DISCUSSION PAPER SERIES

IZA

IZA DP No. 6710

The "Bomb" Risk Elicitation Task

Paolo Crosetto
Antonio Filippin

July 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The “Bomb” Risk Elicitation Task

Paolo Crosetto
Max Planck Institute of Economics

Antonio Filippin
*University of Milan
and IZA*

Discussion Paper No. 6710
July 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 6710
July 2012

ABSTRACT

The "Bomb" Risk Elicitation Task*

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 containing 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 to precisely estimate 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 task and test its robustness in a large-scale experiment. Choices react significantly to the stakes and to the size of the choice set. Our experiment rationalizes the gender gap that often characterizes choices under uncertainty by means of a higher loss rather than risk aversion.

JEL Classification: C81, C91, D81

Keywords: risk aversion, loss aversion, elicitation method

Corresponding author:

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

* 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 Conference (New York) and of a seminar at the MPI in Jena for their comments. All remaining errors are ours.

1. Introduction

Uncertainty is a recurrent feature of the decision-making process in several domains such as investment, insurance, education, tax compliance, and labor market choices. 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 Becker et al. (1964); Eckel and Grossman (2002); Gneezy and Potters (1997); Holt and Laury (2005); Wakker and Deneffe (1996). 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, 2006). However, as Garcia-Gallego et al. (2010) point out, in the economic 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 game 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.

The task proposes to the subjects in a very short time (less than two minutes) 100 lotteries fully described in terms of outcomes and probabilities by a single parameter, the number of boxes collected. Relying on a visual representation in continuous time, the task gives an intuitive and transparent illustration of probabilities and outcomes, making it easy to understand and therefore suitable to be administered also to subjects with low numeracy skills. In fact, as reported in Crosetto and Filippin (2012),[1] the BRET is ranked by subjects as among the simplest tasks of those administered. It allows to precisely estimate both risk aversion and risk seeking, generating a more or less 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.

[1] The paper is currently in a draft form and has not yet been circulated. It is available upon request.

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 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 to wealth effects, to the amount of money at stake, and, to a lower extent, to the size of the choice set. 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 2.5% of the observations in each tail. In a few cases, this slightly affects the results.

The outline of the paper is as follows. In Section 2, we summarize the Bomb Risk Elicitation Task along with the underlying theoretical framework. The experimental procedure is summarized in Section 3, while the evidence obtained from the baseline versions as well as the treatments that we perform in order to validate the BRET are presented in Section 4. Section 5 concludes.

2. The Bomb Risk Elicitation Task

2.1. 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' decision can be formalized as the choice of their favorite among the lotteries

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

[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.

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 lottery 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 to estimate 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.

2.2. *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 Figure 1. The subject 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:

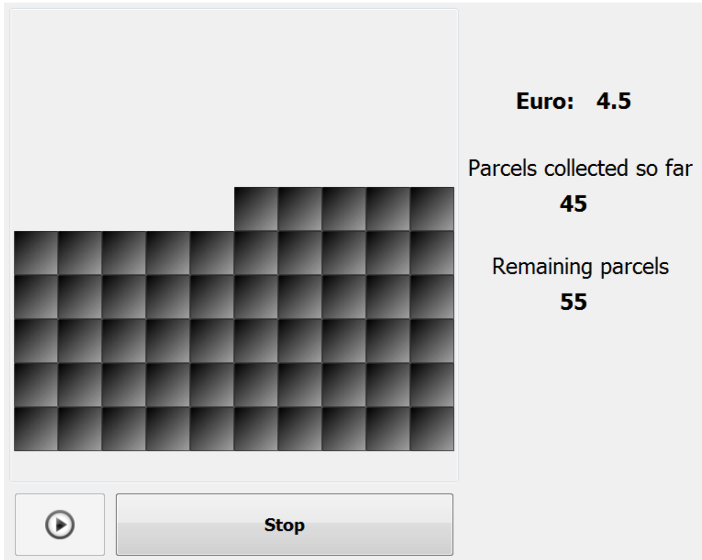


Figure 1: The BRET interface after 45 seconds - dynamic version

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

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 game and faces the same opportunity set.

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.[4] The task allows to precisely estimate 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

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

[4] The comparison of the BRET with the main elicitation tasks in the literature is specifically addressed in a companion paper (Crosetto and Filippin, 2012).

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). 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 4 below) and to compare the BRET with other risk elicitation mechanisms (Crosetto and Filippin, 2012), although increasing the amount of money at stake in the latter case.

3. Experimental Procedure

The experiment was run between March and May 2012 at the laboratory of the Max Planck Institute of Economics in Jena, Germany. A total of 1110 subjects participated, distributed over 38 sessions, lasting at most half an hour.[5] The sample includes mainly, but not only,[6] 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 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 4.4).

The experiment was computerized and the script we used to collect and organize the raw data were programmed in Python (van Rossum, 1995).[7]

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 of the task in order to experience the visual representation of the game. At the end of the trial period, however, there was no draw of the

[5] The experimental protocol was refined using insights from two incentivized pilots (with 29 and 88 subjects, respectively) and a non-incentivized experiment conducted during a scientific showcase exhibition at the Friedrich Schiller University Jena (with 198 subjects).

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

[7] 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`.

bomb's position in order not to provide the subjects with any reference point.[8] 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 a number of additional questions:

1. The Domain Specific Risk Taking (DOSPERT) battery of questions. The DOSPERT (Blais, 2006) risk questionnaire and scale is a validated measure of risk attitude as a personality trait, taking into account and weighting several different domains in which risk attitudes can play a role, i.e., ethical, financial (further decomposed into gambling and investment), health/safety, social, and recreational decisions.
2. A self-reported measure of risk attitudes, the general risk question used in the German Socio-Economic Panel (SOEP) 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%).
3. 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.

4. 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 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.[9] For all these reasons we decided to label the 17 subjects who stopped at $k = 0$, $k = 1$, and $k = 100$ as outliers and eliminate them from the dataset, leaving a total of 1093 observations. The descriptive statistics of the resulting sample are provided in Table 1.

Of the 1093 subjects of the resulting sample, 353 took part in the baseline treatments (Static and Dynamic) described in Section 2, while 740 took part in the additional treatments that we run to explore several possible departures from the baseline version of the task.

[8] We have evidence from a pilot experiment that notification of the explosion generated a huge serial correlation because subjects tended to change their choice as a function of the outcome previously observed. For instance, subjects were prone to a form of gambler fallacy such that they thought they could not be lucky twice in a row and stop the task earlier after having observed a relatively high b.

[9] On average, the subjects who submitted dominated choices reported only a slightly higher perceived difficulty of the task compared to the rest of the sample.

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

Table 1: Demographics of the experimental sample, excluding outliers

The breakdown by treatment of our experiment, including the robustness checks described below in more detail, is summarized in Table 2.

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 size of the field of the boxes, the sum at stake, the order in which the boxes are deleted, and the speed of the deletion process). Finally, we analyze the sensitivity of the task to repetition and wealth effects.[10] We adopted a between-subject design for the majority of treatments except for treatments *Wealth effects* and *Repeated*, where we had to rely on a within-subject design.

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

Table 2: Summary of the treatments (outliers excluded in N)

Table 2 also shows that the number of outliers, 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 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 if we restrict the sample dropping from the pooled sample of 1110 observa-

[10]The number of subjects for the *Repeated* treatment is bracketed in Table 2 because it was performed (unannounced) after one of the first baseline sessions. Hence, these 30 subjects are already counted in the first row.

tions (including the outliers) the about 2.5% most extreme choices in each tail. Given the fact that outliers 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.

4.1. Static and Dynamic Version of the BRET

Results of the Static and Dynamic treatments are summarized in Figure 2. As reported in the upper 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 $(Prob > |z| = 0.58)$.

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 outliers present in the Static sessions (see Table 2). The self-reported degree of complexity of these outliers is also higher $(3.19\, vs\, 2.49)$, though not significantly so. The sensitivity analysis performed without the 5% of extreme choices leads to identical results in the two treatments (see lower table in Fig. 2).

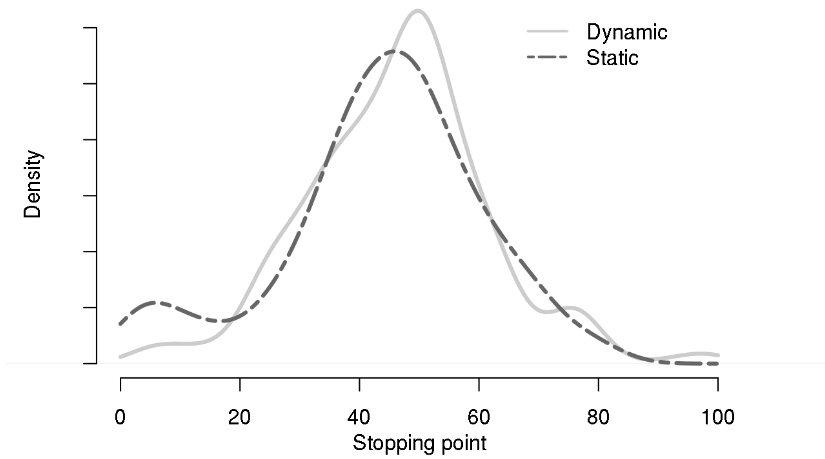
Given the similarity of results and the emergence of comprehension problems in the Static treatment, we concluded that the visual, *Dynamic* task is more likely to produce reliable outcomes, especially when it is administered to subjects characterized by low numeracy skills. We hence used the *Dynamic*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

Assuming a CRRA utility function, Equation 2 allows 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 ($k = 50$), 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 gap in risk aversion can be observed (Table 3). This finding is at odds with other results in the literature, which show that females display a more conservative behavior in decisions under uncertainty, as reported for instance by Charness and Gneezy (2011), Charness and Viceisza (2011) and in the survey by 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.[11] Finally, pooling males and females we observe that the choices in the Dynamic treatment correlate positively and significantly (at 10%) with the SOEP answers, but the variance explained is very low ($R^2 \cong 0.01$).

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

[11]Significantly, a gender gap can also been observed in both treatments in the self-reported DOSPERT scale. Results are not presented for the sake of brevity but are available upon request.



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

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

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. For instance, Gneezy and Potters (1997) and Charness and Gneezy (2010) 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.

A possible reading of the mixed results from the literature on gender differences in risk

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

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

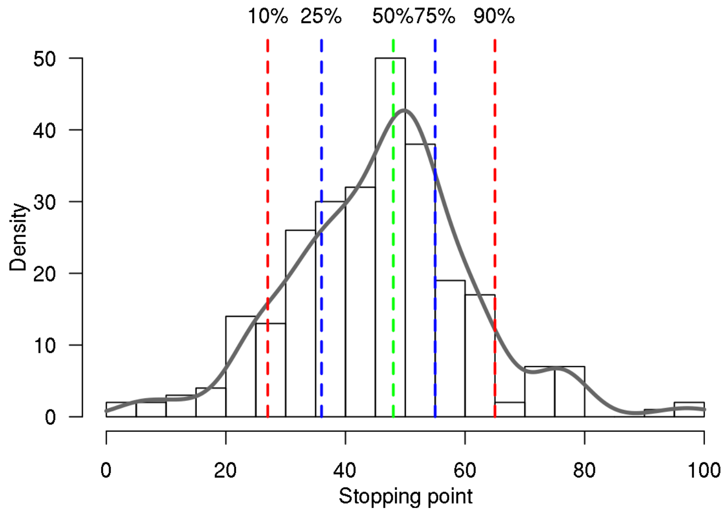


Figure 3: Distribution of decisions in the dynamic version

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 against which evaluating gains and losses if in their choice sets is present an option that guarantees a positive amount with probability one, even though te framing of the experiment only entails gains.

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.[12] 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.

4.2. 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,

[12] 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).

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 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.[13]

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 this 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.[14] 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 Figure 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 Figure 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

[13] 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.

[14] 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.

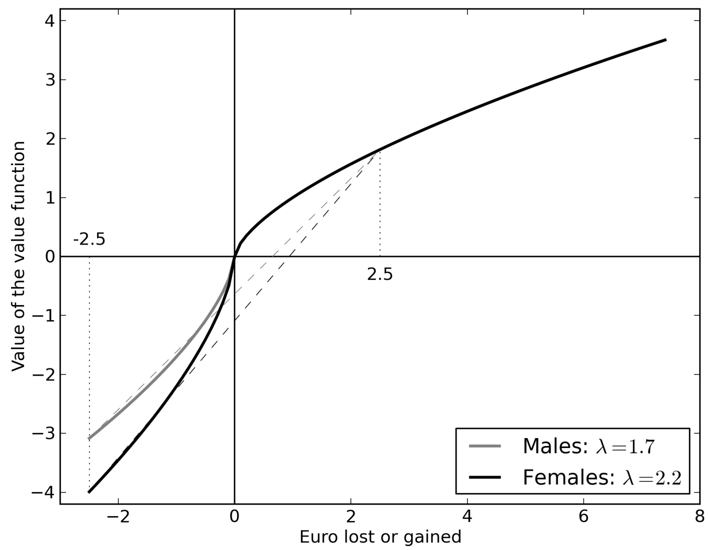


Figure 4: Two prospect Theory value functions, with the same r but different λ

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 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 entering the game 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. If instead they perceive the 2.5 euro as windfall gains such an effect is less likely to emerge. Moreover, the fact that the endowment was automatically put at stake, and subjects had no way of keeping the 2.5 euro with certainty, is also likely to weaken the salience of losses in this treatment. Performing a real effort task prior to the awarding of the 2.5 euro would probably strengthen the results. Finally, it is to be noted that the change of frame does not alter the underlying basic characteristic of the BRET, i.e., that subjects cannot secure any positive amount of money with probability one. We relied on a simple change of frame, giving subjects an additional endowment, as we wanted to reduce to a minimum the modifications compared to the Baseline treatment.[15] In weighting comparability more than the salience of losses, we are identifying in the Loss Aversion treatment a lower bound of the effect of loss aversion; on the other hand, we can be more confident that the observed change in behavior can be ascribed to the loss frame.

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 tasks. 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).

Treatment	N	Mean	Mann-Whitney z	Complexity	Mann-Whitney z
Baseline	269	46.54	$Prob > \|z\| = 0.86$	2.44	$Prob > \|z\| = 0.00$
Loss	135	46.31		3.36	

Table 4: Results and kernel density of decision by treatment, baseline *vs.* loss

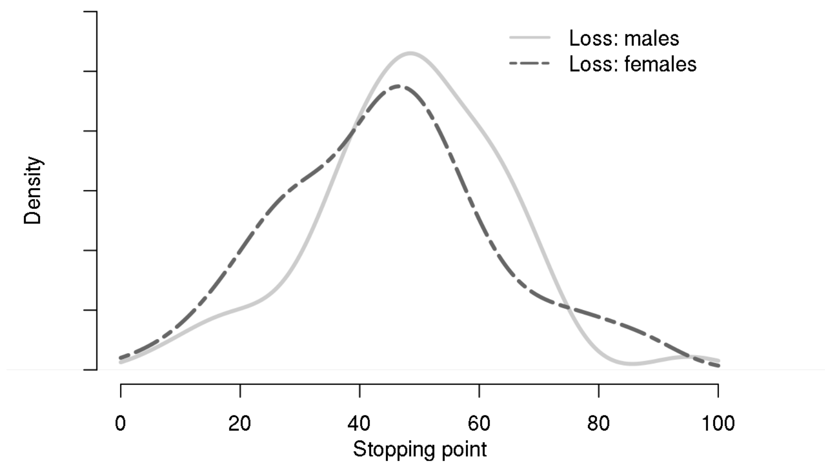
Decomposing the choices by gender (Fig. 5, upper table) confirms our hypothesis. The average choice of females turns out to be sizably lower than that of males, with the difference approaching traditional levels of significance, though not reaching them because of some outliers. Note that, at first glance, 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 session is characterized by puzzling results, with an even wider gap in the SOEP together with a huge reversed gender gap ($\hat{k}^*_m = 44.6, \hat{k}^*_f = 50.9$). 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 about 5%. Similarly, significance is also achieved by performing the sensitivity analysis (Fig. 5, lower table). As we kept manipulations at a minimum, we can say that results are not conclusive but definitely point in the right direction.

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 to identify 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 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 aver-

[15] There is evidence in the literature that even this weak manipulation of subjects' reference point is effective (Battalio et al., 1990; Harbaugh et al., 2010).

sion 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: similarly 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.



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

Figure 5: Stopping time and SOEP self-reported answers by gender: loss treatment

4.3. 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 game 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 game 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 game (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 k^* 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.

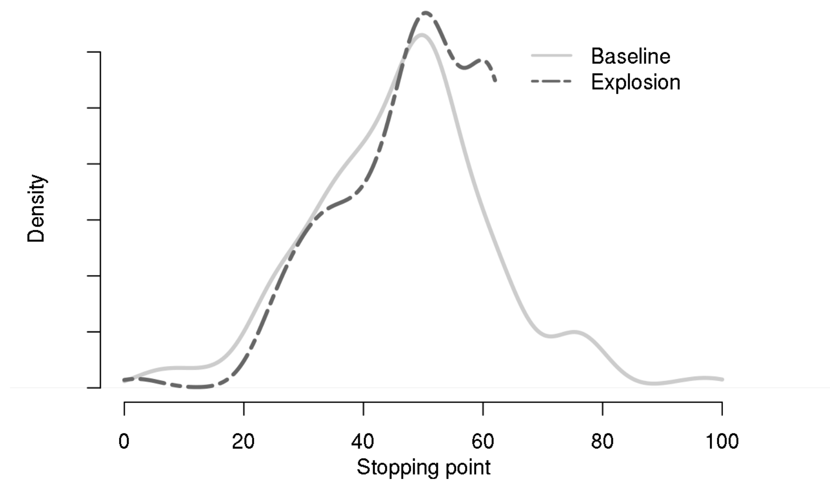
Assuming a generic CRRA utility function $u(x) = x^r$ the task allows to estimate the coefficient of risk aversion solving by numerical computation the following equation

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

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 $k^*_{expl} > k^*_{base}$.

Results show that this is not the case. In fact, as reported in Figure 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.[16] 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

[16] 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	

Figure 6: Results and kernel density of decision by treatment, Baseline *vs.* Explosion

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 a less conservative behaviour.

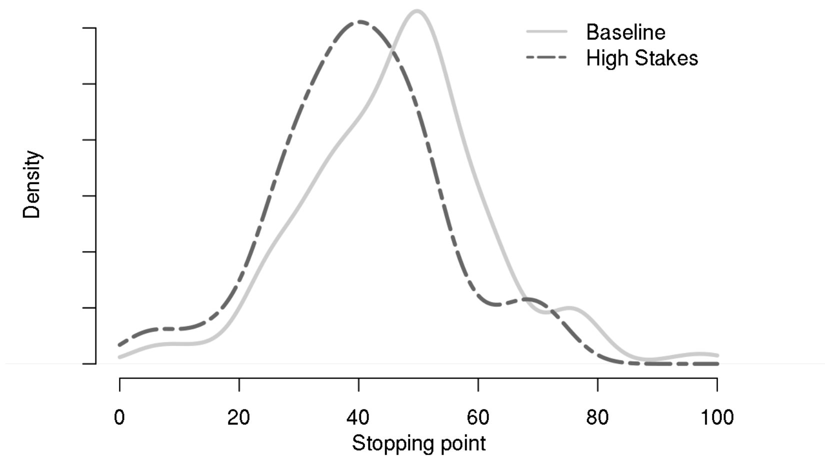
4.4. High Stakes

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 Figure 7, 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.

4.5. 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 to estimate 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*



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

Figure 7: Results and kernel density of decision by treatment, Baseline *vs.* High Stake

paribus. Moreover, manipulating the size of the square can reduce the salience of eventual focal points at $k = 50$.

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. 8, 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 game 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.

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

$$k^* = N\frac{r}{1+r}, \tag{4}$$

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.

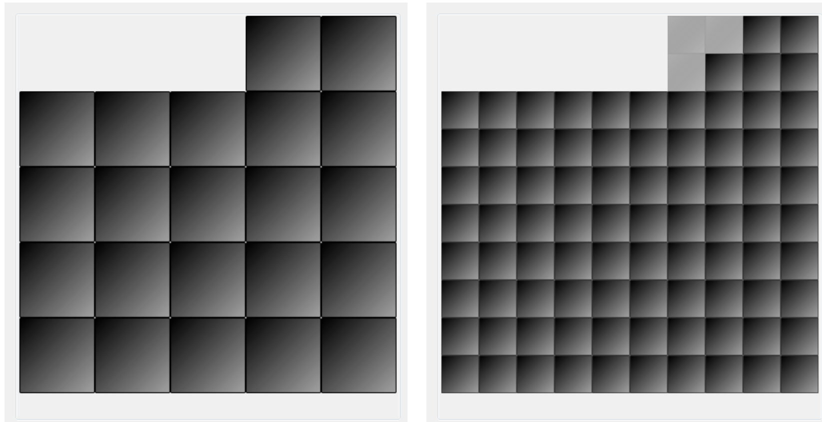


Figure 8: Screenshot after 15 seconds in the 5×5 (left) and Mixed 5×5 (right)

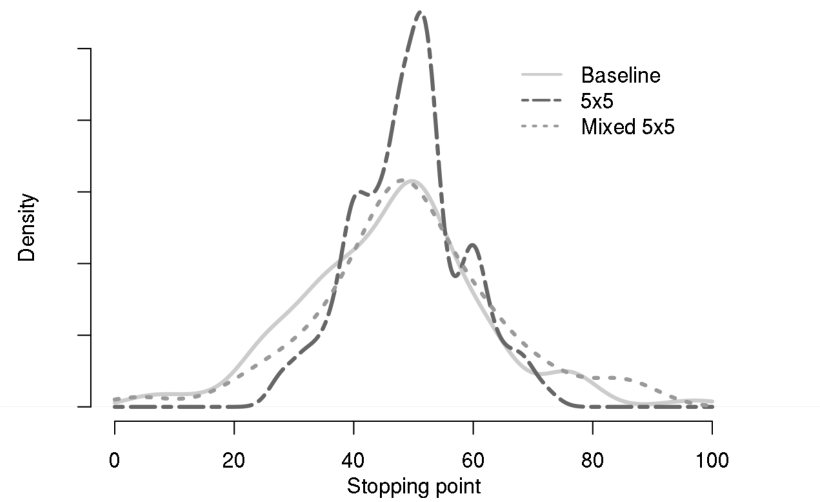
The theoretical prediction is therefore that we should observe $k^*_{20\times20} \geq k^*_{10\times10} \geq k^*_{5\times5}$ with inequality signs driven only by rounding problems, which should be more relevant between 5×5 and 10×10 than between 20×20 and 10×10.

Results are presented in Figure 9 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^*.[17] When analyzing the quantiles, the distribution looks similar to the Baseline except for a much higher concentration around the mean: 52.1% of 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. 8, 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. Therefore, in line with results already known in the literature, subjects seem to tolerate risk more, i.e., to display a lower risk aversion, when their choice is less diluted. In other words, it seems that facing a 4 times bigger probability of a bad outcome once is less painful than facing a proportionally lower probability 4 times. Such a finding would be consistent with a sort of anxiety induced by the mere participation in a lottery, at least within some interval (in our case, there is a significant difference between 25 and 100,

[17] 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.



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

Figure 9: Results and kernel density of decision by treatment, Baseline *vs.* Fast

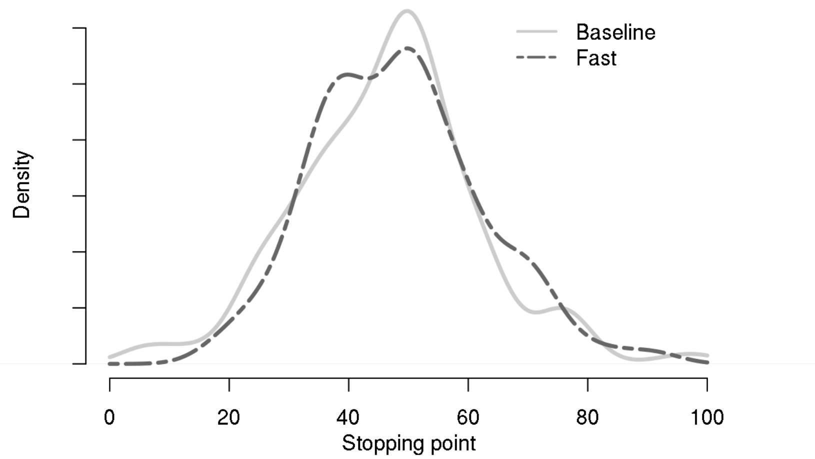
but not between 100 and 400 decisions). In any case, further research is needed to shed light on this counterintuitive result.

4.6. 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 of 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 Figure 10. 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.



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

Figure 10: Results and kernel density of decision by treatment, Baseline *vs.* Fast

4.7. Random Sequence

In the *Random* treatment, subjects face a 10×10 square in which the deletion process follows a random 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.

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,[18] we limited the number of observations to 32.

4.8. 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 payoff-irrelevant multiple decision framework.

After completing a Baseline treatment, in the *Repeated* treatment subjects played five further unannounced paying repetitions of the same task. Earnings were determined after com-

[18] 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.

Treatment	Trial	Average	Wilcoxon test $P > \|z\|$*
One-shot task		44.7	
Repeated	1	46.7	0.543
	2	50.8	0.056
	3	50.0	0.131
	4	46.9	0.802
	5	49.9	0.327
	Mean	48.8	0.063

*The Wilcoxon signed rank test refers to pairwise comparisons of each trial of the repetitions against the one-shot

Table 5: One-shot *vs.* repeated choices, Repeated treatment

pletion of both tasks extracting two positions for the bomb, one for the Baseline and another for one of the repetitions chosen at random.

Choices in the five repetitions are on average always higher than in the single task, significantly so in the second trial and almost significantly in the third trial of the repetitions (see Table 5). The average choice within repetitions exceeds by about 4 boxes the choice in the one-shot task, and despite the low number of observations, such a difference approaches traditional significance levels. This pattern emerges because 20% of the subjects repeat the same choice in all periods, while the others display some variance in the choices, which, on average, is significantly biased towards stopping later in the repeated task. Although we are not controlling for order effects, results seem to imply that subjects play the one-shot and the repeated tasks in a different way, displaying a lower risk aversion when the salience of the single decision is diluted by the repetitions. These subjects thus violate the Reduction Axiom, predicting that behavior should not systematically differ when the compounded lottery implied by the repetitions is played.

The evidence of the Repeated treatment therefore clearly indicates that it is sensible to restrict attention to the one-shot version of the task, which prevents the results from being affected by possible violations of the Reduction Axiom as well as by any kind of serial correlation and hedging, that are more likely to characterize a repeated framework.

4.9. 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 game 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 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. A full comparison of the Balloon Task with the BRET is beyond the scope of this paper and can be found in Crosetto and Filippin (2012). 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.

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 measure risk attitudes. Subjects are presented a game 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 to precisely estimate 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, risk neutral, or slightly risk loving.

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 and, to a lower extent, to the payoff irrelevant manipulation of the size of the choice set. On the other hand, results are sensitive both to wealth effects and repetition of the task, suggesting that when used as control, the BRET should be performed at the beginning of the experiment and one-shot.[19]

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,

[19] 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.

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 as much as possible the measurement error induced by extreme, though not necessarily dominated, choices 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 in choice under uncertainty, for all of which the results of this paper could act as a benchmark. For instance, exogenously inducing different reference points allows 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 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.

References

Battalio, R. C., Kagel, J. H., Jiranyakul, K., 1990. Testing between Alternative Models of Choice under Uncertainty: Some Initial Results. Journal of Risk and Uncertainty 3 (1), 25–50.

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. . W. 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., 01 2010. Portfolio Choice And Risk Attitudes: An Experiment. Economic Inquiry 48 (1), 133–146.

Charness, G., Gneezy, U., 2011. Strong Evidence for Gender Differences in Risk Taking. Journal of Economic Behavior & Organization (0), –.

Charness, G., Viceisza, A., 2011. Comprehension and risk elicitation in the field: Evidence from rural Senegal. IFPRI discussion papers 1135, International Food Policy Research Institute (IFPRI).

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 Comparison of Risk Elicitation Tasks. mimeo.

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. 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. Vol. 1 of Handbook of Experimental Economics Results. Elsevier, Ch. 113, pp. 1061–1073.

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., 2010. The SGG risk elicitation task: Implementation and results. ThE Papers 10/07, 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.

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), 263–91.

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? The 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 University Press.

Wakker, P., Deneffe, D., 1996. Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. Management Science 42 (8), 1131–1150.

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

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

For Online Publication

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.

Appendix 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.

Appendix 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, 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 the 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.

Appendix 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.

Appendix 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 th 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.