cannot be assumed to form any reference point, as their choice set does not include any option that offers a positive amount of money with probability 1; at most, subjects can assure themselves a payoff of 0 by choosing either $k = 0$ or $k = 100$.

Additionally, the task requires subjects to make a single decision, and it is therefore robust to possible violations of the Reduction Axiom that would instead affect the results in case of multiple choices, one of which paid at random (Bernasconi and Loomes (1992), Halevy (2007) and Harrison and Swarthout (2012)). Last but not least, the task is simple. From this point of view the two versions are not identical, because the dynamic BRET is less demanding on a cognitive level as well as better suited to facilitate subjects' comprehension. This reason, together with the fact that the dynamic version is characterized by a richer set of parameters that can be manipulated, makes the visual version in continuous time our preferred choice. We use it both to test the robustness of the task (Section 3 below) and to compare the BRET with other risk elicitation mechanisms (Section 4), although increasing the amount of money at stake in the latter case.

2 Experimental procedure

The experiment was run between March and November 2012 at the laboratory of the Max Planck Institute of Economics in Jena, Germany. A total of 1559 subjects participated, distributed over 52 sessions, lasting at most half an hour. The sample includes mainly, but not only,[4] students from the Friedrich Schiller University Jena, Germany.

We chose to make each box worth 10 euro cents ($\gamma = 0.1$). The resulting average earning was 5 euro, with a minimum corresponding to the show-up fee of 2.5 euro and a maximum of 13.5 euro in a session with high stakes. We kept the average

[4] Even though recruitment for experiments takes place mainly on campus, the subject pool also includes some non-student workers and adults from Jena.

earnings low (the maximum expected value of the task was 2.5 euro) for two main reasons. First, one of the goals of this paper is to provide a tool that can serve as a control for risk attitudes. Using small stakes ensures that the results illustrated in this paper can serve as a useful benchmark for this purpose. Second, we chose to keep the expected earnings around the student reservation wage, given the large number of subjects involved and the short time span required. To control for the effects of increasing the stakes, we ran a specific *High Stakes* treatment ($\gamma = 0.2$, Section 3.5).

The experiment was computerized. Both the experimental software and the script we used to collect and organize the raw data were programmed in Python (van Rossum 1995).[5]

Upon entering the lab, subjects were randomly assigned to a computer. Instructions were then displayed on the screen and read aloud. After clarifying questions were individually addressed, subjects performed one trial period in order to experience the visual representation of the task. At the end of the trial period, however, there was no draw of the bomb's position in order not to provide the subjects with any reference point. To check for the possible effects of a trial period, a specific *No Trial* treatment was run. The paying task was then played one-shot.

At the end of the task, subjects filled in a questionnaire containing two demographic questions (age and gender) plus:[6]

1. A self-reported measure of risk attitudes, the general risk question used in the German Socio-Economic Panel (SOEP, see Wagner et al. 2007) on a $0 - 10$ scale: "How do you see yourself: are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" The validity of this self-reported question to elicit risk attitudes as compared to the results of incentivized lottery-based tasks has been explored by Dohmen et al. (2011), who find that self-reported answers can represent a valid low-cost substitute for incentivized lottery schemes, although the fraction of variance explained is quite low (about 6%).
2. A question to directly measure the perceived complexity of the task on a $0 - 10$ scale. Perceptions might not agree with actual choices, though, as subjects might deem the task easy and then commit trivial errors or dominated choices. Therefore we also rely on objective measures, derived from actual choices, to evaluate the difficulty of the treatments.

3 Results

The task was generally well understood, and subjects raised few questions. In some few cases, subjects reported technical or understanding difficulties or asked for help

[5] A z-Tree version of the BRET, though not used in the sessions, was also developed and tested to ensure the widest portability of the task. The experimental software of the BRET and its source code as well as the z-Tree version are available in the online supplementary material to be found at http://goo.gl/3eogr.

[6] We also administered the DOSPERT risk questionnaire that weights several different domains in which risk attitudes can play a role (gambling and investment among the others), but we do not report results because it shows a correlation with the incentivized choices lower than the SOEP.

Table 1 Demographics of the experimental sample, excluding inconsistent choices

	Age bracket				N
	18 − 22	23 − 27	28 − 32	33 − 60	
Male	234	307	77	26	644
Female	422	387	56	15	880
Total	656	694	133	41	1524

in carrying out the task. For instance, five subjects declared immediately that they stopped the task within the first seconds by mistake; they were given the opportunity to start anew. In other cases, subjects did not report any problems, but nevertheless submitted dominated choices, e.g., never stopping collecting, thereby gaining zero with probability one.[7] For all these reasons we decided to label as inconsistent the 20 subjects who stopped at $k = 0$, $k = 1$, and $k = 100$ in the BRET sessions, as well as the 15 who made inconsistent choices in the Holt and Laury task, and eliminate them from the dataset, leaving a total of 1524 observations, whose descriptive statistics are summarized in Table 1.

We adopted a between-subject design in the whole experiment. Of the 1524 subjects of the resulting sample, 353 participated in the baseline treatments (Static and Dynamic) described in Section 1, while 924 took part in the additional robustness treatments that we ran to explore several possible departures from the baseline version of the task, and 247 were involved in the across-task comparison sessions. The breakdown by treatment of our experiment, including the robustness checks described below in more detail, is summarized in Table 2.

We first ran *Baseline* sessions, testing both the static and dynamic tasks. To check the robustness of the task we ran several treatments. First, we analyze a version of the task framed in terms of losses and gains around a focal starting endowment. We then test whether the metaphor we use to avoid truncation of the data, i.e., the time bomb, affects the results. We do so by implementing a version in which the bomb explodes immediately when collected. We then run robustness checks by manipulating the parameters that characterize the baseline version (the absence of a trial period, the sum at stake, the size of the field of the boxes, the speed of the deletion process and the order in which the boxes are deleted). Moreover, we analyze the sensitivity of the task to repetition and wealth effects.

Finally, we compared the BRET with other three popular tasks, namely the multiple price list *à la* Holt and Laury (2002), the ordered lottery choice task of Eckel and Grossman (2002) and the investment game introduced by Gneezy and Potters (1997).

Table 2 also shows that the number of inconsistent choices, though not high on average, is not uniformly distributed across treatments and can therefore be used as an objective proxy of the complexity of the treatments. Similar insights about the

[7] On average, the subjects who submitted dominated choices reported a higher perceived difficulty of the task compared to the rest of the sample (0.57 vs. 0.42). The difference is however not significant (Mann-Whitney, p-value 0.16).

Table 2 Summary of the treatments (inconsistent choices excluded in N)

	Treatment	N	% Inconsistent	% Extreme
Baseline treatments				
Baseline	Dynamic	269	0.74	4.43
Baseline	Static	84	3.54	10.34
Robustness treatments				
Loss aversion	Inducing a reference point at 2.5€	135	6.25	10.42
Explosion	Bombs explode upon collection	122	0.81	1.63
No trial	No trial period	123	2.38	7.14
High stakes	Box value: 0.2€	87	1.14	4.55
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	Collecting sequence: random	32	0	6.25
Wealth effects	BRET performed after another task	93	1.06	2.13
Repeated	Unannounced repetition of the task for 5 times	61	0	0
Comparison treatments				
Holt-Laury	Multiple price list task	73	17.04*	
Eckel-Grossman	Ordered lottery selection task	88	0	
Gneezy-Potters	Risky investment task	86	0	
Total		1524		

*Inconsistent choices - i.e. multiple switching points along the multiple price list.

difficulty of the different treatments can be derived in the last column of Table 2. It reports the fraction of observations lost in each treatment of the BRET if we restrict the sample, dropping from the pooled sample of 1297 observations (including the inconsistent subjects) the 2.5 % most extreme choices in each tail. Given the fact that inconsistent and extreme choices are more frequent in some sessions, we ran a sensitivity analysis comparing results from the whole and the restricted samples. This sensitivity analysis is suggested by the fact that the wide range of choices available by construction assigns a strong role to extreme choices. Moreover, decisions far in the tails of the distribution sometimes display an inversion of the otherwise negative correlation that characterizes the density and distance from the mode, which also signals a possible anomaly.

The SOEP question and the self-reported complexity index are ordinal scales, and their use in computing averages and correlations requires giving them a cardinal interpretation. To circumvent this problem, both scales were transformed into dummy variables. In order to create the dummy variable, we chose for the SOEP a cut-off point such that the percentage of risk averse and risk loving subjects be consistent with the average choices in the BRET. This implied cutting the SOEP at 6: all observations below this threshold were given the value of 0, and all the ones strictly above the value of 1. For complexity, lacking a precise benchmark, we computed the median of the observed answers, which turned out to be 2, and then we assigned all observations lower or equal to the median a value of zero, the others a value of one.

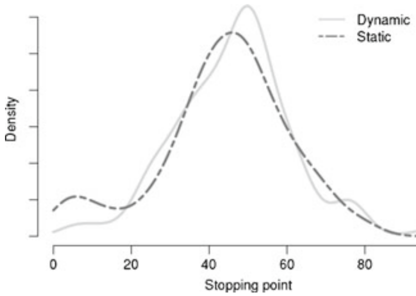
3.1 Static and dynamic version of the BRET

Results of the Static and Dynamic treatments are summarized in Fig. 2. As reported in the figure and table, the average number of boxes collected slightly differs across treatments, although the difference is not significant according to a Mann-Whitney test ($Prob > |z| = 0.34$). Subjects also report a similar perceived complexity ($Prob > |z| = 0.79$).

Looking at the distribution of choices it is immediately evident, however, that in the Static version, there is a cluster of subjects who display a suspiciously high level of risk aversion. This may point to an imperfect comprehension of the task, as captured by the higher number of inconsistent choices present in the Static sessions (see Table 2). The self-reported degree of complexity of these inconsistent subjects is also higher (0.67 vs 0.38), though the difference is only marginally significant ($Prob > |z| = 0.105$). The sensitivity analysis performed without the 5% of extreme choices leads to identical results in the two treatments (see Fig. 2).

Given the similarity of results and the emergence of comprehension problems in the Static treatment, we concluded that the visual, Dynamic task is more likely to produce reliable outcomes, especially when it is administered to subjects characterized by low numeracy skills. We hence used the Dynamic treatment as our baseline in running all the other treatments.

Assuming a CRRA utility function, Eq. 2 allows us to straightforwardly compute the coefficient r of risk aversion, which, according to the results in the Dynamic baseline, turns out to be on average $r \cong 0.85$. The cumulative distribution of choices shows 52.1% of risk averse subjects ($k \leq 49$), 14.7% risk neutral subjects ($k = 50$),



Treatment	N	Mean	Mann-Whitney z	Complexity	Mann-Whitney z
a. full sample					
Static	84	43.7	$Prob > \|z\| = 0.34$	0.376	$Prob > \|z\| = 0.798$
Dynamic	269	46.5		0.414	
b. sensitivity analysis without 2.5% *of obs. in each tail in the whole experiment*					
Static	78	45.8	$Prob > \|z\| = 0.83$	0.376	$Prob > \|z\| = 0.905$
Dynamic	259	46.2		0.384	

Fig. 2 Results and kernel density of decision by treatment, *static* vs. *dynamic*

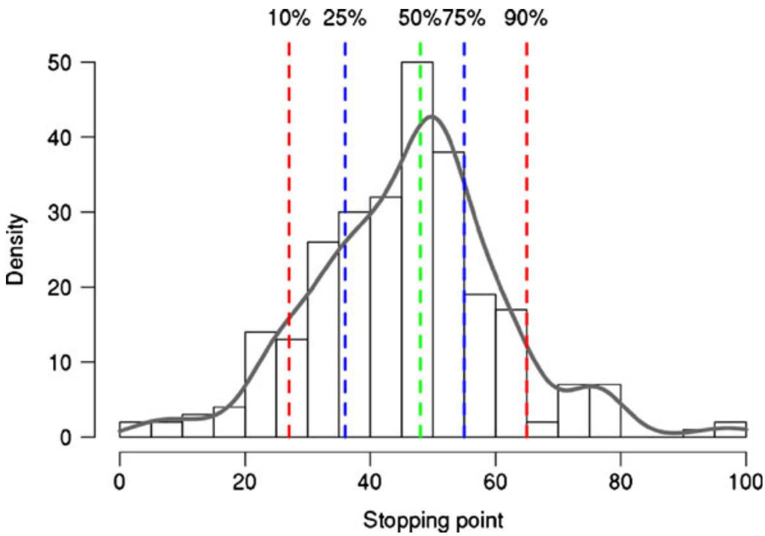


Fig. 3 Distribution of decisions in the dynamic version

and 33.2% risk seekers ($k \geq 51$). Splitting the distribution by quartiles shows that 50% of the observations lie between 36 and 55. The bulk of the subjects are either risk averse, risk neutral, or slightly risk loving (see Fig. 3).

A robust result of both the Static and Dynamic treatments is that no gender difference in risk aversion can be observed (Table 3).[8] This finding is at odds with other results in the literature showing that females display more conservative behavior in decisions under uncertainty (see for instance the surveys by Charness and Gneezy (2012), Croson and Gneezy (2009) and Eckel and Grossman (2008b)). Moreover, in the Dynamic treatment, the answer to the non-incentivized SOEP question shows, within the same subsamples of males and females, a significant gender gap.

A closer look at the contributions in the literature shows that the gender gap in risk attitudes is a recurrent but not a systematic finding and that it is possible to identify some relevant underlying patterns. First, the gender gap is stronger when risk attitudes are measured in a contextual framework and, in particular, within simulated investment decisions (Eckel and Grossman 2008b). Second, some tasks seem to elicit gender differences more systematically than others. Limiting the attention to the tasks compared to the BRET in this paper, Charness and Gneezy (2012) show a strong and systematic gender gap: in their task, subjects are endowed with an amount of money and have to decide how much of it to invest in a risky option, with a 50% chance of losing the amount invested. On the other hand, with the task proposed by Holt and Laury (2002), in which subjects are asked to choose between batteries of lotteries that differ in terms of riskiness, the evidence on a gender gap in risk preferences is mixed.

[8]There is instead a gender difference in the perceived complexity of the tasks, with women finding the task significantly more difficult than men. This is a result that emerges in almost all the treatments.

Table 3 Stopping time in the baseline treatments, breakdown by gender

			BRET		SOEP		Complexity	
		N	Mean	Mann-Whitney z	Mean	Mann-Whitney z	Mean	Mann-Whitney z
Static	Males	30	44.23	Prob > $\lvert z \rvert = 0.66$	0.27	Prob > $\lvert z \rvert = 0.51$	0.19	Prob > $\lvert z \rvert = 0.002$
	Females	54	43.44		0.20		0.53	
Dynamic	Males	105	46.38	Prob > $\lvert z \rvert = 0.66$	0.36	Prob > $\lvert z \rvert = 0.02$	0.25	Prob > $\lvert z \rvert = 0.002$
	Females	164	46.65		0.23		0.46	

A possible reading of the mixed results from the literature on gender differences in risk preferences is that the presence or not of a gap might be correlated with the importance played by losses in the chosen task. In particular, subjects could endogenously form reference points and use them to evaluate lotteries, such that a gain-only choice would then be perceived as a loss-gain one. Formation of reference points could be highly likely in particular if an option that guarantees a positive amount with probability one is present in the choice set, even for tasks completely framed in the gain domain.

In Gneezy and Potters (1997), for instance, a highly significant gender gap is observed: in the task subjects choose between two options, one of which secures up to all of the initial endowment. In Holt and Laury (2002), on the contrary, losses are less salient, and no gender gap is observed. Such an interpretation could also account for the emergence of a gender gap in the Eckel and Grossman (2002) method, in which subjects are asked to choose one of five lotteries characterized by increasing expected value and riskiness. Although framed entirely in the gain domain, this task features a degenerate lottery with no uncertainty, in which subjects could secure a positive amount with probability one; subjects could adopt this as a reference point to compare losses that might be incurred by choosing the riskier prospects.[9] In order to test whether the absence of a gender gap in the BRET can be attributed to the fact that males and females have similar risk aversion, while being characterized by different loss aversion, we ran a framed Loss Aversion treatment. The results are detailed in the next subsection.

3.2 Loss aversion and the gender gap in risk attitudes

There is some evidence in the literature that loss aversion, rather than risk aversion, may account for the observed gender differences. In a study carried out in the Netherlands, exposing 1935 respondents from a household representative panel to hypothetical choices, Booij and de Kuilen (2009) find strong support for the idea that the gender gap is due to loss and not to risk aversion. The paper elicits a series of lottery outcomes that generate indifference relations between prospects in

[9] Indirect evidence for this interpretation can be derived from their finding that the gender gap in risk attitudes is virtually identical in a treatment supposed to be "gain only" and in a payoff equivalent treatment in which losses are explicitly considered (Eckel and Grossman 2008a).

order to estimate the parameters of a gender-specific Prospect Theory value function. When carrying out a parametric exercise to estimate two curvature parameters – one over gains only and one over losses only – females do not appear to have significantly different preferences. But when a loss aversion parameter is estimated by means of mixed gambles, a high and significant difference appears in terms of loss aversion.[10]

In this Section, we implement a version of the BRET that frames the incentive scheme explicitly including the possibility of incurring losses, in order to test whether it is possible to rationalize under Prospect Theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992) the puzzling absence of gender differences in the BRET described in the previous section.

Prospect Theory differentiates from Expected Utility Theory because within a rational decision-making process it pays greater attention to the way in which agents code and process information. It allows subjects to evaluate lotteries in terms of a reference point instead of evaluating them according to the implied final level of wealth. One of the implications of the theory is that it displays risk seeking in the loss domain, i.e., when the final outcome is worse than the reference point, and risk aversion in the gain domain. Crucially, Prospect Theory also features loss aversion, which captures the idea that the same outcome in absolute value is weighted more when it is a loss rather than a gain. The objective function summarizing these characteristics is called value function $V(x)$ and can be formalized in the following way:

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

where x is the change with respect to the reference point, r is the usual coefficient of risk aversion, and λ is the coefficient of loss aversion.[11] A graphical representation of the value function for reasonable levels of the parameters and centered at a reference level of 2.5 euro is displayed in Fig. 4. The BRET is defined entirely on the gain domain, with no possible reference point different than zero: similar average decisions of males and females imply that they are characterized by the same coefficient r. In contrast, when subjects face mixed gambles, the coefficient of loss aversion λ becomes crucial. If females are characterized by stronger loss aversion (higher λ), they could end up making different decisions even if they were characterized by the same r. Such a situation is depicted in Fig. 4, where females characterized by a steeper value function are more likely to turn down lotteries involving losses.

In order to test our hypothesis about gender in the BRET, i.e., that females and males have similar preferences for risk, while having different loss aversions, we framed our task in order to induce a reference point at 2.5 euro in what we called

[10] Evidence along this line is also provided by Brooks and Zank (2005); Schmidt and Traub (2002), who find a relatively larger number of women being classified as loss averse in experimental tasks with multiple individual choices involving mixed gambles. Brooks and Zank (2005) also find a larger fraction of women being risk averse. Gaechter et al. (2010) find no evidence of a gender gap in loss aversion.

[11] Here we omit the possibility of different curvatures in the two domains as well as issues of probability weighting, as this would go beyond the scope of this paper.

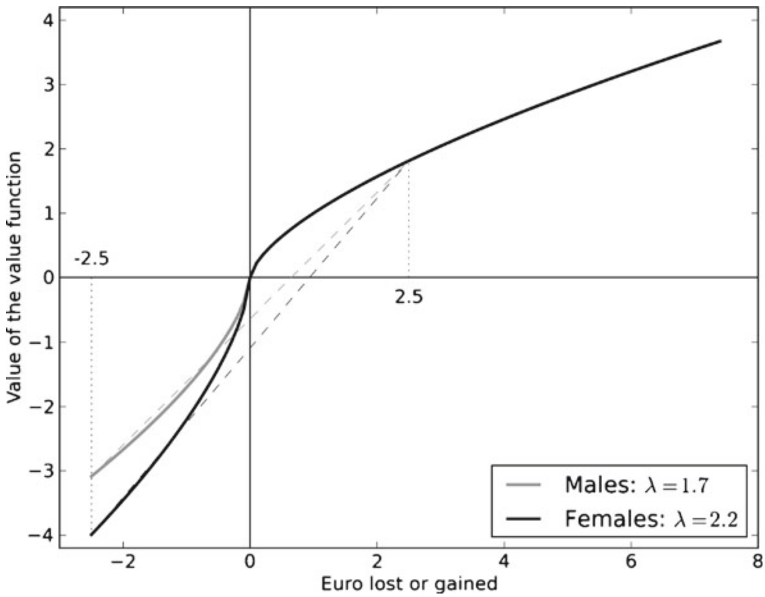


Fig. 4 Two Prospect Theory value functions, with the same r but different λ

the *Loss Aversion* treatment. This reference point was specifically chosen because it corresponds to the maximum expected value of the task and lies in the interval of values that, under Prospect Theory, generates for a common value of r sufficiently different predictions for two different λ.

After entering the lab, subjects found 2.5 euro on their desk. They were told that this amount had been awarded to them on top of the usual show-up fee, but that it would be put at stake in the experiment that was about to take place. In particular, by starting the task they would lose the 2.5 euro entirely. Subsequently, collecting boxes from 1 to 24, their loss with respect to their initial endowment would gradually reduce at the usual rate of 0.1 euro per box, reaching zero upon collecting the 25^{th} box. From that point, subjects entered the domain of gains that were again communicated with respect to the reference point. For instance, after collecting the 37^{th} box, subjects were told they were facing gains of 1.2 euro on top of their initial 2.5 euro endowment rather than total gains of 3.7 euro. Apart from the presence of a reference point, the task was identical to the Baseline in all details.

We can expect the initial endowment to induce an effect only insofar as subjects perceive this amount as theirs and use it as a reference point. The fact that the endowment was automatically put at stake, and subjects had no way of keeping the 2.5 euro with certainty, is likely to weaken the salience of the reference point in this treatment. Performing a real effort task prior to the awarding of the 2.5 euro, as done by Eckel and Grossman (2008a), would probably have strengthened the results, but we wanted to reduce to a minimum the modifications compared to the Baseline treatment. If instead the subjects perceive the 2.5 euro as windfall gains the anchoring to such a reference point is even less likely to emerge. Therefore we relied on a simple change

Table 4 Mean stopping time and complexity by treatment, baseline vs. loss

Treatment	N	Mean	Mann-Whitney z	Complexity	Mann-Whitney z
Baseline	269	46.54	$Prob > \|z\| = 0.86$	0.376	$Prob > \|z\| = 0.009$
Loss	135	46.31		0.518	

of frame, that we tried to make more salient by putting the money on the subject's desk.[12]

Results show that the different framing does not have any effect on aggregate choices (see Table 4). In contrast, it has a sizable effect on the perceived complexity of the task. Subjects find the mixed gamble environment significantly more difficult to understand. Such a perception is also confirmed by the high fraction of extreme choices that characterizes this treatment (see Table 2).

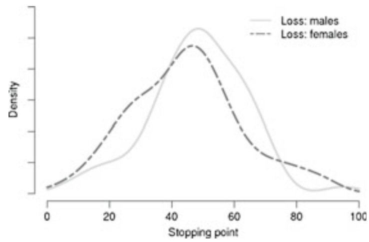
Breaking down the choices by gender (Fig. 5) confirms our hypothesis. The average choice of females turns out to be sizably lower than that of males, with the difference approaching traditional levels of significance, though not reaching them because of some outliers.

Note that, at first glance, these two subsamples are more different than those in the Baseline according to the answers to the SOEP question. However, the difference in the choices cannot be attributed to such an artifact. In fact, one of the sessions is characterized by puzzling results, with an even wider gap in the SOEP (0.58 for males vs. 0.16 for females) together with a reversed gender gap ($\hat{k}^*_m = 47.3, \hat{k}^*_f = 50.9$). Removing the data of this session would make the SOEP results virtually identical to those in Table 3, while at the same time improving the significance of the different choices of k^* at 0.057. Similarly, significance is also achieved by performing the sensitivity analysis (Fig. 5). As we kept manipulations at a minimum, we can say that results are not conclusive but definitely point in the right direction.

Finally, women significantly find the task more complex than men, but the difference between genders (0.6 vs. 0.41) is not higher than the gender difference in complexity of the Baseline treatment (0.46 vs. 0.25); the increase in perceived complexity is the same for men and women.

Focusing on the comparison by gender across treatments, it can be noted that one reason for the presence of a gender gap in this treatment may also be that males stop later when facing mixed gambles. Such a finding can be rationalized by a more pronounced curvature of the value function in the loss domain. Evidence in the literature shows that this is likely to be the case and that, at the same time, females are characterized by an even higher curvature (Booij and de Kuilen 2009; Schubert et al. 1999). Hence the gender gap should not be accounted for by a different curvature in the loss domain but rather by a different loss aversion parameter. This treatment alone, however, does not allow identification of all the parameters of the value function in the loss domain. As the BRET can be easily framed to be defined completely in the loss

[12] A similar procedure to induce a reference point was used, among others, by Harbaugh et al. (2010), who gave real dollar bills to the participants.



	N	Loss treatment Mean	Loss treatment Mann-Whitney	SOEP Mean	SOEP Mann-Whitney	Complexity Mean	Complexity Mann-Whitney
a. full sample							
Males	62	48.26	Prob > |z| = 0.129	0.47	Prob > |z| = 0.003	0.419	Prob > |z| = 0.002
Females	73	44.67		0.22		0.602	
b. Sensitivity analysis without 2.5% *of obs. in each tail in the whole experiment*							
Males	60	48.16	Prob > |z| = 0.057	0.45	Prob > |z| = 0.003	0.4	Prob > |z| = 0.018
Females	69	43.43		0.20		0.608	

Fig. 5 Stopping time, SOEP and complexity self-reported answers by gender: loss treatment

domain, it is straightforward to test if males and females have the same utility curvature in the loss domain. But this is beyond the scope of this paper, and is left for future research.

The last piece of evidence that deserves to be stressed is that the results of the loss aversion treatment improve the significance of the correlation with the SOEP as well as the amount of variance explained, compared to the baseline case. The R^2 turns out to be about 5%, very similar to that found in the basic specification by Dohmen et al. (2011). Note that the gambles proposed by Dohmen et al. (2011) involve choices between lotteries and certain amounts: similar to what has already been stressed when we described the Eckel and Grossman (2008a) method, subjects could in this case use the certain amount as a reference point against which the gamble would imply a possible loss. Stretching such an interpretation even further, the higher correlation of the SOEP with the loss aversion treatment constitutes suggestive evidence that individuals could implicitly frame abstract decisions under uncertainty as possibly involving some losses.

3.3 Truncation of the data: explosion

The main goal of this treatment is to check whether using the "time bomb" as a device to avoid truncation of the data might hamper the salience of risk in the BRET. If this is the case, subjects could postpone their decision to stop simply because they perceive the risk of the explosion as less likely than it really is as the explosion cannot happen during the task but only afterwards. In other words, they might feel that the bad luck of a negative outcome, as determined by the draw of the bomb's position, is less correlated to their decision to stop than it really is. Such an effect would imply that the BRET underestimates subjects' risk aversion.

The *Explosion* treatment differs from the Baseline task insofar as the bomb explodes immediately when collected. The subjects are told that the position of the bomb has been randomly predetermined and that the printout of the random generation of the number is available in a sealed envelope at the experimenters' desk. The bomb was located at $k = 62$; this means that the task could go on until the box containing the bomb (the 62^{nd}) is collected, in which case the bomb explodes and wipes out all earnings, or until the subject presses "Stop", in which case she earns the amount in her provisional account (0.1 euro per box).

From a theoretical point of view, the decision of the subject can be formalized as a sequence of lotteries. Calling $k \in [0, 99]$ the number of boxes already deleted, the subject compares the gains of stopping the task (equal to γk with probability one) against the uncertain outcome of proceeding by one more box, which corresponds to the following lottery:

$$L^{k+1} = \begin{cases} 0 & \frac{1}{100-k} \\ \\ \gamma(k+1) & \frac{99-k}{100-k}. \end{cases}$$

An expected utility maximizer decides the optimal exit point k^*, equalizing the expected gains of collecting an additional box with the (opportunity) cost of losing the certain amount γk^* that would be earned by stopping. The simplest case of a risk neutral agent leads to $k^* = 49.5$, which means that the subject will stop the task after 49 boxes, i.e., before the 50^{th} box is collected. The optimal stopping point changes according to the subjects' risk attitude. The earlier (later) the exit point, the stronger the risk aversion (loving), with half of the boxes being the threshold that separates risk aversion from risk seeking.

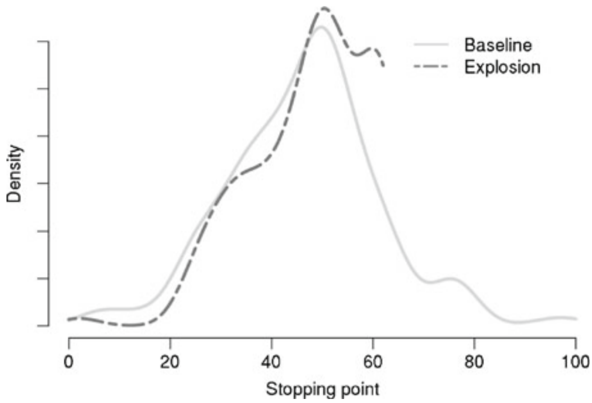
Assuming a generic CRRA utility function $u(x) = x^r$ the task allows estimation of the coefficient of risk aversion solving by numerical computation the following equation

$$k^* : \qquad \frac{99-k}{100-k}(k+1)^r = k^r \tag{3}$$

that leads to an equilibrium relation between k and r virtually identical to the case without an explosion. Therefore, the testable implication is that if in the Baseline treatment the risk is perceived as less salient we should observe that $k^*_{base} > k^*_{expl}$.

Results show that this is not the case. In fact, as reported in Fig. 6, the point estimate is even higher in the Explosion treatment. Although the difference is not significant, it is worth noting that the average in the Explosion treatment is biased downward since all those who would have liked to proceed beyond 62 have been stopped, and there is no way of retrieving their latent desired exit point.[13] The distribution of decisions confirms a similar pattern, with a larger fraction of players not exiting at $k = 61$ in the Explosion than in the Baseline treatment (20.5% vs. 11.9%). A Fisher exact test rejects, at $p = 0.12$, the null hypothesis that the likelihood of not having decided to exit is the same in the two treatments. In any case, we can safely conclude that delaying the draw of the bomb's position in order not to get truncated data does not induce less conservative behavior.

[13] A significantly higher k^* in the Explosion treatment could be rationalized in a Prospect Theory framework for values of the parameter such as those estimated in Tversky and Kahneman (1992). Note that the incentive structure in the Explosion treatment explicitly entails the possibility of incurring a loss in case the bomb is collected as long as subjects adjust their reference point to the amount of money earned every time they take a decision.



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

Fig. 6 Results and kernel density of decision by treatment, Baseline vs. Explosion

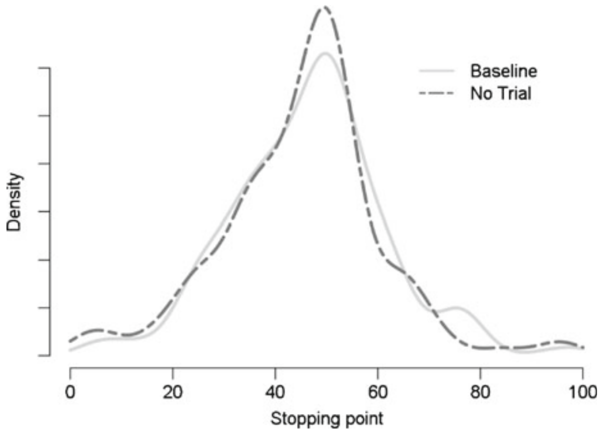
3.4 No trial

In the Baseline Dynamic treatment we used a trial period to let the subjects familiarize themselves with the visual interface before moving to the paying part of the experiment. This trial period was meant for exploratory purposes only. All the information provided stressed its 'trial' nature, and no feedback concerning the position of the bomb in the trial period was given in order to avoid focal points, mean-reversion strategies, or any other confounding factor. The absence of any explosion in the trial period, immediate or postponed, could affect the subsequent behavior nevertheless, as long as it generates a feeling of 'having beaten the bomb', thereby inducing a more risk tolerant behavior in the paying stage.

In order to test if a trial period without feedback can affect subjects' behavior, we ran a *No Trial* treatment in which they did not go through a practice period, but played directly the paying stage. All other details of the experiment were kept identical to the Baseline Dynamic.[14]

Results are shown in Fig. 7. No statistically significant difference in stopping behavior is found. The task is perceived as slightly more difficult, as self-reported complexity is 0.439 in No trial, 0.376 in the Baseline. The difference is, however, not significant (Mann-Whitney, Prob $> |z| = 0.174$).

[14] The alternative of showing the position of the bomb at the end of the trial period is not feasible, as we have clean evidence from a pilot experiment that it generates huge serial correlation. Subjects tend to adjust their choices as a function of the outcomes previously observed. For instance, they are prone to a form of gambler fallacy implying that they could not be lucky twice in a row thereby stopping the task earlier after having observed a relatively high b.



	N	Mean	Std. Dev.	Mann-Whitney z
Baseline	269	46.54	15.138	Prob $> \|z\| = 0.433$
No trial	123	45.19	15.59	

Fig. 7 Results and kernel density of decision by treatment, Baseline vs. No trial

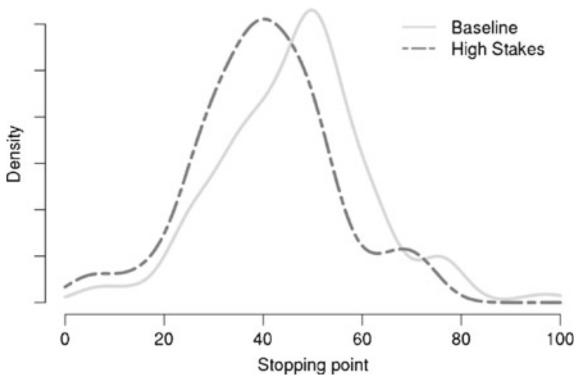
3.5 High stakes

It has already been shown in the literature that in risk elicitation tasks the amount of money at stake affects decisions (e.g., Holt and Laury (2002)). In order to check the robustness of this result in our setting, in the *High Stakes* treatment we doubled the value of each box, setting it to $\gamma = 0.2$ euro. The task has in this case an expected value that peaks at 5 euro when $k = 50$; theoretically choices should not be affected by such a change.

The distribution of stopping points in the High Stakes with respect to the Baseline treatment is reported in Fig. 8, which also summarizes the mean and provides a Mann-Whitney test of equality in distribution. In line with the existing literature, we find that higher stakes significantly increase the average measured degree of risk aversion.

3.6 Size treatments: 20×20, 5×5, and *mixed*

In the Baseline treatment, we opt for a 10×10 square because it greatly simplifies the perception of probabilities by the subjects, at the same time allowing estimation of a fine grid of values for r. However, the size of the square could be freely manipulated. Empirical evidence concerning the Balloon Analogue Risk Task (Lejuez et al. 2002) shows that subjects' behavior is not invariant to the salience of the single decision. In particular, subjects display a lower measured risk aversion the higher the salience of each decision. The counterpart of such a finding in the BRET can be tested by manipulating the size of the square *ceteris paribus*. Moreover, manipulating the size of the square can reduce the salience of a possible focal points at $k = 50$.



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

Fig. 8 Results and kernel density of decision by treatment, Baseline vs. High Stake

In the 5 × 5 treatment, the number of boxes is set to 25, laid out in 5 rows of 5. Each box is worth 40 cents and is deleted every 4 seconds (see Fig. 9, left panel). In this treatment, the center of the task (12.5) is not focal as it cannot be chosen.

In the 20 × 20 treatment, following the same principle, 400 boxes are laid out in 20 rows of 20, each worth 2.5 euro cents and deleted every quarter of a second.

Note that the speed of deletion and the value of each box have been modified in such a way that both the monetary incentives and the length of the task are identical to the Baseline. In this way, we avoid the introduction of confounds in the size treatment, leaving the probability that the bomb is in a given box as the only exogenous manipulation across conditions. Such a probability is equal to 1% in the Baseline, 4% in the 5 × 5, and 0.25% in the 20 × 20 treatment.

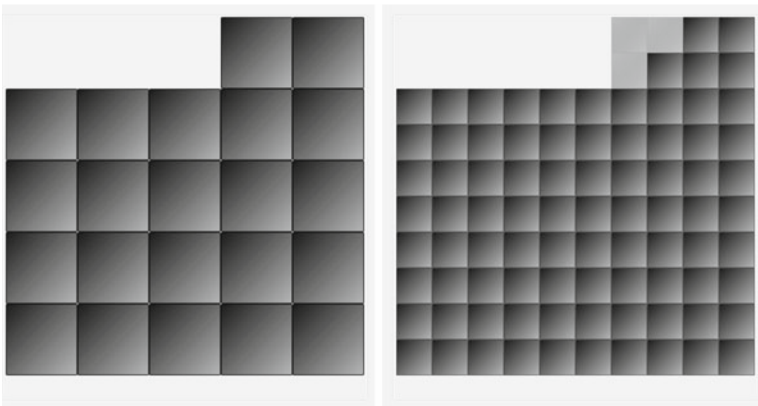
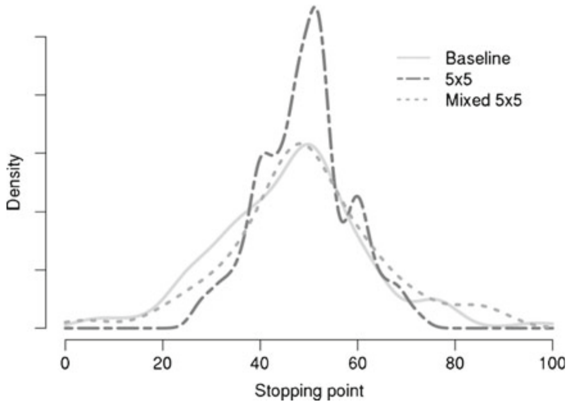


Fig. 9 Screenshot after 15 seconds in the 5 × 5 (*left*) and Mixed 5 × 5 (*right*)



	N	Mean	Std. Dev.	Mann-Whitney z
Baseline	269	46.55	15.14	
20 × 20	32	46.29	21.29	Prob $> \|z\| = 0.879$
5 × 5	92	49.13	9.11	Prob $> \|z\| = 0.067$
Mixed 5 × 5	55	49.40	15.63	Prob $> \|z\| = 0.249$
Kruskal-Wallis test				Prob $> H = 0.269$

Fig. 10 Results and kernel density of decision by treatment, Baseline vs. Fast

Note also that such a manipulation should not play a relevant role from a theoretical point of view, up to rounding problems. Assuming the classic power utility function $u(x) = x^r$ and calling N the total number of boxes, we have that the general solution is equal to

$$k^* = N\frac{r}{1+r'} \quad (4)$$

implying that the same fraction $r/(1+r)$ of boxes should always be collected, up to rounding problems. For instance, a risk neutral agent in the 5×5 treatment is expected to stop after having removed the 12^{th} box since the option 12.5 is not available. The choice $k^* = 12$ would correspond to $k^* = 48$ in the 10×10 where, however, a finer decision is allowed and the equivalent of $k^* = 12.5$, i.e., $k^* = 50$, can be chosen.

The theoretical prediction is therefore that we should observe $k^*_{20\times20} \geq k^*_{10\times10} \geq k^*_{5\times5}$ with inequality signs driven only by rounding problems, which should be more relevant between 5×5 and 10×10 than between 20×20 and 10×10.

Results are presented in Fig. 10 rescaling on a 100 basis the choices in the 5×5 and 20×20 treatments for the sake of comparability. The 20×20 treatment delivers average results virtually identical to the Baseline, which is why we decided to stop gathering further observations at $N = 32$. In contrast, the 5×5 treatment displays a significantly higher k^*.[15] When analyzing the quantiles, the distribution looks similar to the Baseline except for a much higher concentration around the mean: 52.1% of

[15]Note that due to rounding problems the estimated $k^*_{5\times5}$ is actually a lower bound of the choice once reported on a 100 scale.

subjects are risk averse ($k \leq 48$), exactly as in the Baseline. Moreover, it also hints to the fact that the higher mean is due to a lower number of extreme risk averse choices: at $k = 36$ the cumulative distribution records $\sim 25\%$ for the *Baseline* and only $\sim 10\%$ for the 5×5.

We tried to discover whether such a result is simply due to the visual characteristics of the task by setting up an additional condition, called *Mixed* 5×5, in which the structure is identical to the 5×5 while the visual layout is more similar to the 10×10 Baseline treatment. We achieve this goal using a 10×10 square, in which at every second one box is marked for deletion (its color changes to gray), but boxes are deleted in groups of 4 every 4 seconds (See Fig. 9, right panel).

The results show that the visual framework does not account for the different choices. The point estimate is extremely close to the 5×5, although due to a lower number of observations and a larger standard deviation, the difference with respect to the Baseline does not reach a significant level.

When comparing the results of all the Size treatments using a Kruskal-Wallis test, however, no statistical difference appears. The evidence of the size treatments is hence mixed.

3.7 Speed of deletion: *Fast*

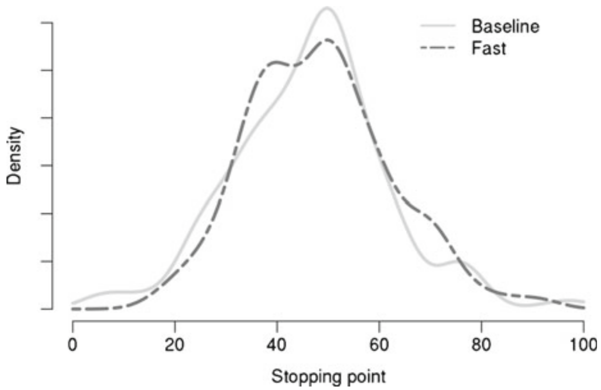
In order to check if patience or boredom have an impact on the choice, we ran a *Fast* treatment, in which a box is collected every half a second, doubling the frequency of draws. This means that the time needed to deplete the box field in the Fast treatment is 50 seconds instead of 100.

The treatment should have an effect only if patience plays a role in the choice, i.e., if subjects get bored waiting for their preferred outcome and hence click the "Stop" button before reaching it. A faster deletion entails other minor effects. First, if the time lag between the moment when the decision is taken and the moment when the mouse is clicked is longer than 0.5 seconds, the recorded result might be slightly higher. Second, subjects have less time to check the information on the screen. In general, the faster the task, the higher the possibility of slight errors and tremblings around the preferred value.

The result of the Fast treatment are presented in Fig. 11. The mean turns out not to be significantly different from the Baseline – though the point estimate is higher, possibly reflecting one of the two explanations above. The full distribution of the preferred stopping times confirms a substantial equivalence with the baseline, showing that the BRET can hence be safely sped up without altering its base results.

3.8 Random sequence

In the *Random* treatment, the deletion process follows a random, instead of arabic, predetermined sequence. The treatment aims to test if the salience of 50 could be a confound driving the results in the Baseline treatment and to introduce some ambiguity in the BRET. If subjects overlook the information about the number of boxes provided to the right of the square, tracking and counting the number of boxes collected becomes more difficult in the Random treatment.



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

Fig. 11 Results and kernel density of decision by treatment, Baseline vs. Fast

Results display a lower average k^* (44.53), not significantly different from the Baseline ($Prob > |z| = 0.398$). Moreover, the amount of risk neutral choices (13.3%) is similar to the baseline. As the results were not distinguishable, and as the task was both perceived as much more complex and was more difficult to run,[16] we limited the number of observations to 32.

3.9 Repeated task

The goal of this treatment is to check whether subjects, as predicted by the theory, behave in the same way in a one-shot task as compared to a multiple decision framework in which one random decision is paid. In the *Repeated* treatment subjects play the same task five times. At the end of the experiment one period is randomly selcted to be payoff-relevant, before the position of the bomb is determined. This $1 - in - k$ payment procedure is routinely employed in experiments, but it generates compound lotteries, which are subject to violations of the Reduction Axiom (Harrison and Swarthout 2012).

Results (see Table 5) show that on average, i.e. collapsing all periods, choices in the *Repeated* treatment are comparable to those in the *Baseline*. This is true also when comparing single repetitions to the Baseline except for repetition 2, which shows a significantly lower average. However, when comparing the five repetitions among themselves, a Kruskal-Wallis test signals that they cannot be said to be equal. We find a significant increasing trend overall, as k_5 is significantly higher than k_1.

[16] The *Random* treatment requires a predetermined sequence of deletion of the 100 cells. Subjects must therefore be made aware that, in this case, the bomb's position cannot be automatically found in the 10×10 square but requires the random sequence to be disclosed for the bomb to be located.

Table 5 Baseline vs. repeated choices

Treatment	N	Repetition	Average	Mann-Whitney test $P > \|z\|*$
Baseline Dynamic	269		46.54	
		Mean	47.16	0.952
		1	44.18	0.160
		2	42.59	0.013
Repeated	61	3	47.87	0.729
		4	50.34	0.303
		5	50.83	0.224
Kruskal-Wallis test (among repetitions)			0.036	
Wilcoxon signed-rank test, period 5 vs. 1			0.002	

*The Mann-Whitney tests refer to pairwise comparisons of each period of the repetitions against the baseline

While on average behavior does not change when aggregating choices across periods, repetitions introduce a sizeable variance of behavior at the individual level. We classify subjects into different types, according to the stability of their choices. *Stable* subjects have the maximum and the minimum choice not more than 10 blocks apart (e.g., 43 − 53). *Increasing* subjects show a constant (weakly) increasing trend, such that the choice in the second period is weakly higher than the one in the first, and so on. Moreover, the difference between the last and the first choice is bigger than 10. Finally, *Rollercoaster* subjects behave errantly, showing a difference higher than 10 between the maximum and the minimum as well as inverting the sign of the difference between any two choices in a row: for instance, they go first up, then down, then up again, etc. The remainder of the subjects features unstable choices ($\max(k) - \min(k) > 10$) that are not classifiable according to the above definitions.[17] Table 6 reports the breakdown by type, showing that only 29.51% of the subjects show some stability in their choices.

Summarizing, we find that the BRET is robust to repetitions, although at the price of a substantial variance of behavior at the individual level.

3.10 Wealth effects

When simple tasks are used at the end of experiments in order to measure risk aversion, they are usually unannounced and take place before the payoff for the main part of the experiment has been computed and shown to subjects. Nonetheless, subjects might know more or less precisely what their gains in the main part of the experiments are going to be. In that case, it is crucial to understand how risk aversion tests respond to wealth effects.

In order to investigate the reaction of the BRET to wealth effects, we carried out a specific treatment in which the BRET was performed unannounced, after subjects

[17] There is no subject displaying a steadily decreasing pattern.

Table 6 Type of subjects in the *Repeated* treatment

Type	N	Percent	Average
Stable	18	29.51	42.7
Increasing	7	11.48	47.22
Rollercoaster	19	31.15	52.47
Other Instable	17	27.87	45.84
Total	61	100	47.16

had played a one-shot version of the Balloon Analogue Risk Task (BART, Lejuez et al. (2002)). In this task, subjects must inflate a balloon that they see on the screen. For every time they pump air into it, they earn money, at the same time increasing the probability of it bursting, in which case they earn zero. What is relevant here is that at the end of the Balloon Task subjects know for sure the payoff they have accumulated, be it zero in case they exploded the balloon, or a positive amount in case they stopped short of the random exploding point. This allows us to test how wealth effects affect the BRET.

Results of a least squares regression of k^* on gains from the previous experiment (ranging from 0 to 6.1 euro) allowing for nonlinearities indicate that there is a significant convex wealth effect. Zero earnings in the previous experiment result in $k^* = 47$; k reaches a minimum of 40 when gains are about 2.7 euro, starting to increase again for higher gains.

4 Comparison of the BRET with other tasks

In order to explore how the BRET compares to other risk elicitation mechanisms we ran additional sessions involving a) a multiple price list, in its Holt and Laury (2005) incarnation (henceforth, HL) b) an ordered lottery choice as implemented by Eckel and Grossman (2002, 2008a) (EG) and c) the Investment Game by Charness and Gneezy (2010), Gneezy and Potters (1997) (CGP). Results are reported in Table 7.[18]

The HL task implements a series of 10 ordered binary choices between a 'safe' and a 'risky' lottery. The outcomes stay constant, while probabilities of the high payoff increase along the 10 choices. A consistent subject should start on the 'safe' side and switch at some point to the 'risky' option. This switching point reveals the coefficient of risk aversion.

In the EG task subjects choose the preferred among 5 binary lotteries characterized by increasing expected value and increasing variance driven by changing outcomes, while probabilities are fixed at 50-50. As the expected value is always increasing, a risk-neutral subject should always choose the lottery with the highest expected

[18] Many other risk elicitation mechanisms have been extensively used in the literature (for a review see Harrison and Rutström (2008)). We focus on the three above as they are fast and easy to implement and they do not result in a too high cognitive load on the subjects, and are, like the BRET, most suitable to be used as controls.

Table 7 Estimates of r and risk categories for all tasks

	N	SOEP	Type of choice	Choice set	Median choice	Median r	Type classification (%)			
							Averse	Neutral	Loving	N.C.
BRET	88	0.28	Stopping point	[0,100]	40	0.67	72.73	11.36	14.77	1.14
HL	88	0.33	Safe choices	[0,10]	6	0.32	70.45	4.55	7.95	17.05
EG	88	0.36	Chosen lottery	[1,5]	2	0.33	81.81	*na*	*na*	18.18
CGP	86	0.34	Amount invested	[0,4]	2.5	0.75	80.23	*na*	*na*	19.77

value. The task hence cannot tell apart slightly risk averse, risk neutral and risk loving preferences (see Appendix B for details, and in particular Table 8 for HL and Table 9 for EG).

In the CGP task, subjects are given an endowment that must be allocated between a safe account, which yields exactly the amount invested in it, and a risky option that yields 2.5 times the amount invested with 50% probability, and zero otherwise. As the expected return from the risky option is bigger than 1, both risk neutral and risk loving subjects should invest the whole endowment.

In all three cases we followed closely design, procedures, and instructions from the original papers except for the level of payoffs, which were very different across tasks. For the sake of comparability we used very similar payoffs, in the order of 5 Euro for the decision of a risk neutral agent. Therefore, we compare the results with our *High Stakes* treatment.

The distribution of choices obtained in all three tasks is in line with the one found in the original papers and in several replications in the literature. In order to compare outcomes of the 4 different tasks, the choices have been converted into risk-aversion parameters assuming a CRRA utility function x^r. Results show that the BRET and CGP measure a lower estimated degree of risk aversion as compared to HL and EG although the subjects recruited for the comparison sessions report a very similar and not significantly different average SOEP choice (Kruskal-Wallis joint test, $p = 0.719$).[19] Moreover, the BRET estimates a higher share of risk neutral and risk loving subjects with respect to all other tasks, and it minimizes the number of observations that cannot be assigned to any category due to inconsistent or dominated choices, or to the impossibility of assigning the observation to one of the three categories (N.C. in Table 7).

5 Conclusion

The paper presents the results of a large-scale experiment aimed at introducing and validating the Bomb Risk Elicitation Task (BRET), an intuitive procedure designed to

[19]For more details and to see the different way in which the tasks map choices into r see Crosetto and Filippin (2012).

measure risk attitudes. Subjects are presented a task in which they are asked to decide at which point to stop collecting a series of 100 boxes, one of which contains a time bomb. Earnings increase linearly with the number of boxes collected but are equal to zero if one of the boxes collected contains the time bomb. The bomb's position is determined at the end of the experiment, thereby avoiding potential truncation of the data, so that subjects are free to choose any number between 0 and 100.

The static version can be run with paper and pencil, while the dynamic version consists of a visual task in continuous time that makes the underlying structure in terms of probability and outcomes of the 100 lotteries even simpler and more intuitive to understand. The dynamic version of the BRET is therefore well suited for subjects with low numeracy skills and particularly appropriate to measure risk attitudes in decision tasks that entail a time dimension.

The BRET implies a single choice, and therefore it is not affected by possible violations of the Reduction Axiom. The absence of any explicit or implicit reference point different from zero ensures a pure measure of risk aversion, unaffected by the degree of loss aversion. This allows us to also contribute to the literature on gender differences in risk attitudes, with results that bring support to the hypothesis that females are characterized by stronger *loss* aversion while displaying a similar degree of *risk* aversion. The task allows precise estimation of both risk aversion and risk seeking. The cumulative distribution of choices in the dynamic BRET shows that 51.3% of subjects are risk averse, 14.1% risk neutral and 34.6% risk seekers. Splitting the distribution by quartiles shows that 50% of the observations lie between 26 and 55 boxes. Hence the bulk of subjects is either risk averse or risk neutral. The fraction of slightly risk loving subjects turns out to be relatively high when compared to other tasks in the literature.

The BRET is quite robust to several changes in its protocol, such as the speed and the order of deletion of the boxes. Choices react significantly only to the amount of money at stake. Results are sensitive to wealth effects, suggesting that when used as control, the BRET should be performed at the beginning of the experiment.[20]

The possible drawback of the wide range of choices available to subjects is the sensitivity of results to outliers and decisions far in the tails of the distribution. Sensitivity analysis, deleting from the pooled sample the 2.5% of the observations in each tail, slightly affects the results only in a few cases. Nevertheless, this suggests that even greater attention should be paid to reduce the measurement error induced by extreme choices (by means, e.g., of control questions or confirmation screens) in order to make sure that they are deliberate and do not reflect an imperfect comprehension of the task.

The BRET is highly flexible and can be easily modified to test for a host of different issues. For instance, exogenously inducing different reference points allows us to fully estimate the parameters of a Prospect Theory value function. Moreover, the task can easily accommodate ambiguity if the visual and numerical information about the probability of explosion is minimized, randomizing the order of deletion

[20] Note that the BRET itself does not induce wealth effects as long as the position of the bomb is determined at the very end of the experiment.

and increasing its speed at the same time. Finally, changing the frame of the choice from stopping the automatic process to actively deciding how long to proceed collecting boxes offers the possibility to test if subjects suffer from illusion of control. All these and other issues are left for future research but testify to the extreme flexibility of the Bomb Risk Elicitation Task.

Acknowledgments We are grateful to the Max Planck Institute of Economics (Jena) for financial and logistic support and to Denise Hornberger, Nadine Marmai, Florian Sturm, and Claudia Zellmann for their assistance in the lab. We would like to thank Alexia Gaudeul and Gerhard Riener for helpful suggestions and participants at the ESA 2012 Conferences in New York and Cologne and at the Nordic Conference on Behavioral and Experimental Economics (Bergen), participants of seminars at the MPI Jena and at the University of Stirling, as well as an anonymous referee, for their comments. All remaining errors are ours.

Appendix A. Estimates of r for the BRET, assuming CRRA $u(k) = k^r$

K	r	K	r	K	r
1	$0 \le r \le 0.014$	**36**	$0.551 \le r \le 0.574$	**71**	$2.39 \le r \le 2.508$
2	$0.015 \le r \le 0.025$	**37**	$0.575 \le r \le 0.599$	**72**	$2.509 \le r \le 2.636$
3	$0.026 \le r \le 0.036$	**38**	$0.6 \le r \le 0.625$	**73**	$2.637 \le r \le 2.773$
4	$0.037 \le r \le 0.046$	**39**	$0.626 \le r \le 0.652$	**74**	$2.774 \le r \le 2.921$
5	$0.047 \le r \le 0.058$	**40**	$0.653 \le r \le 0.68$	**75**	$2.922 \le r \le 3.081$
6	$0.059 \le r \le 0.069$	**41**	$0.681 \le r \le 0.709$	**76**	$3.082 \le r \le 3.255$
7	$0.07 \le r \le 0.08$	**42**	$0.71 \le r \le 0.739$	**77**	$3.256 \le r \le 3.444$
8	$0.081 \le r \le 0.092$	**43**	$0.74 \le r \le 0.769$	**78**	$3.445 \le r \le 3.651$
9	$0.093 \le r \le 0.104$	**44**	$0.77 \le r \le 0.801$	**79**	$3.652 \le r \le 3.878$
10	$0.105 \le r \le 0.117$	**45**	$0.802 \le r \le 0.834$	**80**	$3.879 \le r \le 4.129$
11	$0.118 \le r \le 0.129$	**46**	$0.835 \le r \le 0.869$	**81**	$4.13 \le r \le 4.406$
12	$0.13 \le r \le 0.142$	**47**	$0.87 \le r \le 0.904$	**82**	$4.407 \le r \le 4.715$
13	$0.143 \le r \le 0.155$	**48**	$0.905 \le r \le 0.941$	**83**	$4.716 \le r \le 5.062$
14	$0.156 \le r \le 0.169$	**49**	$0.942 \le r \le 0.98$	**84**	$5.063 \le r \le 5.453$
15	$0.17 \le r \le 0.183$	**50**	$0.981 \le r \le 1.02$	**85**	$5.454 \le r \le 5.898$
16	$0.184 \le r \le 0.197$	**51**	$1.021 \le r \le 1.061$	**86**	$5.899 \le r \le 6.41$
17	$0.198 \le r \le 0.212$	**52**	$1.062 \le r \le 1.105$	**87**	$6.411 \le r \le 7.003$
18	$0.213 \le r \le 0.226$	**53**	$1.106 \le r \le 1.15$	**88**	$7.004 \le r \le 7.7$
19	$0.227 \le r \le 0.242$	**54**	$1.151 \le r \le 1.197$	**89**	$7.701 \le r \le 8.53$
20	$0.243 \le r \le 0.257$	**55**	$1.198 \le r \le 1.247$	**90**	$8.531 \le r \le 9.534$
21	$0.258 \le r \le 0.273$	**56**	$1.248 \le r \le 1.298$	**91**	$9.535 \le r \le 10.776$
22	$0.274 \le r \le 0.29$	**57**	$1.299 \le r \le 1.352$	**92**	$10.777 \le r \le 12.351$
23	$0.291 \le r \le 0.307$	**58**	$1.353 \le r \le 1.409$	**93**	$12.352 \le r \le 14.412$
24	$0.308 \le r \le 0.324$	**59**	$1.41 \le r \le 1.469$	**94**	$14.413 \le r \le 17.229$
25	$0.325 \le r \le 0.342$	**60**	$1.47 \le r \le 1.531$	**95**	$17.23 \le r \le 21.309$
26	$0.343 \le r \le 0.36$	**61**	$1.532 \le r \le 1.597$	**96**	$21.31 \le r \le 27.76$
27	$0.361 \le r \le 0.379$	**62**	$1.598 \le r \le 1.666$	**97**	$27.761 \le r \le 39.532$
28	$0.38 \le r \le 0.398$	**63**	$1.667 \le r \le 1.739$	**98**	$39.533 \le r \le 68.274$
29	$0.399 \le r \le 0.418$	**64**	$1.74 \le r \le 1.816$	**99**	$r \ge 68.275$
30	$0.419 \le r \le 0.438$	**65**	$1.817 \le r \le 1.898$		
31	$0.439 \le r \le 0.459$	**66**	$1.899 \le r \le 1.985$		
32	$0.46 \le r \le 0.481$	**67**	$1.986 \le r \le 2.077$		
33	$0.482 \le r \le 0.503$	**68**	$2.078 \le r \le 2.174$		
34	$0.504 \le r \le 0.526$	**69**	$2.175 \le r \le 2.278$		
35	$0.527 \le r \le 0.55$	**70**	$2.279 \le r \le 2.389$		

Appendix B. Experimental instructions

The experimental instructions were originally drafted in English, then translated into German to enable us to run the experiments in the Max Planck Institute's lab in Jena, Germany. In what follows, we will report the original, English versions of the instructions for the Baseline versions and robustness controls. The German versions are available in the additional online material at http://goo.gl/3eogr.

B.1 Baseline BRET, static

First screen Welcome to the Experiment. In the experiments all payoffs are expressed in euro. For your punctuality you receive 2.5 euro. The experiment consists of one short task, followed by a questionnaire. Should you have any questions or need help, please raise your hand. An experimenter will then come to your place and answer your questions in private.

Second screen On the sheet of paper on your desk you see a field composed of 100 numbered boxes. Behind one of these boxes a time bomb is hidden; the remaining 99 boxes are empty. You do not know where the time bomb is. You only know that it can be in any place with equal probability.

Your task is to choose how many boxes to collect. Boxes will be collected in numerical order. So you will be asked to choose a number between 1 and 100.

At the end of the experiment, we will randomly determine the number of the box containing the time bomb by means of a bag containing 100 numbered tokens.

If you happen to have collected the box in which the time bomb is located – i.e., if your chosen number is greater than, or equal to, the drawn number – you will earn zero. If the time bomb is located in a box that you did not collect – i.e., if your chosen number is smaller than the drawn number – you will earn an amount in euro equivalent to the number you have chosen divided by ten.

In the next screen you will be asked to indicate how many boxes you would like to collect. You confirm your choice by hitting OK.

B.2 Baseline BRET, dynamic

The following instructions, containing no changes, were used in the Baseline dynamic, Wealth effect, and Repeated treatments. Only slight parameter and word changes were needed to adapt them to the Fast, High Stake, 5×5, 20×20, Mixed 5×5, No Trial and Random treatments. The slight changes are indicated within brackets in the text.

First screen: as in Baseline static

Second screen On the sheet of paper on your desk you see a field composed of 100 {5×5*: 25;* 20×20*: 400*} numbered boxes.

You earn 10 euro cents {5×5*: 40 euro cents;* 20×20*: 2.5 euro cents*} for every box that is collected. Every second {*Fast: half a second;* 20×20*: quarter of a* *second;* 5×5*: four seconds*} a box is collected {*Mixed* 5×5*: marked for collection*}, starting from the top left corner {*Random: following the sequence reported on the sheet of paper*}. Once collected {*Mixed* 5×5*: Once 4 boxes have been marked for collection*}, the box disappears {*Mixed* 5×5*: 4 boxes disappear*} from the screen, and your earnings are updated accordingly. At any moment you can see the amount earned up to that point.

Such earnings are only potential, however, because behind one of these boxes a time bomb is hidden that destroys everything that has been collected.

You do not know where the time bomb is. You only know that it can be in any place with equal probability. Moreover, even if you collect the bomb, you will not know it until the end of the experiment.

Your task is to choose when to stop the collecting process. You do so by hitting 'Stop' at any time. At the end of the experiment, we will randomly determine the number of the box containing the time bomb by means of a bag containing 100 {5×5*: 25;* 20×20*: 400*} numbered tokens. If you happen to have collected the box in which the time bomb is located, you will earn zero. If the time bomb is located in a box that you did not collect, you will earn the amount of money accumulated when hitting 'Stop'.

We will start with a practice round. After that, the paying experiment starts. *No Trial: Please note that there will be no trial period. You will take only one decision, and this will be payoff relevant. You will see now the main interface of the experiment. Take your time to observe it, before you click on 'Start'.*

B.3 Explosion

First screen: as in Baseline static

Second screen On the sheet of paper on your desk you see a field composed of 100 numbered boxes.

You earn 10 euro cents for every box that is collected. Every second a box is collected, starting from the top left corner. Once collected, the box disappears from the screen, and your earnings are updated accordingly. At any moment you can see the amount earned up to that point.

Behind one of these boxes a bomb is hidden that destroys everything that has been collected. You do not know where the bomb is. You only know that it can be in any place with equal probability.

Your task is to choose when to stop the collecting process. You do so by hitting 'Stop' at any time.

If you collect the box in which the bomb is located, the bomb will explode and you will earn zero. If you stop before collecting the bomb, you gain the points accumulated that far. The position of the bomb in the paying round has been randomly determined beforehand, and the documentation of the drawing process is available in a sealed envelope at the experimenters' desk.

We will start with a practice round. The practice round is only meant to demonstrate how the experiment works: there will be no explosion. After that, the paying experiment starts.

B.4 Loss aversion

First screen Welcome to the Experiment. In the experiments all payoffs are expressed in euro. For your punctuality you receive 2.5 euro. On top of that, you have received an initial endowment of 2.5 euro, which you can see on your desk. PLEASE NOTE that this is NOT the show-up fee that will be paid at the end of the experiment, but it is given to you on top of that. Please keep this additional amount of money on your desk. The experiment consists of one short task, followed by a questionnaire. Should you have any questions or need help, please raise your hand. An experimenter will then come to your place and answer your questions in private.

Second screen The initial endowment of 2.5 euro that lies in front of you will be at stake during the experiment, according to the following rules.

On your screen you will see a field composed of 100 boxes. Every box is worth 10 euro cents, which you receive for every box collected. Every second a box is collected, starting from the top-left corner following the sequence reported on the sheet of paper. Once collected, the box disappears from the screen.

You start losing all of the 2.5 euro. Your losses are then reduced by 10 euro cents for every box collected. If you collect enough boxes, you will not only offset the losses, but you will also earn additional money, at the same value of 10 euro cents for each box. At any moment you can see the amount of your losses or gains with respect to your initial 2.5 euro.

Such gains or losses are only potential, however, because behind one of these boxes a time bomb is hidden that destroys everything that has been collected.

You do not know where the bomb is. You only know that it can be in any place with equal probability. Moreover, even if you collect the bomb, you will not know it until the end of the experiment.

Your task is to choose when to stop the collecting process. You do so by hitting ‘Stop’ at any time.

At the end of the experiment we will randomly determine the number of the box containing the bomb by means of a bag containing 100 numbered tokens. If you happen to have collected the box in which the bomb is located, you will lose all of your initial 2.5 euro. If the bomb is located in a box that you did not collect, you will earn your initial 2.5 euro plus or minus the gains or losses that you had accumulated when hitting ‘Stop’.

We will start with a practice round. After that, the paying experiment starts.

B.5 HL (modeled on instructions from Holt and Laury 2002)

You will be asked to make 10 choices. Each decision is a paired choice between “Option A” and “Option B”. For each decision row you will have to choose between Option A and Option B. You may choose A for some decision rows and B for other rows, and you may change your decisions and make them in any order.

Even though you will make ten decisions, only one of these will end up affecting your earnings. You will not know in advance which decision will be used. Each decision has an equal chance of being relevant for your payoffs.

Table 8 The 10 lotteries chosen for the HL treatment

	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 €

Now, please look at Decision 1 at the top. Option A pays 4 euro if the throw of the ten-sided die is 1, and it pays 3.2 euro if the throw is 2-10. Option B yields 7.7 euro if the throw of the die is 1, and it pays 0.2 euro if the throw is 2-10.

The other Decisions are similar, except that as you move down the table, the chances of the higher payoff for each option increase. In fact, for Decision 10 in the bottom row, the die will not be needed since each option pays the highest payoff for sure, so your choice here is between 4 or 7.7 euro.

To determine payoffs we will use a ten-sided die, whose faces are numbered from 1 to 10. After you have made all of your choices, we will throw this die twice, once to select one of the ten decisions to be used, and a second time to determine what your payoff is for the option you chose, A or B, for the particular decision selected.

Table 9 The 5 lotteries chosen for the EG treatment

	Choice	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€

B.6 EG (modeled on instructions from Eckel and Grossman 2008a)

You will be asked to select from among five different gambles the one gamble you would like to play. The five different gambles will appear on your screen. You must select one and only one of these gambles. Each gamble has two possible outcomes (Event A or Event B), each happening with 50% probability.

Your earnings will be determined by: 1) which of the five gambles you select; and 2) which of the two possible events occur.

At the end of the experiment, we will roll a six-sided die to determine which event will occur. If a 1, 2, or 3 is rolled, then Event A will occur. If 4, 5, or 6 are rolled, then Event B will occur.

B.7 CGP (modeled on instructions from Gneezy and Potters 1997)

You will be given 4 euros and will be asked to choose the portion of this amount (between 0 and 4 euros, in cents) that you wish to invest in a risky option. The money not invested is yours to keep.

There is a 50% chance that the investment in the risky asset will be successful. If it is successful, you receive 2.5 times the amount invested; if the investment is unsuccessful, you lose the amount invested.

The roll of a 6-sided die determines the value of the risky asset. You will be asked to choose 3 success numbers; if one of these numbers is rolled, the risky investment is successful; if not, it is not successful.

After the decisions are made the die will be rolled and then you will be paid the amount not invested plus 2.5 times the investment if it is successful and plus zero if it is not.

References

Becker, G., DeGroot, M., Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*, 226–236.

Bernasconi, M., & Loomes, G. (1992). Failures of the reduction principle in an Ellsberg-type problem. *Theory and Decision*, *32*(1), 77–100.

Blais, A.-R., & Weber, E.U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, *1*, 33–47.

Booij, A., & de Kuilen, G.V. (2009). A parameter-free analysis of the utility of money for the general population under prospect theory. *Journal of Economic Psychology*, *30*(4), 651–666.

Brooks, P., & Zank, H. (2005). Loss averse behavior. *Journal of Risk and Uncertainty*, *31*(3), 301–325.

Charness, G., & Gneezy, U. (2010). Portfolio choice and risk attitudes: an experiment. *Economic Inquiry*, *48*(1), 133–146.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50–58.

Cox, J.C., Roberson, B., Smith, V.L. (1982). *Theory and behavior of single object auctions*. Greenwich: JAI Press.

Crosetto, P., & Filippin, A. (2012). *A theoretical and experimental appraisal of five risk elicitation methods. Jena Economic Research Paper 2013-009*. Jena: Friedrich-Schiller University and the Max-Planck-Institute of Economics.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–74.

Dave, C., Eckel, C., Johnson, C., Rojas, C. (2010). Eliciting risk preferences: when is simple better? *Journal of Risk and Uncertainty*, *41*(3), 219–243.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., Wagner, G.G. (2011). Individual risk attitudes: measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

Eckel, C.C., & Grossman, P.J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, *23*(4), 281–295.

Eckel, C., Weber, E., Wilson, R.K. (2003). Four ways to measure risk attitudes.

Eckel, C.C., & Grossman, P.J. (2008a). Forecasting risk attitudes: an experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, *68*(1), 1–17.

Eckel, C.C., & Grossman, P.J. (2008b). Men, women and risk aversion: experimental evidence. *Handbook of experimental economics results* (Vol. 1, Ch. 113, pp. 1061–1073). Elsevier.

Gaechter, S., Johnson, E.J., Herrmann, A. (2010). *Individual-level loss aversion in riskless and risky choices.* Discussion papers 2010-20, The Centre for Decision Research and Experimental Economics. School of Economics: University of Nottingham.

Garcia-Gallego, A., Georgantzis, N., Jaramillo-Gutiérrez, A., Parravano, M. (2012). *The lottery-panel task for bi-dimensional parameter-free elicitation of risk attitudes*. Tech. rep., Department of Economic Theory and Economic History of the University of Granada.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, *112*(2), 631–45.

Grether, D.M., & Plott, C.R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, *69*(4), 623–38.

Halevy, Y. (2007). Ellsberg revisited: an experimental study. *Econometrica*, *75*(2), 503–536.

Harbaugh, W., Krause, K., Vesterlund, L. (2010). The fourfold pattern of risk attitudes in choice and pricing tasks. *The Economic Journal*, *120*(545), 595–611.

Harrison, G.W. (1990). Risk attitudes in first-price auction experiments: a Bayesian analysis. *The Review of Economics and Statistics*, *72*(3), 541–46.

Harrison, G.W., & Rutström, E.E. (2008). Risk aversion in the laboratory. In J.C. Cox, & G.W. Harrison (Eds.), *Risk aversion in experiments* (Vol. 12, pp. 41–196). Bingley: Emerald Group Publishing Limited.

Harrison, G.W., & Swarthout, J.T. (2012). The independence axiom and the bipolar behaviorist. Experimental economics center working paper series 2012-01, experimental economics center. Andrew Young School of Policy Studies, Georgia State University. Unpublished working paper.

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

Holt, C.A., & Laury, S.K. (2005). Risk aversion and incentive effects: new data without order effects. *American Economic Review*, *95*(3), 902–912.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Kaivanto, K., & Kroll, E.B. (2011). Negative recency, randomization device choice, and reduction of compound lotteries. Working Paper Series in Economics 22, Karlsruhe Institute of Technology (KIT). Department of Economics and Business Engineering.

Karni, E., & Safra, Z. (1987). Preference reversal and the observability of preferences by experimental methods. *Econometrica*, *55*(3), 675–85.

Lejuez, C., Read, J., Kahler, C., Richards, J., Ramsey, S., Stuart, G., Strong, D., Brown, R. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75.

Schmidt, U., & Traub, S. (2002). An experimental test of loss aversion. *Journal of Risk and Uncertainty*, *25*(3), 233–49.

Schubert, R., Brown, M., Gysler, M., Brachinger, H. (1999). Financial decision-making: are women really more risk-averse? *American Economic Review*, *89*(2), 381–385.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

van Rossum, G. (1995). Python reference manual. CWI Report CS-R9525.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Wagner, G.G., Frick, J.R., Schupp, J. (2007). The german socio-economic panel study (soep): scope, evolution and enhancements. SOEP papers on multidisciplinary panel data research 1, The German Socio-Economic Panel (SOEP), DIW Berlin.

Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*, *42*(8), 1131–1150.