Safe options induce differences in risk attitudes

Paolo Crosetto, Antonio Filippin

Safe options induce gender differences in risk attitudes

Paolo Crosetto, Antonio Filippin

Working paper GAEL n° 05/2017

GAEL
Grenoble Applied Economic Laboratory
Consumption – Energy – Innovation

Safe options induce gender differences in risk attitudes

Paolo Crosetto

Antonio Filippin

May 2017

JEL : C81 ; C91 ; D81

https://gael.univ-grenoble-alpes.fr/

contact : agnes.vertier@inra.fr

Safe options induce gender differences in risk attitudes☆

Paolo Crosetto[a], Antonio Filippin[b,c]

[a] *GAEL, INRA, CNRS, Grenoble INP, Univ. Grenoble-Alpes, 38000, Grenoble, France*
[b] *University of Milan, Department of Economics, Via Conservatorio 7, 20122 Milano, Italy*
[c] *Institute for the Study of Labor (IZA), Schaumburg-Lippe-Str. 5-9, 53113 Bonn, Germany*

Abstract

Gender differences in risk attitudes are frequently observed, although recent literature has shown that they are context dependent rather than ubiquitous. In this paper we try to rationalize the heterogeneity of results investigating experimentally whether the presence of a safe option among the set of alternatives explains why females are more risk averse than males.

We manipulate three widely used risk elicitation methods finding that the availability of a safe option causally affects risk attitudes. The presence of a riskless alternative does not entirely explain the gender gap but it has a significant effect in triggering or magnifying (when already present) such differences.

Despite the pronounced instability that usually characterizes the measurement of risk preferences, we show estimating a structural model that the effect of a safe option is remarkably stable across tasks. This paper constitutes the first successful attempt to shed light on the determinants of gender differences in risk attitudes.

JEL Classifications: C81; C91; D81

Keywords: Gender differences, Risk attitudes, Experiment, Safe option

☆We are grateful to the Max Planck Institute of Economics (Jena) for financial and logistic support and to Eicke Hauck, Iolanda Viceconte, Claudia Zellmann, and Ivan Soraperra for their help in running the experiments. We would like to thank Maria De Paola, Frank Heinemann, Doron Sonsino, and participants to seminars at the University of Goettingen and at the University of Economics (Prague), as well as to participants to the 2013 CESifo workshop in Venice, the 8th Nordic Conference in Behavioral and Experimental Economics in Stockholm, the 2013 EUI Alumni Conference in Florence, and the 2014 IMEBESS conference in Oxford for useful suggestions. All remaining errors are ours.

Contact: `paolo.crosetto@inra.fr` (Paolo Crosetto), `antonio.filippin@unimi.it` (Antonio Filippin)

May 30, 2017

1. Introduction

A large body of literature in experimental economics and psychology reports gender differences in risk attitudes. Consistently with the evidence reported by some surveys (Byrnes et al., 1999; Croson and Gneezy, 2009; Eckel and Grossman, 2008b) a consensus has developed that females are more risk averse than males. Recent contributions have nonetheless found these findings to be less robust than previously thought (Nelson, 2016) and context-dependent. For instance, females tend to behave in a less risk averse way when not exposed to mixed-gender environments, be it in the context of female high schools (Booth and Nolen, 2012) or tutorials at the college level (Booth et al., 2014). Even within the more specific realm of choices over lotteries, the likelihood of observing gender differences in risk attitudes heavily depends on the elicitation method adopted, as recently shown by Filippin and Crosetto (2016). Gender differences are nearly always found using some tasks, whereas only seldom or never in others.

Strikingly, the literature in economics and psychology does not provide an explanation of why females are more risk averse than males. Up to our knowledge, the only attempt to rationalize the determinants of gender differences in risk attitudes can be found in the evolutionary literature. Evolution is a process in which preferences may adapt to the environment. Adaptation is slow and its consequences can last longer than the causes. As other phenomena observed nowadays such as different rates of labor force participation (Alesina et al., 2011), behavioral traits like risk attitudes may also have evolutionary roots. Trivers (1972) develops a theory of sexual selection based on parental investment and Dekel and Scotchmer (1999) extend the theory to risk aversion. The number of offspring a female is expected to have is nearly linear in the amount of resources available. Males, on the other hand, are characterized by a lower parental investment. Therefore, by competing over mates they can exploit a convex mapping from resources to reproductive success. Such differences might have driven the evolution of risk-seeking behavior in males but not in females.[1] Dreber and Hoffman (2010) present a survey of contributions supporting Trivers' theory using gender differences among a wide range of species. Even though their survey is suggestive of a possible extension to humans, Dreber and Hoffman (2010) acknowledge that the mechanisms triggering gender differences in humans has not yet been clearly identified. They suggest that males may be selected to fear the worst case scenario less, or to feel greater hope in anticipation of the best case scenario.

The evolutionary explanation is suggestive of the reason why differences emerged, but cannot account for the heterogeneity of the results documented in the laboratory. Filippin and Crosetto (2016) argue that two characteristics correlate with the likelihood of observing gender differences: a) the availability of a safe option among the set of alternatives, and b)

[1] Studies involving apes are used to address the origins of human behavior. Several contributions show that risk attitudes are indeed shaped by evolution, although without providing evidence along a gender perspective. These studies find for instance that bonobos are more risk averse than chimpanzees (Heilbronner et al., 2008; Rosati and Hare, 2013). The two species derive from a common ancestor but evolved differently, with chimpanzees developing a riskier foraging strategy. Differences in risk taking between the two species are nowadays observed within captive populations, i.e. among subjects that are fed and do not need to undertake any foraging strategy.

the presence of fixed $50\% - 50\%$ probabilities. Unfortunately, these features do not change independently across task. Moreover, although all risk elicitation tasks are ultimately built on choices among lotteries, they may differ in several other aspects such as the number of choices, their mathematical vs. visual representation, the interval of preferences that can be estimated, and so on. Hence, it is not possible to identify the role played by the different features of the task using the results already available in the literature, and an experimental investigation becomes necessary.

In this paper we test whether the availability of a riskless alternative, defined as a degenerate lottery in which a positive amount of money can be obtained with certainty, induces gender-specific behavior and can account for the heterogeneity of results observed. To isolate the role of riskless alternatives we carry out a set of controlled experiments involving 1085 subjects in which we add a safe option to (or remove from) three well-known and widely used tasks, keeping all else equal: a multiple price list (Holt and Laury, 2002), an ordered lottery selection (Eckel and Grossman, 2002), and the Bomb task (Crosetto and Filippin, 2013). Replicating the manipulation proves necessary since risk elicitation tasks vary in several dimensions and map choices into risk aversion parameters in different ways. Moreover, risk attitudes elicited with different tasks display a notoriously low consistency (Crosetto and Filippin, 2016; Deck et al., 2013; Isaac and James, 2000; Menkhoff and Sakha, 2017; Nielsen et al., 2013; Reynaud and Couture, 2012). We hence build new ad-hoc tasks in which safe options are carefully manipulated and we do so over several tasks.

To the best of our knowledge this paper is the first attempt to systematically analyze and provide causal evidence about a determinant of gender differences in risk attitudes.[2] We find evidence that the availability of a riskless alternative indeed makes gender differences more prominent. The effect is significant and consistent in sign and magnitude over the three different risk elicitation methods once results are compared using structural estimation. We cannot conclude that the presence of a safe option is the only factor affecting the emergence of a gender gap in risk elicitation. Nonetheless, safe options seem to play a consistent and robust role across different tasks. Even though the task employed do not allow us to test competing theories against each other, we also try to provide some hints about which one might (or not) rationalize the results observed.

The structure of the paper is as follows. We introduce the general structure of our experimental design in Section 2. Sections 3, 4, and 5 present in detail the baseline version of each task and the corresponding manipulation, and then test the presence of gender differences using both non-parametric, and, when supported by the data, parametric tests. Section 6 compares the results across risk elicitation methods using a structural estimation approach. Section 7 provides an overview of how different decision theories might account for our findings, and section 8 concludes.

[2]The careful reader may have noticed that the Allais paradox is per se a test of the role of a riskless alternative. The experimental literature using the Allais paradox, however, is not informative towards our research goal because usually results are not displayed by gender. The few exceptions are Petit et al. (2011) who find that females are more prone to the paradox and choose more often the safe alternative, and Da Silva et al. (2013) who do not detect gender differences.

2. Design

Risk attitudes are a latent construct, and as such can only be indirectly and imperfectly measured. Given the heterogeneous features of the different elicitation methods it is not surprising that measures of risk attitudes tend to be volatile across domains. Such an instability of results has been emphasized also along a gender dimension, thereby imposing a careful design to our experiment. In particular, we need to a) exogenously manipulate the presence of a safe option in a task *ceteris paribus*, i.e. keeping its structure unchanged; and b) replicate the exercise in more than one task, because the heterogeneity of results renders the generalization of results from a single elicitation method a questionable exercise.

The first question to answer is therefore which elicitation methods should be used. We believe that in order to be robust our findings must rely upon tasks delivering clear and different results along a gender perspective. Significant gender differences are a systematic finding in an ordered lottery choice task *à la* (Eckel and Grossman, 2002, henceforth: *EG*) and in the Investment Game by Gneezy and Potters (1997). In contrast, gender differences are rarely found and, when found, small in magnitude in the most widely used risk elicitation task, (Holt and Laury, 2002, henceforth: *HL*). Finally, the behavior of males and females is indistinguishable when preferences are elicited with the Bomb Risk Elicitation Task (Crosetto and Filippin, 2013, henceforth: *BRET*).

The likelihood of observing gender differences strongly correlates with the presence and focality of a safe alternative. *EG* and the Investment Game by Gneezy and Potters (1997) present a focal safe option in the form of a degenerate lottery yielding the same payoff irrespective of the random event. *HL* does not provide an explicit safe option but allows the subject to get a minimum payoff with probability 1. Finally, the *BRET* does not allow the subjects to earn any positive amount of money with probability one. The Investment Game and *EG* share many features besides the relationship between the availability of a riskless alternative and the gender gap. They both feature fixed $50\% - 50\%$ probabilities, they both cannot identify risk neutral and risk loving preferences, and in both gender differences are a nearly ubiquitous finding. Since in the Investment Game a safe option is present but virtually never chosen, and given the similarity of the two tasks, we decided to drop the Investment Game from consideration and focus on *EG*, *HL*, and the *BRET*.

We create new versions of each task either introducing (*HL* and *BRET*) or removing (*EG*) a safe option. Our aim is to reduce changes to a minimum, in order to preserve all the idiosyncratic characteristics of each task but still be able to causally identify the role played by the riskless alternative. Towards this goal we assume that agents are characterized by classic CRRA preferences:

$$U(x) = \frac{x^{1-\rho}}{1-\rho},$$

where ρ represents the coefficient of relative risk aversion. This assumption a) allows us to build a treatment version of each task that is isomorphic to the baseline condition under the null assumption that the safe option is irrelevant; b) helps to make results comparable across tasks.

Another source of heterogeneity in the results might stem from the repetition of the choice. It has been shown that (part of) the subjects make choices that are even negatively correlated over time (Isaac and James, 2000).[3] We hence opt for a pure between-subject experiment in which each subject participates in only one experimental condition. A grand total of 1085 subjects took part to our 6 conditions (3 tasks times 2 treatments design). The distribution of subjects by condition and the breakdown by gender are detailed in Table 1.

Task	N_{task}	**Version**	$N_{condition}$	**Males**	**Females**
Holt and Laury (2002)	344	Baseline: *HL*	179	84	95
		Treatment: *HLsafe*	165	79	86
Bomb Risk Elicitation Task	462	Baseline: *BRET*	271	106	165
		Treatment *BRETsafe*	191	73	118
Eckel and Grossman (2002)	279	Baseline: *EGsafe*	145	67	78
		Treatment: *EGnosafe*	134	57	77

Table 1: Distribution of the 1085 subjects by task and gender

2.1. *Experimental procedures and details*

The experimental sessions were run in 2014 (*BRETsafe*, *HLsafe*, *HL* extra sessions) at the Laboratory of the Max Planck Institute of Economics, and in 2016 (*EGnosafe* and *EGsafe* extra sessions) at the Laboratory of the Friedrich Schiller University, both in Jena, Germany.[4]

The experimental procedures were identical for all tasks.[5] Subjects entered the laboratory, instructions were both read aloud and available on screen. The English translation of the original instructions in German is available in Appendix A. Control questions about the experimental procedure and tasks were asked, and subjects were allowed to continue only after having replied correctly to all questions. Then the subjects faced the task, one shot.[6]

After all subjects had completed the task, they were exposed to a short questionnaire including demographics and a self-reported measure of the perceived complexity of the task. The randomization in the assignment to the six conditions should guarantee a balanced distribution of risk attitudes. However, in order to allow us to control for possible unbalances in mid-size samples like those we gathered, we exposed the subjects to the SOEP self-reported measure of attitude toward risk (introduced in the risk elicitation literature by Dohmen et al., 2011).

[3]Crosetto and Filippin (2013) show that a roller-coaster behavior is observed even when repeating the same task several times. Menkhoff and Sakha (2017) report that if the subjects fail to properly reduce the compound lottery generated by within-subjects designs, instability and inconsistencies are to be expected.

[4]The data of the baseline *BRET* are the same as in Crosetto and Filippin (2013). Part of the baseline *HL* and *EGsafe* data are the same as in Crosetto and Filippin (2016). However, given the focus on gender of this study, we needed to increase the overall sample size to support a gender comparisons. We hence planned three additional *HL* and *EGsafe* sessions. The treatment conditions are entirely original data.

[5]The custom experimental software for each task, written in Python, is available upon request.

[6]A trial run of the task was provided for the *BRET* and *BRETsafe* tasks. The presence of such a trial does not affect the results, see Crosetto and Filippin (2013).

3. Experiment 1: Multiple Price List

3.1. The task

Baseline condition: the classic Holt and Laury task (HL). The multiple price list format is a general procedure used to elicit values from a subject. Applied to risk, it consists of giving the subject an ordered list of binary choices between lotteries. The most widely known implementation hes been provided by Holt and Laury (2002), which is, to date, the most popular risk elicitation mechanism according to the number of citations.

In the *HL* task subjects face a series of choices between pairs of lotteries, ordered by increasing expected value. The set of possible outcomes is common to every choice, with one lottery safer (i.e., with lower variance) than the other. The increase in expected value across lottery pairs is obtained by increasing the probability of the 'good' event (see Table 2). At the end of the experiment, one row is randomly chosen for payment, and the chosen lottery is played to determine the payoff.

	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 €

Table 2: The classic Holt and Laury task

The subjects make a choice for each pair of lotteries, switching at some point from the safe to the risky option as the probability of the good outcome increases. The switching point captures their degree of risk aversion. For instance, a risk-neutral subject should start with Option A, and switch to B from the fifth choice on. Never choosing the risky option or switching from B to A are not infrequent and are regarded as inconsistent choices. They can be rationalized only adding as stochastic component.

In the *HL* condition subjects choose between lotteries characterized by uncertain outcomes with the exception of row 10 in which two sure amounts are compared. The lottery played at the end of the experiment is selected randomly. Therefore, there is no way subjects can avoid uncertainty, so that this condition can be used to build a pure measure of risk aversion in which certainty effects play no role.

Treatment condition: HLsafe. We introduce a safe option in the *HL* task by replacing Option A with a sure amount (see Table 3). Note that the amount changes across rows, in order to eliminate any difference in the fundamentals between the two treatments, except the availability of a safe choice. Every lottery proposed as Option A in Table 2 has been replaced with its certainty equivalent for an agent characterized by a CRRA utility function and a risk aversion parameter such that she would switch to Option B in that row. For instance, a subject should switch in the 6^{th} row if his risk aversion coefficient is ($\rho \in [0.15, 0.41]$). Hence, the safe amount in the sixth row (3.7) has been derived as the certainty equivalent of the lottery (4 € with $p = 0.6$; 3.2 € with $p = 0.4$) assuming ρ equal to the midpoint of the interval $[0.15, 0.41]$. Under the null assumption that the safe option does not matter, the two conditions *HL* and *HLsafe* are isomorphic.[7]

	Option A	Option B			
1	3.3 €	1/10	7.7 €	9/10	0.2 €
2	3.4 €	2/10	7.7 €	8/10	0.2€
3	3.5 €	3/10	7.7 €	7/10	0.2€
4	3.5 €	4/10	7.7 €	6/10	0.2€
5	3.6 €	5/10	7.7 €	5/10	0.2€
6	3.7 €	6/10	7.7 €	4/10	0.2€
7	3.7 €	7/10	7.7 €	3/10	0.2€
8	3.8 €	8/10	7.7 €	2/10	0.2€
9	3.9 €	9/10	7.7 €	1/10	0.2€
10	4 €	10/10	7.7 €	0/10	0.2€

Table 3: The HL task with a safe option

The goal of this treatment is to exogenously manipulate the presence of a safe option within a set of otherwise equivalent alternatives. Such a manipulation can be considered rather weak, however. In fact, the multiple price format is also likely to induce a comparison of risky alternatives across rows. Moreover, the necessity of maintaining a direct comparability across conditions imposes to substitute each Option A with a different safe amount. Both factors are likely to dilute the impact of the introduction of a safe choice in every row.

3.2. Results

In *HL* a higher degree of risk aversion induces the decision maker to choose Option A for a larger number of times. The switching point from Option A to Option B is therefore used to measure risk aversion. Unfortunately, it is not infrequent to observe that participants never choose Option B (i.e. they prefer 4 € to 7.7 €) or switch from Option B to Option A. Such patterns of decision are regarded as inconsistent because they cannot be rationalized by a deterministic expected utility maximizer. In our data such patterns of decision are displayed by 28 subjects (18 females, 10 males) in the Baseline *HL* and by 6 subjects (equally divided by gender) in *HLsafe*.[8] In this section we present the descriptive statistics and the basic results removing the inconsistent subjects. The behavior of multiple switchers can be analyzed in a structural model including a stochastic component (see Section 6 below).

[7] Other methods used to compute an amount row by row as similar as possible to the corresponding lottery in Treatment 1 would deliver virtually identical results. For instance, using the expected value of the lottery would deliver slightly different amounts in only two out of ten rows. In contrast, using the same amount in all the ten lotteries would change the underlying incentives across conditions.

[8] The fraction of inconsistent subjects in our data (15.6%) is in line with the literature, as reported by Filippin and Crosetto (2016).

Figure 1 displays the distribution of the switching point by gender and treatment. No clear pattern can be identified in the Baseline *HL* (left panel). In contrast, in the *HLsafe* treatment it clearly emerges that males are more represented than females in the risk seeking and risk neutral domains (switching points 3 to 5), while the opposite occurs in the risk aversion domain (switching points 6 to 7).

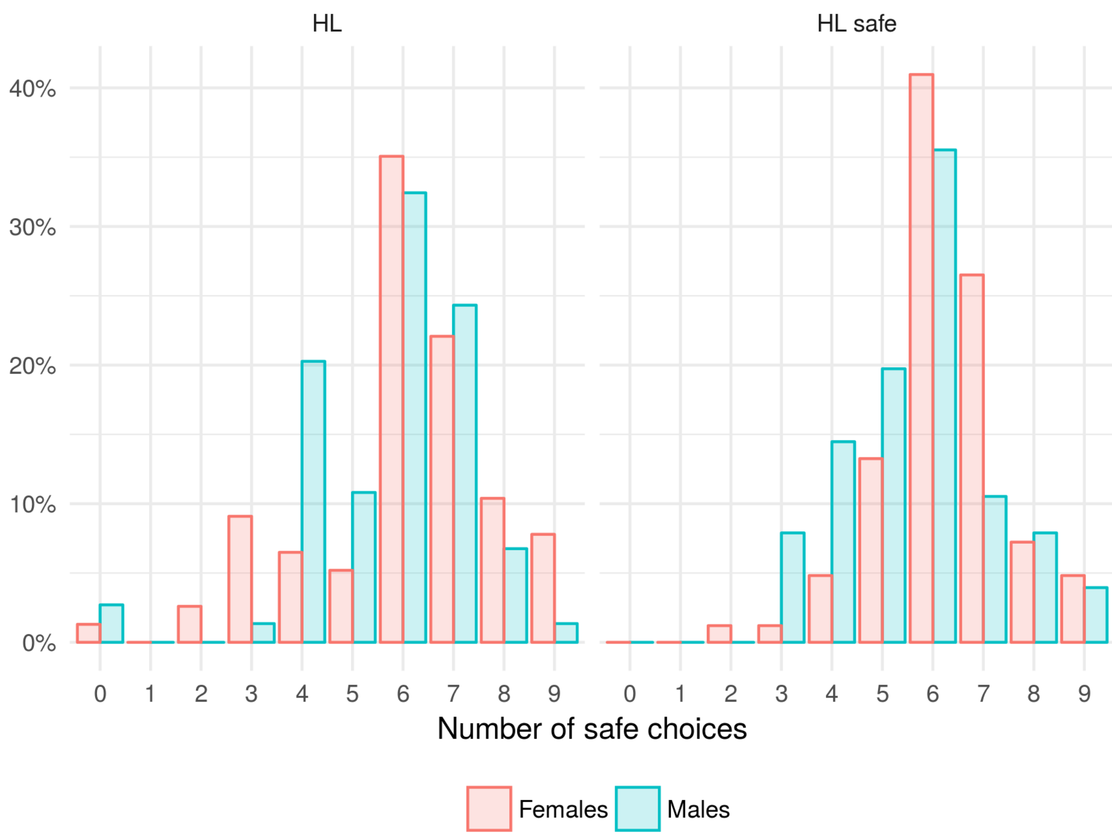

Figure 1: Distribution of the switching point by gender and treatment in the Holt and Laury task

The main advantage of the Holt and Laury task is that it has been built on the expected utility model. Therefore, the different switching points map almost linearly into coefficients of relative risk aversion (see Crosetto and Filippin, 2016, Figure 3). As a result, using the average choice to summarize the results does not introduce distortions in the analysis.

Table 4 shows that significant gender differences appear only when introducing a safe option. In the Baseline *HL* the average switching points of males and females do not significantly differ according to a Mann Whitney test ($p = .18$). This evidence is perfectly in line with the meta-analysis of Filippin and Crosetto (2016), showing that the point estimate typically differs by gender but does not reach traditional significance levels given the usual sample size of lab experiments. Strictly speaking, we cannot conclude that the two groups are characterized by the same degree of risk aversion because what we observe could be a false negative.

In the *HLsafe* treatment, the difference in the average switching point is instead highly

		N	Average # safe choices	Std. Dev.	Cohen's d	Mann Whitney	Norm. test^{o}	t-test*
Baseline: HL	Males	74	5.70	1.61	.19	.183	< 0.001	
	Females	77	6.03	1.84				
	Diff. (M-F)		*-0.33*					
Treatment: HLsafe	Males	76	5.66	1.47	.43	.004	0.636	0.008
	Females	83	6.24	1.26				
	Diff. (M-F)		*-0.58*					

^{o}skewness-kurtosis normality test; *one-tailed

Table 4: Average switching point by gender and treatment in the Holt and Laury task

significant ($p = .004$). We cannot reject normality for the *HLsafe* treatment, and we can thus run a t-test, that is also highly significant ($p = 0.008$). Furthermore, the magnitude of the difference is also larger, as captured by the Cohen's d (Cohen, 1988).[9] Hence, the comparison across treatments shows that the introduction of a riskless alternative causally affects gender differences in the Holt and Laury task.

4. Experiment 2: The Bomb Risk Elicitation Task (BRET)

4.1. The task

Baseline condition: the classic BRET. Our baseline condition uses the dynamic version of the *BRET*, a risk elicitation task introduced by Crosetto and Filippin (2013). Subjects face a 10×10 square in which each cell represents a box. 99 boxes are empty, while one contains a time bomb. Every second one box is automatically collected. (see Figure 2).

The subjects have to decide how many boxes to collect, i.e. $k^* \in [0, 100]$, by clicking the Stop button. 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 his earnings. In contrast, if $k_i^* < b$, subject i leaves the minefield without the bomb and receives 10 euro cents for every box collected.

The *BRET* interface provides a visual representation of probabilities that allows subjects to keep track of how many boxes have been collected and how many are left. Subjects' decision can be formalized as the choice of their favorite among the set of 101 lotteries fully described by the parameter $k \in [0, 100]$, which summarizes the trade-off between the amount of money that can be earned and the likelihood of obtaining it:

[9]Cohen's d is a measure of the size of an effect that is independent of the sample size. It is computed as

$$d = \frac{\bar{X}_f - \bar{X}_m}{\sigma},$$

where $\bar{X}_m$ and $\bar{X}_f$ are the average group choices and σ is the pooled standard deviation. Cohen (1988) indicates thresholds for interpreting his d: referring to aggregate differences, 0.2 should be considered a small effect, 0.5 a medium effect, and from 0.8 on a large effect.

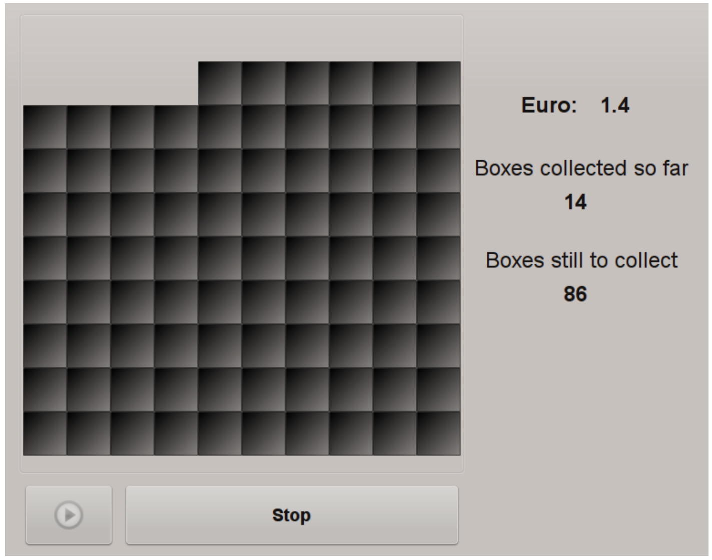

Figure 2: The BRET interface after 14 seconds

$$L_{BRET} = \begin{cases} 0 & \frac{k}{100} \\ \\ k & \frac{100-k}{100} \end{cases}$$

The degree of risk aversion negatively correlates with the choice of k and a risk-neutral subject should choose $k^* = 50$. The *BRET* does not provide safe options as the only amount that can be secured with certainty is zero, by chosing $k = 0$ or $k = 100$. Hence, the choice of k implies a comparison between uncertain amounts only and it can therefore be used to build a pure measure of risk aversion in which certainty effects play no role.

Treatment condition: BRETsafe. The baseline version of the *BRET* usually finds similar behavior across gender. If a safe option induces gender differences we should expect to observe them in the *BRETsafe* condition, in which a riskless alternative is made available by preventing the time bomb to be in the first 25 boxes. In other words, by choosing $k \leq 25$ subjects can secure a positive amount without incurring any risk. Figure 3 displays the graphical interface of the *BRETsafe* after 14 seconds.

For instance, by choosing $k = 20$ the subject earns for sure the value of 20 boxes (2 euro) because the time bomb can only be in $b \in [26, 100]$. In contrast, if the choice is $k = 40$, the underlying lottery implies earning either 4 euro with probability $(100 - 40)/75$ or nothing with probability one fifth ($(40 - 25)/75$). More generally, each lottery is then characterized by:

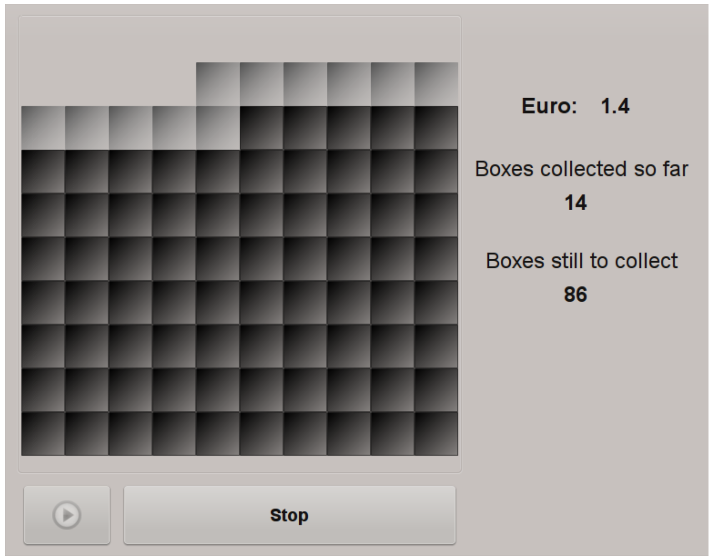

Figure 3: The Safe-BRET interface after 8 seconds

$$L^k = \begin{cases} \left[\; k \quad \text{with prob. } 1 \right. & \text{if } k \leq 25 \\ \left[\begin{array}{ll} 0 & \text{with prob. } \frac{k-25}{75} \\ k & \text{with prob. } \frac{100-k}{75} \end{array}\right. & \text{if } 25 \leq k \leq 100 \end{cases}$$

Note that for $k \geq 25$ the expected utility in the *BRETsafe* condition is a linear transformation of the Baseline under the reasonable assumption that $u(0) = 0$. Therefore, an expected utility maximizer should make the same choice in the two conditions as long as his optimal choice is $k \geq 25$. The only effect of the safe manipulation is that of inducing the more risk averse subjects to choose the highest safe option $k = 25$.[10] Any choice $k < 25$ violates the monotonicity assumption and would be irrational. As a result, we can expect to observe a slightly higher average choice in the *BRETsafe*. In any case, we are not interested in a point prediction of the average behavior across treatments, but only in the different effect that a safe alternative can induce along a gender dimension.

4.2. Results

As explained above, the number of boxes k captures the degree of risk aversion in the *BRET*. In this section we present the descriptive statistics and the basic results eliminating two subjects making dominated options - one stopping after one box and one collecting all 100 boxes, both in the Baseline *BRET* treatment.

Figure 4 shows a kernel density of the choices by gender and treatment. In the Baseline *BRET* (left panel) the two distributions nicely overlap with the exception that females tend

[10] Assuming a CRRA utility function only the subjects characterized by $\rho \geq .658$ should opt for the safe option.

to make more disperse choices. Looking at the *BRETsafe* (rigth panel) two things are immediately evident. First, female choose the safe option $k = 25$ relatively more often than males, although the difference in the frequency is not significant according to a Fisher exact test ($p = .542$). Second, the distribution of males is now shifted to the right as compared to that of females. In particular, males are now overrepresented in the risk loving domain.

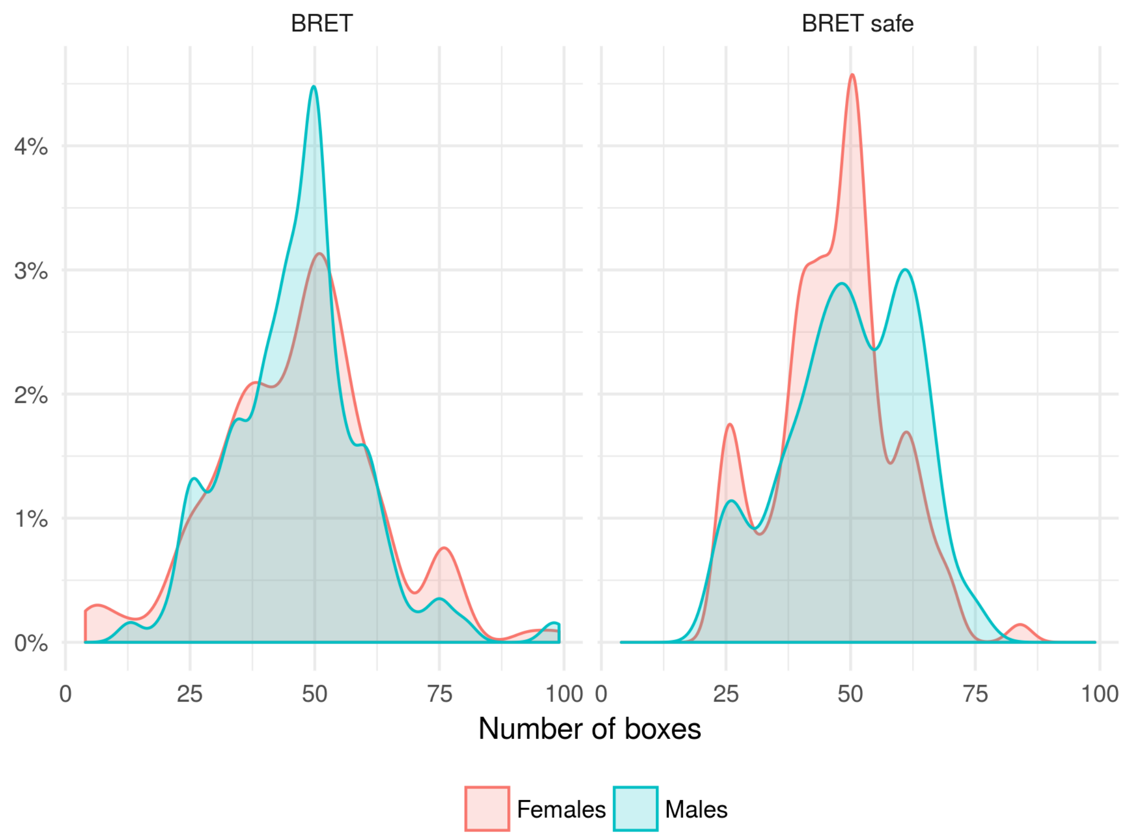

Figure 4: Kernel density (bandwidth adjustement 0.7) of the choices by gender and treatment in the BRET

Under a consequentialist approach like the one underlying the design of the treatment, it is rather surprising to observe that choices are affected over and above the likelihood of opting for the riskless alternative. In other words, there seems to be two effects. The first is a classic certainty effect that consists in opting for the riskless alternative. The second is an indirect effect occurring through choices not directly affected by the manipulation of the safe option. The two things trivially coincide using binary choices such as in the Allais paradox. However, in a more complex environment like those implemented in this paper things can be more subtle. The introduction of a safe option, even though not directly chosen, can make subjects more risk tolerant, inducing them to choose a more risky lottery than what would have been chosen otherwise.[11] Analyzing this kind of effects goes however beyond the goal

[11] The intuition is pretty similar to the Give vs. Take manipulation in Dictator Games. Bardsley (2008) and List

of this paper, as the focus of our research is to test whether females are more risk averse when a safe option is available. The comparison of the choices in the two experimental conditions indicates that this is the case also in the *BRETsafe*, at the same time suggesting that the indirect effect of the safe option (i.e., through choices other than the safe one) may even be prevalent.

		N	Average Choice	Std. Dev.	Cohen's d	Mann Whitney	Norm. test^o	t-test*
Baseline: BRET	Males	105	46.38	13.3	-.018	.675	.025	
	Females	164	46.65	16.3				
	Diff. (M-F)		-.27					
Treatment: BRETsafe	Males	73	49.79	12.7	.254	.079	.526	.045
	Females	118	46.72	11.8				
	Diff. (M-F)		3.07					

^{o}skewness-kurtosis normality test; *one-tailed

Table 5: Average number of boxes chosen by gender and treatment in the BRET

The *BRET* entails 101 possible choices and the kernel density above provides a partial picture because it necessarily smooths the actual distribution. Hence, in Table 5 we report the average choices, which confirm that introducing a safe option in the *BRET* generates significant and sizable gender differences.[12] In the baseline version of the task the behavior of males and females is indistinguishable, and differently from the *HL* case here the point estimate is virtually identical. In the *BRETsafe* females turn out to be relatively more risk averse than males according both to a Mann-Whitney test ($p = .079$) and also to a t-test ($p = .045$) performed as the normality of the distributions is not rejected. Moreover, the safe option induces differences that are considerable in size as captured by the Cohen's d.

5. Experiment 3: Ordered Lottery Selection

5.1. The task

Baseline condition: the classic Eckel and Grossman task (EGsafe). In ordered lottery selection tasks, subjects make a single choice picking one out of an ordered set of lotteries. This method has been first introduced in the literature to measure risk preferences by Binswanger (1981). A popular version is the one proposed by Eckel and Grossman (2002, 2008a), which has often been referred to in the literature about gender differences. In the *EGsafe* task subjects make their choice from a set of 5 lotteries characterized by a linearly increasing expected value as well as a larger and larger variance (see Table 6, panel (a)).

The risk-reward trade-off is induced by manipulating the outcomes of each lottery while keeping the probability of each outcome fixed at 50%. A risk-neutral subject should choose

(2007) show that the possibility of taking affects not only those whose choice was truncated by the lower bound of zero in the Give framework. In contrast, the whole distribution, including the counterparts of those who give a positive amount, shifts towards more selfish decisions once taking is a practicable alternative.

[12]The *BRET* is characterized by a hyperbolic relationship between k and ρ (see again Crosetto and Filippin, 2016, Figure 3). The analysis of the results using risk aversion coefficients is made in Section 6.

Panel (a): Original EG

	Event	Probability	Outcome
1	A	50%	4 €
	B	50%	4 €
2	A	50%	6 €
	B	50%	3 €
3	A	50%	8 €
	B	50%	2 €
4	A	50%	10 €
	B	50%	1 €
5	A	50%	12 €
	B	50%	0 €

Panel (b): EG nosafe

	Event	Probability	Outcome
1	A	50%	4.5 €
	B	50%	3.6 €
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 €

Table 6: Variations of the EG task used in the paper

lottery 5, as it yields the higher expected value, and the same holds both for a risk loving agent and for a slightly risk averse one. Increasing degrees of risk aversion induce choices with lower expected returns. Crucially towards our goal, the menu of choices includes a degenerate lottery that if chosen allows the subjects to secure a positive amount – in this case 4 Euro – without incurring any risk.

Treatment condition: EGnosafe. In case of the *EG* task the experimental manipulation is trivial and it simply amounts to proposing a version of the task in which the safe choice is replaced with an equivalent risky choice (see Table 6, panel (b)). Like in the case of the Holt and Laury task, the amounts have been chosen in such a way to be isomorphic to the original task for agents characterized by a CRRA utility function. In both condition the cutoff level of risk aversion that makes the agent switch from Lottery 1 to Lottery 2 is $\rho = 2$.

If the availability of a safe alternative triggers gender differences, we should observe that the behavior of males and females is more similar in this treatment than in the baseline condition. In case the safe option is the only determinant of gender differences their behavior should become indistinguishable.

5.2. Results

Both versions of the *EG* task entail a small number of alternatives, ordered from the safer to the riskiest, that are fully described in Figure 5. The Baseline *EGsafe* (right panel) displays a huge difference in the average choice of males and females, as usually found in the literature. Part of the gap derives from female choosing more often the safe option: 20.5% against only about 6%, a difference that is statistically significant according to a 1-sided Fisher exact test ($p = .010$).[13] In contrast, males are disproportionately more likely

[13]The reason why gender differences in the certainty effect are significant here but not in the *BRETsafe* may be due to the different salience of the safe option. In the *EGsafe* task the safe option is one out of only five alternatives, and moving from lottery 1 to lottery 2 makes a clear difference in terms of risk incurred. In contrast,

to choose the riskiest alternative (32.8% vs. 7.7%). Not surprisingly, the distribution of choices by gender turns out to be significantly different according to a Mann-Whitney test ($p = 0.001$).

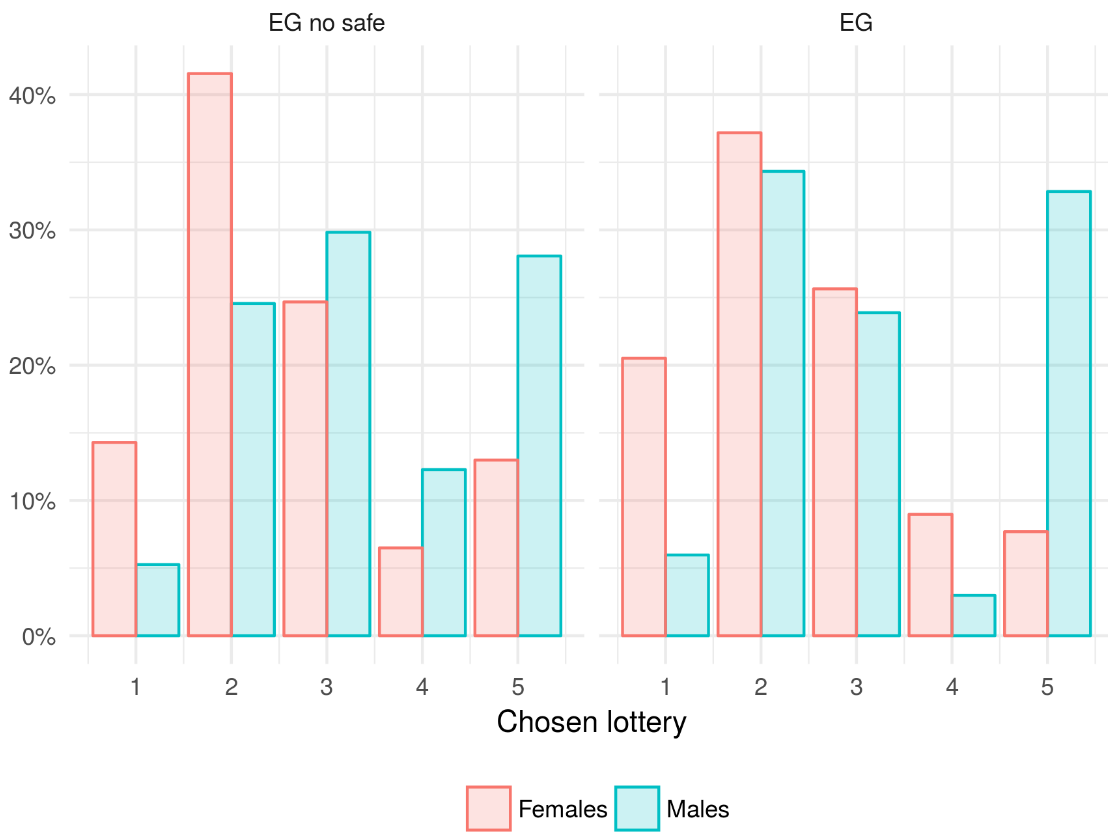

Figure 5: Distribution of lottery choices by gender and safe option in the EG task

Removing the riskless alternative in the *EGnosafe* treatment affects the distribution of the choices, without however changing the overall picture. The equivalent (according to EU theory) safer lottery in this case involves a small risk and is chosen by a lower fraction of females (14.3%). This fraction is still higher than that of males which stays about unchanged, though less significantly so according to a 1-sided Fisher exact test ($p = .078$). As already noted for the *BRETsafe*, the availability of a safe option may significantly affect the decisions even without immediately translating into a certainty effect. In the *EGnosafe* we can see that females are relatively more likely to choose lottery 5 than in the Baseline EG. However, in the *EGnosafe* gender differences are still evident, as confirmed by a Mann-Whitney test ($p = 0.001$).

In case of the *EG* task computing average choices is not a fruitful exercise for two reasons.

in the *BRET* the salience of the riskless alternative is likely diluted by the fact that the risk of the bad outcome increases at a very low rate (1.33% per additional box.)

The first is that the menu of lotteries is ordinal but has no cardinal meaning. The second is that unfortunately the mapping from choices to coefficients of relative risk aversion is highly non-linear (once more, see Crosetto and Filippin, 2016, Figure 3). Therefore, a quantitative assessment of the effect of the experimental manipulation is left to the structural model in the next section. Nevertheless, a bird's-eye view of the results is enough to note that removing the safe option does not eliminate gender differences. While the introduction of a riskless alternative induces gender differences in the elicitation methods where usually they are not observed (*HL* and *BRET*), the reverse pattern does not hold in the *EG* task.

6. Maximum Likelihood estimation

The evidence gathered with our experiments is rather informative. On the one hand, results in the *HLsafe* and *BRETsafe* conditions show that the availability of a safe option plays a significant role in inducing gender differences in risk attitudes. On the other hand, the findings of the *EGnosafe* treatment indicate that there are other determinants at work because females behave in a more risk averse manner also without the safe option. In this section we build an encompassing framework that allows us to compare the results obtained with the different tasks. We can thus provide a quantitative assessment of the role played by the availability of a riskless alternative across tasks. Towards this goal we have to tackle three problems.

The first issue is that the choices in the three tasks are not directly comparable. However, the fact that all tasks rely upon lotteries allows us to use coefficients of risk aversion as a common metric. The specific inner working of each elicitation method may affect the measurement of a latent construct such as risk attitudes. Nevertheless, we rely upon what all the different methods capture in common to estimate a structural model and to investigate whether the availability of a riskless alternative has a comparable impact across tasks.

In order to obtain comparable structural estimations across different tasks we need to make assumptions about the form of the utility function, error term, and the shape of the tasks themselves. We assume the standard CRRA utility function in which ρ represents the coefficient of relative risk aversion. We further assume that when subjects compare pairs of lotteries, they make an error ε. Following Hey and Orme (1994) we model this error as being normally distributed, $\varepsilon \sim N(0, \mu^2)$. By denoting the two lotteries as 'left' (L) and 'right' (R), the subjects choose R if and only if

$$\Delta EU = EU_R - EU_L > \varepsilon,$$

else they choose L. From the assumption of normality of the error term it follows that the probability of choosing R corresponds to the cumulative distribution function of a standard normal distribution

$$Pr(R) = \Phi\left(\frac{EU_R - EU_L}{\mu}\right).$$

The stochastic component allows us to analyze the behavior of the subjects switching from Option B to Option A in the *HL* conditions. We instead keep out of the sample the subjects making dominated choices, such as Option A in row 10 of *HL*, or choosing 100 boxes in the *BRET*.

To apply this method we need to reshape the tasks as providing a series of binary choices among lotteries. The *HL* task does not require any change, since it is built from the start on comparisons of pairs of lotteries. *EG* and the *BRET* need further assumptions and data transformation. Following Dave et al. (2010) and as done in Crosetto and Filippin (2016) we reshape the *EG* and the *BRET* as implicitly containing a series of choices among pairs of lotteries. In the *EG* task, we thus interpret a choice of lottery 3 as implying not only that 3 was preferred to all other lotteries, but also that 3 was preferred to 2, and 2 to 1. In other words, we assume preferences to be smooth and single-peaked. Similarly, in the *BRET*, we interpret a subject having chosen, say, 35 boxes as preferring 1 box to 0, then 2 to 1, then 3 to 2, and so on, up to the point where he preferred 35 boxes to 34, but not 36 to 35. The transformed data consists of 4 comparisons of pairs of lotteries for each *EG* subject, and 100 such comparisons for each *BRET* subject. To take into account the different probabilities and amounts at stake in the safe and non-safe versions of the tasks, we build new datasets for each condition separately.

The second issue is to exclude that the differences observed across experimental conditions may be due to confounding factors, such as probability weighting. In fact, subjects may maximize Expected Utility having a distorted perception of the objective probabilities and make different choices across experimental manipulations in a direction that correlates with the presence of the riskless alternative. The tasks used in our experiment are not ideal to estimate the shape of a probability weighting function, which in fact turns out to be rather different than that usually found in the literature. Even borrowing the functional form and the value of the parameters estimated by (Fehr-Duda et al., 2006), however, we wind that probability weighting could rationalize our results only in the *HLsafe* and only in part. The effect of probability weighting in the *BRETsafe* is instead counter-intuitive and characterized by a discontinuity in the choices not mirrored by the data. Finally, probability weighting would predict that females should be more likely to choose the safe option in the *EGsafe*, something that we indeed observe, but that the safe option should be chosen more frequently also by males, something contradicted by our evidence.[14]

The third issue to tackle is the possible heterogeneity of the subject pool across conditions. The randomization procedure allows us to make the assumption that groups are ex-ante equal, and the size of our treatment groups is relatively large for the standards in the laboratory. However, we can further control whether groups are balanced exploiting the self-reported degree of risk tolerance collected through the SOEP question. We want to check whether possible unbalances influence the estimate of gender differences across experimental conditions. In contrast, heterogeneity across treatments or tasks is less relevant toward our goal, as long as it characterizes males and females in a similar way.[15] Table 7 shows that females report to be consistently more risk averse than males across all our experimental conditions, with the exception of the *EGsafe* in which they are not statistically

[14]Using cumulative weights would instead not be an informative exercise. The reason is that the distortion of probabilities would imply a sort of pessimism or optimism (i.e. the better outcome perceived not as likely as the worse one), which could not be disentangled from risk aversion when dealing with binary lotteries (l'Haridon and Vieider, 2016).

[15]This is the case in the *HLsafe* and *BRETsafe*, where subjects seem to be more risk averse than their counterparts in the corresponding baseline condition.

different from males. We therefore decide to include the SOEP question in our estimates below.[16]

Task	Condition		SOEP mean	SOEP st.dev.	MW p-value
HL	Baseline	M	5.56	1.98	$< .001$
		F	4.55	2.02	
	HLsafe	M	5.08	2.26	.003
		F	4.1	1.87	
BRET	Baseline	M	5.36	1.95	.036
		F	4.84	2.06	
	BRETsafe	M	5.15	2.31	.013
		F	4.34	1.89	
EG	EGnosafe	M	5.39	2.29	.029
		F	4.57	2.02	
	EGsafe	M	5.39	2.06	.179
		F	4.95	2.19	

Table 7: Descriptive statistics of the SOEP risk question, by task, condition and gender

After making possible a direct comparison across experimental conditions, we estimate a structural model with maximum likelihood (ML) following Harrison and Rutström (2008). We allow the risk aversion parameter ρ to be affected by gender and by the answer to the self-reported risk-assessment SOEP question, and the noise parameter μ to be affected by gender and by the self-reported perceived complexity of the task.[17]

Results are reported in Table 8. First of all, our ML estimations confirm that the risk aversion coefficient (ρ) as well as the noise parameter (μ) vary widely across tasks. In addition, the constant (ρ) is comparable within the *HL* and *EG* tasks, meaning that the behavior of males does not differ across the safe and no-safe conditions. In contrast, the constant ρ tends to zero in the *BRETsafe*, reflecting that males move from risk aversion in the Baseline condition to almost risk neutrality in the *BRETsafe* treatment.

Focusing on our research question, the results in Table 8 confirm the effect of our experimental manipulation described in the summary statistics of sections 3-5 above. In two cases – *HL* and *BRET* – the introduction of a riskless alternative is enough to trigger significant

[16]One could argue that including the SOEP question in the model is not a good idea because it conceptually coincides with what we aim to estimate, i.e. risk aversion. While this variable has been shown to correlate with risk preferences elicited in an incentivized manner, the variance explained is indeed very low (Dohmen et al., 2011). Hence, there is no risk of overcontrolling. In any case, the results obtained excluding the SOEP variable are very similar (available upon request).

[17]Like in the case of the SOEP variable, results (available upon request) are robust to the exclusion of the control for the self-reported complexity.

	(1) BRET	(2) BRETsafe	(3) HL	(4) HLsafe	(5) EGnosafe	(6) EGsafe
ρ	0.21***	0.049	0.72***	0.67***	0.45***	0.46***
	(0.06)	(0.07)	(0.1)	(0.08)	(0.09)	(0.08)
ρ_{female}	-0.024	0.13*	0.019	0.12**	0.12**	0.20***
	(0.07)	(0.08)	(0.07)	(0.06)	(0.06)	(0.05)
ρ_{SOEP}	-0.011	-0.0086	-0.049***	-0.044***	-0.028**	-0.030**
	(0.008)	(0.007)	(0.02)	(0.01)	(0.01)	(0.01)
μ	0.16***	0.33***	0.67***	0.60***	0.27***	0.36***
	(0.03)	(0.06)	(0.1)	(0.07)	(0.05)	(0.06)
μ_{female}	0.036	-0.12*	0.059	-0.13	0.027	0.098
	(0.05)	(0.07)	(0.1)	(0.09)	(0.08)	(0.09)
μ_{easy}	0.00065	-0.00035	0.059*	0.051*	0.029*	-0.012
	(0.004)	(0.006)	(0.04)	(0.03)	(0.02)	(0.02)
N	208	191	164	164	134	145
Log likelihood	-5274	-4132	-545	-425	-296	-321

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Maximum likelihood structural estimations, Fechner error specification

gender differences. In the *EG* case, both the safe and non-safe versions of the task feature a significant gender effect. However, the ML estimations allows us to derive additional insights. First, in the *EGnosafe* the size and significance of the ρ_{female} coefficient is lower than in *EGsafe*, thereby showing that eliminating the safe option has an effect also in this task. Second and foremost, a comparison across tasks shows that the availability of a riskless alternative causes an increase of the gender difference of about the same magnitude. In fact, we observe a shift of ρ of about 0.1 with respect to the same task implemented without a safe option (0.15 in the *BRET*, 0.1 in *HL*, and 0.08 in the *EG*). This rather stable effect of the safe option in explaining gender differences across task is particularly remarkable, given the fact that risk elicitation measures are very volatile and task-dependent, as already explained in the introduction to this paper.

The noise parameter μ does not show appreciable differences between the two different versions of each task, with the exception of the *BRETsafe* in which females display a weakly significant lower coefficient. In any case, females are never characterized by a significantly higher degree of noise than males, suggesting for instance that numeracy does not play a significant role in rationalizing gender differences in risk attitudes.

7. Discussion

Our experiment shows that the availability of a safe option causally induces females to behave in a less risk tolerant manner relatively to males. This result marks a significant leap forward in the literature, as it constitutes the first successful attempt of shedding some light on the determinants of gender differences in decisions under risk.

The next obvious question is why the availability of a safe option triggers or magnifies gender differences in risk attitudes. Our experiment has been designed with the goal of testing whether a riskless alternative does play a role or not, and this endeavor imposed the use of classic elicitation methods for two reasons. First, it was necessary to compare the results of our manipulation against a sufficiently established consensus. Second, the notorious inconsistency of risk measurements across tasks required the replication of the results across several elicitation methods. Unfortunately, these tasks are not equipped to disentangle the different determinants of the observed effect of the safe option. Identifying the ultimate cause requires to purposely build a menu of choices leading to sufficiently different predictions for different theories under reasonable values of the parameters (as done for risk and ambiguity by Hey and Orme, 1994; Hey and Pace, 2014). Nevertheless, in what follows we discuss some of the possible explanations in the light of our data with the hope of providing suggestive evidence that could lead to further interesting developments in this branch of the literature.

The genuine interpretation of the results of Table 8 implies that a safe option shocks the risk aversion parameter within an Expected Utility framework. However, the inclusion of a safe option may also induce a change in the whole decision making process. In this case, the risk aversion coefficient would simply capture the effects of factors that are out of the realm of Expected Utility theory.

The first obvious candidate is a certainty effect, defined as a disproportional preference for an outcome without risk. For instance, Andreoni and Sprenger (2012) describe qualitatively this effect as $u-v$ preferences featuring a discontinuity at certainty in a similar way as the $\beta-\delta$ model of quasi-hyperbolic discounting. In the light of our results, this explanation would mean that the possibility of removing any risk is disproportionately preferred by females more than males. The testable implication, i.e. females should choose the safe option more frequently, finds limited support in our data. We have seen above that, although this pattern is observed in the *BRETsafe* and the *EGsafe*, most of the effect is due to the change of behaviour among choices that involve some risk.

Another possible mechanism is suggested by Prospect Theory (Kahneman and Tversky, 1979). When a task includes a degenerate lottery in which a positive amount can be secured with certainty, subjects could use such a safe option as a reference point against which to compare the outcomes of the risky alternatives. In this case positive amounts, when lower than the safe alternative, would be perceived as losses. As long as women are characterized by a higher coefficient of loss aversion they should display a less risk tolerant behavior in decisions that feature a safe alternative.[18] Our results lend some support to this interpretation (see Table 9). In the *HLsafe* females display a significantly higher degree of loss aversion. In the *BRETsafe* the λ_{female} coefficient is also positive, but it does not reach traditional significance levels. Note that the Prospect Theory model cannot be meaningfully estimated in the *EGsafe* because it would only rationalize extreme choices (lottery 1 and 5).

Other explanations are instead inconsistent with our results. For instance, simulations with reasonable values of the parameters show that the safe-choice manipulation does not

[18]From this point of view the evidence in the literature is not conclusive. Booij and de Kuilen (2009); Brooks and Zank (2005); Schmidt and Traub (2002) find that females are characterized by a higher degree of loss aversion, while Gaechter et al. (2010) do not.

	(1) BRETsafe		(2) HLsafe	
ρ	-0.026	(0.05)	0.80***	(0.2)
ρ_{female}	0.046	(0.04)	0.10	(0.2)
ρ_{SOEP}	-0.013	(0.01)	-0.18***	(0.04)
λ	1.13***	(0.1)	1.85***	(0.1)
λ_{female}	0.19	(0.1)	0.45**	(0.2)
μ	0.39***	(0.04)	1.44***	(0.2)
μ_{female}	-0.041	(0.05)	-0.019	(0.2)
μ_{easy}	-0.00046	(0.01)	0.13*	(0.08)
Observations	191		164	
Log likelihood	-4101		-429	

St.err. in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: ML structural estimations: prospect theory with safe option as a reference point, Fechner error

induce significant changes in the choices within Salience (Bordalo et al., 2012) and Regret Theory (Loomes and Sugden, 1982).

Summarizing, the safe-option manipulation induces surprisingly consistent results across risk elicitation methods. However, once trying to map the effect of the safe option into popular models of decision under risk we encounter the typical heterogeneity that characterizes this branch of the literature. Loss aversion and certainty effects seem to matter to a different extent across elicitation methods. As already mentioned, providing a test of the different theories goes beyond the scope of this paper. What our results suggest are simply possible directions for future research. On the one hand, loss aversion indicates that women take less risks when there is actually something at stake and that can be lost. On the other hand, the certainty effect shows that women have a stronger inclination to avoid any risk, when possible.[19] Adopting an evolutionary perspective, it is possible that parental investment may have shaped gender differences not only through the linear *vs.* convex mapping from resources to reproductive success, as posited by Dekel and Scotchmer (1999). Other determinants, for instance pregnancy, may also have induced a more pronounced aversion to losses and a stronger tendency to avoid any risk when possible.

8. Conclusion

This paper contributes to the growing literature that studies gender differences in decisions under risk with two main goals. The first is to shed some light on the determinants of such differences. The second is to rationalize the heterogeneous results observed so far. In particular, we test experimentally the role played by the availability in the choiceset of

[19]An important implication of these findings is that gender difference in decisions under risk should not be directly attributed to risk aversion within an Expected Utility framework. Estimating parameters of a utility function using a task that contains a safe alternative, like for instance in Booth et al. (2014), may deliver biased estimates.

a riskless alternative, defined as a degenerate lottery in which a positive amount of money can be obtained with certainty.

We carefully manipulate three well-known and widely used risk elicitation methods, adding to or removing from the menu of choices a riskless alternative *ceteris paribus*. The results of our between-subject design provide evidence showing that a safe option plays indeed a significant role. When subjects face the original *HL* multiple price list featuring two lotteries, one safer and one riskier, the observed choices of males and females do not differ significantly. In contrast, when we manipulate exogenously the *HL* task by substituting each safer lottery with an equivalent (under EU) sure amount significant gender differences emerge. Similarly, gender differences appear in the *BRET* only when the original task is manipulated introducing the possibility of securing a positive amount. Our design also includes a task (EG) which entails a safe option in the commonly used version, and that we substitute in our treatment condition with an equivalent (under EU) lottery with small variance. In this case we observe that removing the riskless alternative does not cause the gender differences to disappear, thereby signalling that there are other determinants at work.

We then make the choices across tasks comparable mapping them onto CRRA risk-aversion coefficients and we estimate a structural model with maximum likelihood. The results confirm the non-parametric analysis, and the comparison across tasks shows that the availability of a riskless alternative induces gender difference of about the same magnitude. When a safe option is available females display an increase of the risk aversion coefficient of about 0.1 (0.15 in the *BRET*, 0.1 in *HL*, and 0.08 in the *EG*). This stable effect of the safe option is particularly remarkable given the fact that measures of risk preferences are usually highly volatile and task-dependent.

Our experiment shows that the availability of a safe option causally induces females to behave in a less risk tolerant manner relatively to males. This result *per se* marks a significant leap forward in the understanding of gender differences in decisions under risk, as it constitutes the first successful attempt of shedding some light on their determinants. Our results also provide a roadmap for future research in terms of models that are suitable to rationalize the effect of the safe option. A proper test of these models against each other requires to purposely build a menu of lotteries capable of identifying the different determinants, but our results suggest that loss aversion and certainty effects likely play a role. Females may take less risks when there is actually something at stake and that can be lost and they display a stronger inclination to avoid any risk when possible.

References

Alesina, A., Giuliano, P., Nunn, N., 2011. Fertility and the Plough. American Economic Review 101 (3), 499–503.

Andreoni, J., Sprenger, C., 2012. Risk preferences are not time preferences. American Economic Review 102 (7), 3357–76.

Bardsley, N., June 2008. Dictator game giving: altruism or artefact? Experimental Economics 11 (2), 122–133.

Binswanger, H. P., 1981. Attitudes Toward Risk: Theoretical Implications of an Experiment in Rural India. The Economic Journal 91 (364), pp. 867–890.

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.

Booth, A., Cardona-Sosa, L., Nolen, P., 2014. Gender differences in risk aversion: Do single-sex environments affect their development? Journal of Economic Behavior & Organization 99 (0), 126 – 154.

Booth, A. L., Nolen, P., 2012. Gender differences in risk behaviour: does nurture matter?*. The Economic Journal 122 (558), F56–F78.

Bordalo, P., Gennaioli, N., Shleifer, A., 2012. Salience Theory of Choice Under Risk. The Quarterly Journal of Economics 127 (3), 1243–1285.

Brooks, P., Zank, H., 2005. Loss Averse Behavior. Journal of Risk and Uncertainty 31 (3), 301–325.

Byrnes, J. P., Miller, D. C., Schafer, W. D., 1999. Gender differences in risk taking: A meta-analysis. Psychological bulletin 125 (3), 367.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. L. Erlbaum Associates.

Crosetto, P., Filippin, A., August 2013. The 'bomb' risk elicitation task. Journal of Risk and Uncertainty 47 (1), 31–65.

Crosetto, P., Filippin, A., 2016. A theoretical and experimental appraisal of four risk elicitation methods. Experimental Economics 19 (3), 613–641.

Croson, R., Gneezy, U., June 2009. Gender Differences in Preferences. Journal of Economic Literature 47 (2), 448–74.

Da Silva, S., Baldo, D., Matsushita, R., 2013. Biological correlates of the allais paradox. Applied Economics 45 (5), 555–568.

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.

Deck, C., Lee, J., Reyes, J. A., Rosen, C. C., 2013. A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. Journal of Economic Behavior & Organization 87, 1–24.

Dekel, E., Scotchmer, S., 1999. On the evolution of attitudes towards risk in winner-take-all games. Journal of Economic Theory 87 (1), 125–143.

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.

Dreber, A., Hoffman, M., 2010. Biological basis of sex differences in risk aversion and competitiveness. Tech. rep., Citeseer.

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.

Fehr-Duda, H., Gennaro, M. D., Schubert, R., 2006. Gender, financial risk, and probability weights. Theory and Decision 60 (2), 283–313.

Filippin, A., Crosetto, P., 2016. A reconsideration of gender differences in risk attitudes. Management Science 62 (11), 3138–3160.

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.

Gneezy, U., Potters, J., 1997. An Experiment on Risk Taking and Evaluation Periods. The Quarterly Journal of Economics 112 (2), 631–45.

Harrison, G. W., Rutström, E. E., 2008. Risk Aversion in the Laboratory. In: Cox, J. C., Harrison, G. W. (Eds.), Risk Aversion in Experiments. Vol. 12 of Research in Experimental Economics. Emerald Group Publishing Limited, pp. 41–196.

Heilbronner, S. R., Rosati, A. G., Stevens, J. R., Hare, B., Hauser, M. D., 2008. A fruit in the hand or two in the bush? divergent risk preferences in chimpanzees and bonobos. Biology Letters 4 (3), 246–249.

Hey, J. D., Orme, C., November 1994. Investigating Generalizations of Expected Utility Theory Using Experimental Data. Econometrica 62 (6), 1291–1326.

Hey, J. D., Pace, N., 2014. The explanatory and predictive power of non two-stage-probability theories of decision making under ambiguity. Journal of Risk and Uncertainty 49 (1), 1–29.
URL `http://dx.doi.org/10.1007/s11166-014-9198-8`

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

Isaac, R., James, D., 2000. Just who are you calling risk averse? Journal of Risk and Uncertainty 20 (2), 177–187.

Kahneman, D., Tversky, A., 1979. Prospect Theory: An Analysis of Decision under Risk. Econometrica 47 (2), 263–91.

l'Haridon, O., Vieider, F., Apr. 2016. All Over the Map: Heterogeneity of Risk Preferences across Individuals, Prospects, and Countries. Economics & Management Discussion Papers em-dp2016-04, Henley Business School, Reading University.

List, J. A., 2007. On the Interpretation of Giving in Dictator Games. Journal of Political Economy 115, 482–493.

Loomes, G., Sugden, R., 1982. Regret Theory: An Alternative Theory of Rational Choice under Uncertainty. Economic Journal 92 (368), 805–24.

Menkhoff, L., Sakha, S., 2017. Estimating risky behavior with multiple-item risk measures. Journal of Economic Psychology 59, 59 – 86.
URL `http://www.sciencedirect.com/science/article/pii/S0167487016304196`

Nelson, J. A., 2016. Not-so-strong evidence for gender differences in risk taking. Feminist Economics 22 (2), 114–142.

Nielsen, T., Keil, A., Zeller, M., 2013. Assessing farmers risk preferences and their determinants in a marginal upland area of vietnam: a comparison of multiple elicitation techniques. Agricultural Economics 44 (3), 255–273.

Petit, E., Tcherkassof, A., Gassmann, X., 2011. Anticipated regret and self-esteem in the Allais paradox. Cahiers du GREThA 2011-25, Groupe de Recherche en Economie Thorique et Applique.

Reynaud, A., Couture, S., 2012. Stability of risk preference measures: results from a field experiment on french farmers. Theory and Decision 73, 203–221.

Rosati, A. G., Hare, B., 05 2013. Chimpanzees and bonobos exhibit emotional responses to decision outcomes. PLOS ONE 8 (5), 1–14.

Schmidt, U., Traub, S., 2002. An Experimental Test of Loss Aversion. Journal of Risk and Uncertainty 25 (3), 233–49.

Trivers, R., 1972. Parental investment and sexual selection. Sexual Selection & the Descent of Man, Aldine de Gruyter, New York, 136–179.

Appendix A. Experimental instructions

HL and HLsafe

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.

Now, please look at Decision 1 at the top. Option A pays {*HL*: 4 euro if the throw of the ten sided die is 1, and it pays 3.2 euro if the throw is 2-10; *HLsafe*: 3.3 euro in any case}. 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.

BRET and BRETsafe

On a sheet of paper on your desk you see a square composed of 100 numbered boxes. Behind one of these boxes hides a mine; all the other 99 boxes are free from mines. You do not know where this mine lies. You only know that the mine can be in any place between {*BRET*: 1; *BRETsafe*: 26} and 100 with equal probability.

You earn 10 eurocents for every box that is collected. After you hit Start in the corresponding square on your screen, 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.

Such earnings are only potential, however, because behind one of these boxes hides the time bomb that destroys your earnings in case it is collected. You do not know where this time bomb lies. You only know that the time bomb {*BRET*: can be in any place between 1 and 100 with equal probability; *BRETsafe*: is not in the boxes from number 1 to 25, while it can be in any place between 26 and 100 with equal probability}. Moreover, even if you collect the time 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 {*BRET*: 100 tokens numbered from 1 to 100 ; *BRETsafe*: 75 tokens numbered from 26 to 100 }.

If you happen to have harvested the box where the mine is located i.e. if your chosen number is greater than or equal to the drawn number you will earn zero. If the mine is located in a box that you did not harvest i.e. if your chosen number is smaller than the drawn number you will earn in euro an amount equivalent to the number you have chosen divided by ten.

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

EG and EGnosafe

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.