Essays in Experimental Economics and the Improvement of Judgment and Decision Making

Essays in Experimental Economics and the Improvement of Judgment and Decision Making

A dissertation presented

by

Alexandra van Geen

to

The Department of Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachussets

April 2014

Advisor: Iris Bohnet Alexandra van Geen

Essays in Experimental Economics and the Improvement of Judgment and Decision Making

ABSTRACT

This dissertation presents essays on the relationship between judgment and decision making and public policy, with a focus on gender diversity. The gender difference in career advancement is the likely result both of decisions made on the supply side (i.e. female and male job candidates) as well as decisions on the demand side (i.e. evaluators). These essays explore the behavioral foundations of decision making processes on both sides, and also make recommendations on how to use these behavioral insights to improve decisions, as well as increase gender diversity.

The first essay, *Opening the Black Box of Gender Differences in Risk Taking*, experimentally tests the drivers behind increased risk aversion of females. The essay demonstrates that this gender difference is not driven by differences in probability weighting but by differences in the valuation of outcomes. In addition, in environments where probabilities and outcomes are not common knowledge, females expect a higher chance of a bad outcome to occur, and dislike the bad outcome more. Controlling for the expected probabilities and the valuation of outcomes eliminates the gender difference in risk preferences, except for monetary risks.

The second essay, *Risk in the Background: How Men and Women Respond,* experimentally tests the gender-specific effect of risk in the decision environment on subsequent risk taking. It demonstrates that females increase risk taking after an increase in income and in the presence of an unrealized risk with strictly positive

potential outcomes, but there are no such effects for males. Males increase risk taking after winning a lottery, while females do not.

The third essay, *When Performance Trumps Gender Bias*, proposes a new intervention to overcome gender bias in hiring: an "evaluation nudge" in which people are evaluated jointly rather than separately. It demonstrates that evaluators are more likely to base their decisions on individual performance in joint than in separate evaluation and on group stereotypes in separate than in joint evaluation, making the proposed intervention the profit-maximizing evaluation procedure.

Contents

Acknowledgments

I am deeply grateful for my thesis committee: Iris Bohnet, Max Bazerman and Richard Zeckhauser. They have guided, inspired and supported me in many invaluable ways. Their integrity, dedication, and intellectual curiosity will be a continuous inspiration for the rest of my career. Their insights have shaped my views and my work. Iris in particular encouraged me to work on issues related to gender diversity. I could have never imagined how much this would alter my perspective, and how much of a passion this would become.

I am indebted to many others for support and valuable feedback on parts of this thesis, including Alberto Abadie, Paola Cecchi, Maria Acevedo, Chris Avery, Mahnaz Islam, Muriel Niederle, Andrea Robbett, Andrej Svorencik and Tyler VanderWeele. Especially Johanna Mollerstrom, Kim Peijnenburg, Olga Shurchkov and Anat Bracha have been of great support during the final stages of my dissertation.

Thanks are due to the staff of the Harvard Decision Science Lab, in particular Eric Mattison and Gabe Mansur for their great assistance during the many experiments that I ran there. Financial support from the Harvard Women and Public Policy Program and the Harvard Program on Negotiation is gratefully acknowledged. Being part of the community of Women and Public Policy Program has made my experience at Harvard a richer one.

I thank the Harvard Kennedy School for providing such an inspiring and wonderful environment and Nicole Tateosian, the director of the Ph.D. program for answering all of my many questions.

Finally, I am grateful for all the amazing people in my life, especially Vivien van Geen, Maaike Bol, Hilde Herrema, Thijs Jansen, Megon Magdanz, Kim Peijnenburg, Ineke Thierauf, Suzan van Otterdijk and Marije van Roekel for all their love, support, and patience.

1. Executive Summary

Young boys and young girls continue to have very different professional futures lying ahead of them. Even in the modern societies that pride themselves on the legal enforcement of equal rights and opportunities, women on average make lower wages, are less likely to receive promotions, and are less likely to reach positions of power. This holds in politics and business, as well as in academia. To understand why women continue to have a smaller voice, and thus a smaller impact, in the shaping of our present and future, we have to consider both the demand and the supply side of labor.

On the supply side, females may behave differently than males do in ways that makes them less geared for professional success. There are many reasons for gender differences in behavior, including differences in preferences, opportunities, expectations, capacities or inclinations. On the demand side, females may be in lower demand for positions of power if employers expect females to have lower potential or if employers inherently prefer to not work with females. Demand and supply are not independent: the expectations and decisions of employers will be influenced by gender differences in behavior and performance, and will in turn also affect the decisions of potential employees.

In economics there is growing awareness that decision makers often do not succeed in making rational and optimal decisions. Human beings are affected by biases and heuristics, oftentimes without realizing this. This implies that decisions cannot always be trusted to be reflections of a successful utility maximization process. Biases and heuristics can keep the individual from making the decisions that maximize his or her welfare. From an economic perspective, this implies an inefficiency. These behavioral tendencies also have implications for gender diversity in employment. For example on the demand side, if employers have no desire to discriminate but do so unconsciously because they are driven by heuristics or biases beyond their control, this can keep employers from hiring the best candidate, as well as keeping the best candidate from developing her full potential. Similarly, on the supply side, if it is behavioral tendencies and not preferences that keeps females from choosing directions that would prove fruitful for them, this is an inefficiency.

This dissertation presents three essays concerned with the relationship between gender, bounded rationality, and decision making. All three essays use experimental methods to determine hurdles for attaining gender diversity. The first two essays take the supply side perspective and yield insights in what is driving gender differences in risk preferences. The focus of the third essay is on the demand side and proposes and tests an intervention based on behavioral insights to reduce the impact of biases and heuristics of employers in their evaluation of job candidates.

The first essay, *Opening the Black Box of Gender Differences in Risk Taking,* uses experiments with a large range of risk attitude elicitation procedures to establish the nature of gender differences in risk aversion. The essay demonstrates that females are more risk averse in most, but not all, contexts. The gender difference for risk taking in the financial context is especially pronounced. The gender difference in risk aversion is driven by risk aversion over the potential outcomes associated with a risk and not by a gender difference in the weighting of the objective probabilities associated with the outcomes. There are two additional sources for higher risk aversion of females in situations where outcomes and probabilities are not objectively known. In these instances decision makers have to rely on their own judgments and expectations to estimate the risk, and females are more pessimistic about the likelihood that the good outcome will occur, and expect higher disutility when the low outcome occurs. These gender differences in risk aversion cannot be explained by differences in cognitive capacities, expected future income, anxiety, or self-reported overconfidence and optimism.

The second essay, *Risk in the Background: How Men and Women Respond,* builds on the first essay and investigates gender differences in the effect of risk in the decision environment. In particular it studies how the presence of unrealized risk and realizations of risk affect subsequent risk preferences. The essay demonstrates that there are stark gender differences. Females increase risk taking after an increase in income and in the presence of an unrealized risk with strictly positive potential outcomes, but there are no such effects for males. Males increase risk taking after winning a lottery, while females do not. The essay illustrates that risk preferences as well as gender differences in risk preferences are not static. Taken as a whole, the results are consistent with those that argue for a testosterone mediated effect of winning on risk preferences; they are also in line with decision models in which higher baseline risk aversion implies increased income sensitivity.

The third essay, *When Performance Trumps Gender Bias: Joint Versus Separate Evaluation,* is joint work with Iris Bohnet and Max Bazerman. This essay fits in the strand of research that uses behavioral insights to design policies to improve decision making. It proposes a new intervention to overcome gender bias in hiring, promotion, and job assignments: an "evaluation nudge" in which candidates are evaluated jointly rather than separately regarding their future performance. This intervention is inspired by findings in behavioral decision research suggesting that people make more rational choices when examining products jointly rather than separately. The intervention succeeds in making evaluators more likely to base their decisions on individual performance than on group stereotypes, and enables employers to make profit-maximizing decisions.

2. Opening the Black Box of Gender Differences in Risk Taking

2.1. Introduction

Studies have found females to be more risk averse than males in a large set of contexts and elicitation procedures, but we know little about what is causing this difference. Risk attitude is known to affect many important economic decisions - including financial investments and occupation choice - and therefore has far reaching consequences, including on a person's expected earnings (Bonin et al., 2007) and his likelihood of financial ruin. It is important to understand whether the increased risk aversion of females, and the reduced risk aversion of males is an expression of behavioral tendencies, such as biases and heuristics, or if it is the result of underlying preferences. If behavioral tendencies are responsible, this may be a motivation for policy makers to step in. A mapping of the drivers behind this gender gap in risk preferences may then inform us about the appropriate tools to address this 'behavioral' risk aversion and its potential consequences, such as the under-representation of females in high paying occupations, or the increased likelihood of bankruptcy for male entrepreneurs (Agarwal et al., 2013).

In a world of known outcomes and probabilities there are three main sources of risk averse choices. Firstly, they can occur because of *concave utility over outcomes*. In the presence of diminishing utility over earnings - represented by a concave utility function - fair bets will be rejected. Risk averse people are willing to give up the potential of receiving a good outcome in order to prevent the bad outcome from happening. This behavior is conform with rational preferences, such as stipulated in Expected Utility Theory. We know from behavioral research that people do not just care about absolute changes in earnings, but also care how they relate to a reference point. For gains the concave utility function over outcomes appears to be a pretty accurate description. but for losses not so much. Research suggests that we have a strong aversion to losing and will accept a lower certainty equivalent in order to prevent the possibility of a state of loss (Kahneman and Tversky, 2000). For risks that involve the possibility to lose money as compared to the reference point, *loss aversion* can be an additional source of risk averse decisions. Thirdly, risk aversion can occur through the *weighting of probabilities*. Behavioral research suggests that people do not weight probabilities linearly and that probability weighting can be a driver behind risk averse choices. For example, if people overweight the probability associated with the bad outcome in a particular risk, this will result in more risk averse decisions. An environment with unknown outcomes and probabilities can give rise to an additional source of risk aversion. If decision makers do not know the objective outcomes and probabilities they have to rely on their *expectations*. These estimates can be biased, for example when people are too optimistic about the probability of the good outcome to occur, this can increase risk taking.

I employ experiments to examine gender differences in these drivers of risk aversion. Using incentivized lottery based risk attitude elicitation methods covering positive outcome, mixed outcome and negative outcome lotteries I test for gender differences in risk aversion over outcomes, risk aversion through probability weighting, as well as loss aversion. Subjects also participate in the Dospert task in order to test for an additional effect of expectations. This task features a decision environment where outcomes and probabilities are not experimentally imposed, allowing me to compare subjects' stated willingness to take a particular risk with their expectations of the potential outcomes and the associated probabilities. The task features risky decisions involving financial, ethical, health, ethical, gambling as well a social contexts, enabling me to test for gender specific context effects on risk preferences.

I find that females are generally more risk averse, except for risks in a social or ethical context. Risk aversion through outcomes and expectations appear to be the two responsible drivers. For lotteries involving strictly positive outcomes I find significant gender differences in risk aversion over outcomes, with females being more risk averse than males. I find no gender differences in risk aversion over outcomes for the other lotteries: Females are not less risk loving for lotteries involving the negative domain and do not display more loss aversion when evaluating mixed lotteries. There is also no gender difference in probability weighting. In the environments were outcomes and probabilities are not experimentally imposed and subjects have to rely on their expectations, females' increased pessimism reduces their risk taking. The monetary context is special both because the gender gap in risk preferences is most pronounced and because this gap cannot be fully explained by expectations or by risk aversion over outcomes (and not by risk aversion through probability weighting or loss aversion either). This residual gap in risk preferences is consistent with gender specific expressions of biases, heuristics, or process utility in the financial context.

I consider a number of explanations to understand what may give rise to these gender differences in risk preferences. I test for an effect of overconfidence and optimism since if females are less overconfident or optimistic this may negatively effect their expectations as well as their tolerance to risk. I also test for an effect of anxiety and an effect of cognitive reflection or proficiency with probabilities as these may have a similar effect on risk preferences. I consider an effect of income, as lower expected income of females may imply increased risk aversion in the presence of decreasing absolute risk aversion. Although I find significant differences in reported optimism, overconfidence, expected future income, as well as experimentally elicited cognitive reflection, proficiency with probabilities and anxiety, these traits cannot significantly explain the gender differences in risk aversion over outcomes or the gender differences in risk aversion through expectations.

This paper proceeds as follows. Section II discusses the conceptual framework. Section III describes the experimental design, Section IV reports the experimental results and Section V concludes.

2.2. Conceptual Framework

The economic theory on risk preferences has become an increasingly complicated one. Expected Utility Theory has an elegant explanation for risk aversion by allowing rational agents to have concave utility functions over wealth. Behavioral insights have forced theorists to develop alternative models that allow for non-linear probability weighting and for reference point dependence. As outlined in the introduction these behavioral tendencies can give rise to additional drivers of risk aversion. In this paper we consider 5 potential drivers of risk preference.

Driver 1: Females are more risk averse because of diminishing marginal utility over earnings. Risk preferences may differ because females have more concave utility functions over wealth or experimental earnings than males do. The concavity of the utility function plays an important role in many decision models including Expected Utility Theory, Expected Utility over Income, and is also consistent with Prospect

Theory's description valuation over gains.[1]

Driver 2: Females are more risk averse because they are more intolerant towards potential losses than males are. The ability to account for loss aversion is an important part of the appeal for Prospect Theory. The aversion to losses may cause people to accept lower certainty equivalents. When females are more loss averse than males this will result in reduced tolerance for mixed outcome risks.

Driver 3: Females are more risk averse because of the way they weight the probabilities associated with the outcomes of the risk. Behavioral research suggest that probabilities are not weighted linearly, in certain situations this may affect risk preferences. Probability weighting is therefore an additional potential driver of gender differences in risk preferences.

Lottery based risk attitude elicitation tasks allow us to experimentally identify the gender differences in risk aversion through all three channels, but this does require the inclusion of negative and mixed outcome lotteries. The inclusion of negative and mixed lotteries is difficult from an experimental perspective, as we generally do not want subjects to become worse off because of the experiment. Most research on gender difference in risk preferences has therefore focused on decisions involving lotteries with strictly positive outcomes, and is unable to determine the effects of probability weighting and loss aversion on risk preferences. These studies overwhelmingly find females to be more risk averse (for overviews see Eckel and Grossman, 2008; Croson and Gneezy, 2009) on the positive domain although this difference is less pronounced for Swedish children and for girls from single-sex schools

[1]Risk aversion through the valuation of outcomes refers to the second order risk preference over outcomes. Higher order risk attitudes - such as prudence, edginess and especially temperance- may also affect risk taking. (Noussair and Trautmann, 2011) study the effect of these higher order risk attitudes on risk taking and find that temperance predicts the riskiness of portfolio choices. Temperance refers to vulnerability to background risk and is implied by decreasing absolute risk aversion. The findings on gender differences in higher order risk preferences is not robust, with some studies reporting no gender differences in temperance (Deck and Schlesinger, 2012)and some reporting higher temperance for females (Ebert and Wiesen, 2010).

(Cárdenas et al., 2012; Nolen and Booth, 2012). There has been little research on gender differences in preferences for negative outcome and mixed outcome gambles, but in risks framed as losses results are inconclusive, with sometimes females being *less* risk averse than males (Brachinger et al., 2000; Eckel and Grossman, 2008; Powell and Ansic, 1997). Fehr Duda et al. (2006) is one of the few that examines gender differences in probability weighting. They study a large variety of risky gambles and overall they find no significant gender difference in the probability weighting function. They do find that women appear to be less sensitive to changes in probabilities and that females are more pessimistic in the positive domain, but there is no gender difference in the effect on probability weights.

Driver 4: Females are more risk averse because they have more pessimistic expectations than males. In environments where probabilities and outcomes are not common knowledge, subjects are to rely on their own expectations. In as far as these expectations differ across genders this may give rise to an additional source of gender differences in risk preferences. Males are generally found to be more overconfident and optimistic (Barber and Odean, 2001; Reuben et al., 2013; Gysler et al., 2002) and this may affect their willingness to accept risk. Lottery based risk attitude elicitation tasks are not suitable for testing the effects of expectations. Harris et al. (2006) use the Dospert task to measure the effect of expectations and probabilities on risk taking in four contexts (health, gambling, recreational and social) and find that expectations do indeed differs across genders, with females generally being pessimistic both in the expected outcomes as well as in probabilities. Gender differences in expectations and valuations do explain part of the gender gap in risk preferences, but do not explain all of it. It is possible that their particular approach (inquiring about expected probabilities using a 5 point likert-scale, rather than actual probabilities) did not allow them to capture the full effect of expectations. But there is also the possibility that this part of the gender difference in risk taking is not driven by differences in the perception of probabilities and outcomes.

Driver 5: Females are more risk averse because of reasons that are not related with the perception of the outcomes and likelihood of the risk. There are other factors that can affect risk preferences. For example, even when a particular risky behavior that is presented to us is objectively enjoyable, our norms or ethics may prevent us from taking risk. In a lot of cultures gambling is considered inappropriate, taking ethical risks may be unacceptable from a moral point of view. Alternatively, the process of risk taking in a domain may be enjoyable to us irrespective of the outcomes and the probabilities. We enjoy gambling partly for the process, some of us may enjoy actively investing for the same reason. These 'other' factors may represent norms, biases and heuristics, or process utility. These factors may differ across genders, for example if our beliefs on what constitutes acceptable behavior differs across genders, such as when investment is considered a stereotypically male behavior.

The next section will discuss how the experimental design enables me to determine the importance of each of these factors.

2.3. Experimental Design

160 students were recruited from the subject pool of the Harvard Decision Science Laboratory at Harvard University. 50% of subjects were male, about 2/3 were American, the average age was 21-22 years, and a bit more than half were Caucasian. The show up fee was $10. All subjects were identified by code numbers and remained anonymous. The experiments were programmed and conducted with the software Z-Tree Fischbacher (2007).

Subjects started with a risk attitude elicitation task that is an adaptation of the Dospert Survey that was developed and validated by Blais et al. (2002). In this task subjects reported the likelihood of partaking in risky activities in several contexts (i.e. financial, social, recreational, gambling and health). Similar to Harris et al. (2006) subjects also stated the expected probability for the good outcome to occur and scale their valuation of the good outcome and the bad outcome. Different from Harris et al. (2006) we did not use a Likert scale but asked subjects to report the actual expected probabilities, in addition we inquired about the perceived riskiness of the activity, and included risks with a financial and ethical context. This task enables us to estimate the effect of expectations *(Driver 4)* as well as the effect of drivers not related to expectations or valuations of outcomes (*Driver 5*). The instructions for the first task are reported in the Appendix.

After the Dospert Survey subjects participated in additional survey based tasks that informed us about several personal characteristics. Notably, subjects participated in the Generalized Anxiety Disorder Questionnaire-IV, a questionnaire used in clinical settings to diagnose generalized anxiety disorder (Cashman-McGrath et al., 2002). Thereafter they participated in a Cognitive Reflection Test (Frederick, 2005) and the Berlin Numeracy Test (Cokely et al., 2012) to measure cognitive reflection and risk literacy.[2] Instructions are in the Appendix.[3]

Next all subjects participated in the second risk attitude elicitation task to estimate the effect of risk aversion over outcomes (*Driver 1*), loss aversion (*Driver 2*) as well risk aversion through probability weighting (*Driver 3*). This lottery based elicitation method is similar to Abdellaoui et al. (2011) with the exception that I also include mixed-outcome and negative-outcome lotteries. Subjects could earn a

[2]These survey based tasks took about 45 minutes to complete and all subject received $15 for their time.

[3]At this point, subjects were assigned to one of 4 treatments and depending on the treatment they now receive either a certain sum of $2 (T1) , $30 (T2) , $0 (T3) or an uncertain sum represented by a lottery that yields either $2 or $30 both with 50% probability (T4) (for a more detailed description see van Geen (2014)). These treatments are not the focus of the present study, but throughout the analysis we will control for gender specific treatment effects.

maximum of $15 and lose a maximum of $15. The task comprises 15 lists of 7 decisions, each list being associated with one particular lottery, with a fixed probability for the high outcome and the low outcome to occur. In each of the 7 decisions, the lottery is contrasted with a different fixed sum. The lowest fixed sum equals the lowest possible lottery outcome, and the highest fixed sum equals the highest possible outcome. After subjects made their decisions, one of the 15x7 decisions is randomly selected for payment. (See the Appendix for a screenshot of one of the lottery lists in the lottery list task). This method is found to be more tractable and less sensitive to the bias that is associated with procedures that rely on comparing non-degenerate lotteries (Bosch-Domènech and Silvestre, 2012; Abdellaoui et al., 2011).

Thereafter all subjects participated in a hypothetical large stake gamble task, as in Anderson and Mellor (2009) featuring questions on the willingness to make a risky investment in the scenario of a recently acquired inheritance. Anderson and Mellor (2009) find risk attitudes derived through this method to be correlated both with real life risk taking and with risk attitudes elicited from incentivized low stake lottery based decision tasks. This task serves as an additional check for the empirical validity of our results in the lottery based risk elicitation task which -because of financial constraints- has relatively low stakes.

Next all subjects were informed about the selected decision in the second risk attitude elicitation task, the outcome of the associated lottery, and their payoffs and participated in an additional task.[4] Most important for us they filled out a short demographic questionnaire that included questions on expected future income, parental income, self-reported risk aversion, loss aversion, confidence, optimism and

[4] Subjects in T4 learned of the outcome of the lottery.

their gender. The total experiment took about 1.5 hours.

2.4. Results: Drivers of Gender Differences in Risk Preferences

2.4.1. Drivers of risk taking in environments where outcomes and probabilities are common knowledge

Tab. 2.1 depicts the risk preferences in the lottery list task by gender. Males are about 20% less risk averse in the positive domain ($F_{1,159} = 3.97, p < 0.05$) and about 16% less risk averse in the mixed domain, but this is not significant ($F_{1,159} = 1.53, p = 0.22$). The reduction in the gender difference in risk aversion for lotteries in the mixed domain implies that is there is no evidence for increased loss aversion (*Driver 2*) of females. There is no gender difference in the negative domain ($F_{1,159} = 1.08, p = 0.30$).

Table 2.1.: Risk Preferences in Lottery List Task: Nr of Safe Choices

	Male	Female
Positive domain	14.154***	18.308 ***
	(0.907)	(1.879)
Mixed domain	14.308***	17***
	(1.173)	(1.837)
Negative domain	21.077***	22.462***
	(0.944)	(0.944)

Number of choices for the safe option in the Lottery List Task by domain. Linear regression controlling for gender specific treatment effects. Robust standard errors in parentheses.*** Significance at the 1 percent level ** Significant at the 5 percent level. * Significant at the 10 percent level.

The design of the first risk attitude elicitation task allows us to further explore the origin of this gender difference in risk aversion. In Tab. 2.2 I first estimate a CRRA power utility function for the lotteries on the positive domain so that $u(x) = x^{(1-a)}/(1-a)$, whilst controlling for gender and treatment effects. In this function a reflects the coefficient of relative risk aversion, with $a > 0$ indicating risk aversion. The results are displayed in the first row of Tab. 2.2. The second row of Tab. 2.2 depicts the estimates of a power utility function for the lotteries on the negative domain with $u(x) = -(-x)^{(1-b)}/(1-b)$, for which $b > 0$ reflects risk lovingness in the negative domain. The results indicate that females are significantly more risk averse in the positive domain ($\chi^2(1) = 6.86, p = 0.01.$), and significantly less risk loving in the negative domain ($\chi^2(1) = 3.71, p = 0.05.$)

Table 2.2.: Risk Preferences in Lottery List Task: EUT Specification

	Male	Female
Positive domain (a)	0.368***	0.643***
	(0.076)	(0.073)
Negative domain (b)	0.656***	0.459***
	(0.065)	(0.079)

Maximum likelihood Estimates of baseline CRRA (a)on the positive domain and risk lovingness (b) on the negative domain. Controlling for treatment effects, gender and treatment interactions, Clustered standard errors in parentheses.*** Significance at the 1 percent level ** Significant at the 5 percent level. * Significant at the 10 percent level.

Next I allow for probabilistic risk aversion (*Driver 3*) and estimate a RDEU model. Results are reported in Tab. 2.3. The estimates of a and b can be interpreted as risk aversion and risk lovingness in the positive and negative domain respectively. w is the weighting of the probability that the low outcome occurs. In the positive domain the objective probability of the low outcome is 0.25, and in the negative domain the probability of the low (most negative) outcome is 0.75. The estimates of a in the first row of Tab. 2.3 show that in the positive domain females are significantly more risk averse over outcomes ($\chi^2(1) = 3.59, p = 0.06.$) In absolute terms the coefficient of relative risk aversion is 0.22 higher for females, implying a coefficient of relative risk aversion for females that is 187% of that of males. The estimates in the second row of Tab. 2.3 are indicative of risk aversion in probabilities in the positive domain: both genders overweight the probability that the low outcome occurs. In the negative domain there is again overweighting of the probability of the low outcome, suggesting that there is risk aversion in probabilities even when there is risk lovingness over outcomes. We find no gender difference in probability weighting in the positive domain ($\chi^2(1) = 0.22, p = 0.64.$). Across genders these estimates of risk aversion over outcomes and probabilities in the positive domain are similar to Abdellaoui et al. (2011)

In the negative domain the absolute difference for the coefficient of relative risk aversion is about 0.23 but this is not significant ($\chi^2(1) = 2.18, p = 0.14$) and there is again no significant gender difference in the weighting of probabilities ($\chi^2(1) = 0.01, p = 0.93.$)

Table 2.3.: Risk Preferences in Lottery List Task, RDEU Specification

	Male	Female
Positive domain		
a	0.251***	0.469***
	(0.083)	(0.081)
w	1.398***	1.545***
	(0.144)	(0.279)
Negative domain		
b	0.465***	0.240 ***
	(0.267)	(0.097)
w	1.702 ***	1.732***
	(0.234)	(0.214)

Maximum likelihood estimates of CRRA value function of RDEU modelon the positive domain and the negative domain. w is probability of thelow outcome (p=0.25 on the positive domain, p=0.75 on the negative domain.) Controlling for treatment effects, gender and treatment interactions, Clustered standard errors in parentheses. *** Significance at the 1 percent level.

2.4.2. Drivers of risk taking in environments where outcomes and probabilities are not common knowledge

Tab. 2.4 reports the gender differences in the risk preferences in the Dospert task. Across contexts we see that males tend to be more likely to engage in risky behavior and perceive given risks to be less risky. This gender difference is especially pronounced in the investment and gamble contexts where males are respectively 23% and 14% less risk averse. There is no gender difference in the social and ethical domains. Males tend to expect a higher probability for the high outcome to occur than females. Males dislike the bad outcome less than females do, but they enjoy the high outcome the same. This pattern applies to all contexts except for the ethical and social contexts where there is no significant gender difference in risk taking or perceived riskiness. These results are in line with the findings of Harris et al. (2006).

We next study the drivers of these gender difference in risk preferences in the Dospert task. In Columns 1 and 4 of Tab. 2.5 we control for the context specific effects. In columns 2 and 5 we additionally control for the perceived outcomes and the estimated probability of the good outcome to occur. We find that these are significantly related to risk preference for both genders. Columns 3 and 6 include the interactions of the probability and the expected utility associated with the outcome. We find significant interaction effects, consistent with an Expected Utility type model. In column 7 we estimate the model for both genders simultaneously, and obtain an absolute gender gap of 0.20. Once we control for the perceived outcome and probabilities in column 8 this gender gap is no longer significant. This implies that on the aggregate the perceived outcome and probabilities can fully explain the overall gender difference in risk preference. Context specific gender differences in

Table 2.4.: Dospert Questions by Decicion Context

	Social		Recreational		Gambling		Health		Ethical		Investment	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Prob to take risk	3.407	3.502	2.755**	2.412**	1.897***	1.412***	2.663**	2.375**	1.833	1.870	3.295***	2.774***
	(0.059)	(0.057)	(0.105)	(0.094)	(0.112)	(0.085)	(0.087)	(0.070)	(0.070)	(0.073)	(0.105)	(0.102)
Perceived riskiness	2.168	2.299	3.404**	3.678**	3.859***	4.220***	3.598**	3.658**	3.526	3.540	2.436**	2.683**
	(0.072)	(0.065)	(0.082)	(0.067)	(0.091)	(0.086)	(0.080)	(0.056)	(0.084)	(0.068)	(0.076)	(0.079)
Prob high outcome	0.445	0.454	0.547***	0.466***	0.336**	0.279**	0.419**	0.357**	0.431**	0.367**	0.596	0.560
	(0.014)	(0.014)	(0.019)	(0.020)	(0.124)	(0.017)	(0.022)	(0.017)	(0.021)	(0.019)	(0.017)	(0.018)
Utility high outcome	2.804*	2.960*	2.971	2.816	3.699	3.558	2.002**	1.796**	2.643	2.706	3.564	3.707
	(0.079)	(0.066)	(0.097)	(0.105)	(0.124)	(0.145)	(0.067)	(0.062)	(0.096)	(0.094)	(0.084)	(0.091)
Disutility low outcome	2.511***	2.809***	4.021***	4.297***	3.731***	4.082***	4.115***	4.459***	3.893	4.020	3.013***	3.509***
	(0.072)	(0.078)	(0.092)	(0.082)	(0.092)	(0.091)	(0.085)	(0.059)	(0.091)	(0.069)	(0.093)	(0.090)

Standard errors in parentheses. *** Significance of gender difference at the 1 percent level (Wilcoxon rank sum test) ** Significance at the 5 percent level. * Significance at the 10 percent level.

the effects of outcomes and probabilities are ruled out in context specific regressions (not reported). In column 9 we allow for gender specific effects of the probabilities or outcomes, and find none. This is in line with our earlier results for the lottery based risk attitude elicitation task that pointed at a lack of gender differences in probability weighting. In column 9 we additionally allow for residual context-dependent gender effects. We find that in the investment context males take more risk, thus even after controlling for a perceived outcomes and probabilities. These findings are suggestive of 'other' forces at work in this context (*Driver 5*). There may be gender differences in the familiarity effect when females are less familiar with the investment context, causing them to be less willing to invest irrespective of the perceived utility.

Table 2.5.: Risk Taking in Dospert Task

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Female	Female	Female	Male	Male	Male	Both	Both	Both
Social context	1.13***	0.31***	0.33***	0.74***	0.09	0.10	0.94***	0.19**	0.31***
	(0.076)	(0.086)	(0.085)	(0.093)	(0.105)	(0.106)	(0.062)	(0.069)	(0.086)
Recreation context	0.04	-0.38***	-0.40***	0.09	-0.33***	-0.34***	0.06	-0.36***	-0.38***
	(0.101)	(0.087)	(0.085)	(0.104)	(0.087)	(0.088)	(0.072)	(0.062)	(0.087)
Gamble context	-0.96***	-1.24***	-1.16***	-0.77***	-1.09***	-0.98***	-0.87***	-1.17***	-1.24***
	(0.097)	(0.107)	(0.109)	(0.126)	(0.130)	(0.126)	(0.079)	(0.083)	(0.107)
Ethics context	-0.50***	-0.77***	-0.75***	-0.83***	-1.03***	-1.01***	-0.66***	-0.91***	-0.77***
	(0.066)	(0.063)	(0.061)	(0.079)	(0.077)	(0.076)	(0.052)	(0.051)	(0.063)
Investment context	0.40***	-0.49***	-0.48***	0.63***	-0.20	-0.16	0.51***	-0.36***	-0.49***
	(0.114)	(0.103)	(0.101)	(0.136)	(0.140)	(0.141)	(0.088)	(0.088)	(0.103)
Value of high outcome		0.20***	0.08**		0.22***	0.05		0.21***	0.20***
		(0.023)	(0.029)		(0.029)	(0.040)		(0.018)	(0.023)
Value of low outcome		-0.16***	-0.06		-0.20***	-0.17***		-0.19***	-0.16***
		(0.028)	(0.044)		(0.034)	(0.046)		(0.022)	(0.028)
Prob of high outcome		1.73***	0.31		1.51***	0.38		1.62***	1.73***
		(0.137)	(0.473)		(0.124)	(0.357)		(0.094)	(0.137)
Prob of high outcome			0.27***			0.35***			
*value of high outcome			(0.054)			(0.085)			
Prob of high outcome			-0.18**			-0.06			
*value of low outcome			(0.091)			(0.066)			
Male							0.20**	0.06	0.32
							(0.071)	(0.060)	(0.240)
Prob of high outcome									-0.22
*male									(0.185)
Value high outcome*									0.02
male									(0.037)
Value Low outcome*									-0.04
Male									(0.044)
Social x male									-0.22*
									(0.135)
Recreation x male									0.05
									(0.123)
Gamble x male									0.16
									(0.168)
Ethics x male									-0.26**
									(0.099)
Invest x male									0.29*
									(0.174)
Constant	2.38***	2.12***	2.70***	2.66***	2.43***	2.91***	2.42***	2.26***	2.12***
	(0.070)	(0.167)	(0.329)	(0.087)	(0.174)	(0.245)	(0.062)	(0.127)	(0.166)
N	3280	3280	3280	3120	3120	3120	6400	6400	6400

Linear regressions. Dependent Variable: Likelihood to take risk. Robust (clustered) standard errors in brackets (Nr of subjects =1600). ***<0.001; **<0.05 *<0.01. The control is the health context.

2.4.3. Results: relating the risk attitude elicitation tasks

Next I contrast the risk preference elicitation tasks to see how the results compare. Beside the Dospert task and the Lottery List Elicitation Task I include additional risk elicitation methods: I include a large stake hypothetical inheritance task, and subjects self-reported risk aversion and loss aversion. Like Anderson and Mellor

(2009) I categorized the responses in the inheritance task in 8 categories, category 8 being most risk averse[5]. For the self-reports I use a 6 point scale, with 6 being extremely risk/loss averse. For both tasks we find females to be significantly more risk averse and they report to be more loss averse, details are in the Appendix.

Tab. 2.6 displays the correlation across elicitation tasks by gender. The risk attitudes in the Dospert questions tend to be significantly correlated across contexts for both genders. The correlation between the health and recreation contexts and the correlation between investment and gambling contexts being the highest. There are some notable gender differences in the relations across contexts. For males risk taking in the gambling context is not correlated to risk taking in the social context. Suggesting that for males these reflect distinct traits. Correlation between risk taking in the health and recreation context and the correlation between risk taking in the ethical and recreation contexts is much higher for males than for females. For females correlation with risk taking in the investment context is higher than that for males for all contexts, suggesting that for males risk taking in investment scenarios is not so much related to general risk preferences.

The relation between lottery domains are similar across genders. Risk preferences in the positive and mixed domains of the Lottery List Task are strongly positively correlated, whereas risk preferences in the positive and mixed domains are negatively correlated with risk preferences in the negative domain. A potential explanation is that if the willingness to accept risk in the negative domain is largely driven by diminishing disutility of losses which may be correlated with diminishing returns of gains.

[5]A small percentage of the subjects provided inconsistent answers and could therefore not be categorized. In the original data there were no males that replied “don’t know” to at least one of the questions, but there were several females that did. The exclusion of these observations may therefore bias our results. I therefore recategorized the data and included those subjects who were inconsistent or who answered ’dont know’ to at least one question in the lowest category that included the most risky gamble that they accepted.

Relating the different risk attitude elicitation methods yields some interesting insights in the nature of gender differences in risk preferences. Overall, risk preferences in monetary based risk attitude elicitation tasks (Dospert investment, LLT task, and the Inheritance task) are strongly correlated for males but not females. For females the risk taking in the positive domain lottery questions in the LLT is strongly correlated with risk taking in recreation contexts, for males it is strongly correlated with risk taking in the investment contexts. Risk taking in the negative domain of the LLT is negatively correlated to risk taking in investment contexts for males. This may be because of a positive correlation of diminishing disutility of losses and diminishing utility of gains. Risk taking in the LLT task in lotteries involving mixed outcomes is positively correlated with risk taking in investment contexts for males. Surprisingly risk taking in mixed lotteries of the LLT task *negatively* correlated with risk taking in investment contexts for females. It is possible that for females the losses in the mixed lotteries play a different role for preferences than they do for males, but this is pure speculation. The inheritance task is significantly correlated with the decisions in Dospert task involving the investment context for males but not females. This task is also significantly correlated with the Lottery List Task for males, again with a negative correlation with the negative domain of this task, suggesting that men who are more risk averse in lotteries involving positive outcomes tend to be more risk seeking in the negative domain. For females these tasks are not related, except for the negative domain of the lottery list task; more risk taking in this domain implies more risk taking in the inheritance task.

The last two rows show that self reported risk aversion is significantly correlated with risk taking in the recreation context for both genders, and that for males it is also correlated with the gambling context. Unexpectedly for females it is *negatively* correlated with risk taking in the investment context. Self reported loss aversion for males is positively correlated with risk aversion in the recreation context. For females self-reported loss aversion is positively and significantly correlated with risk aversion in the positive and in the mixed domain. If higher risk aversion and loss aversion go hand in hand for women this may explain the negative correlation of both the self reported risk aversion and the risk aversion in the mixed lotteries of the LLT task with risk taking in the investment context.In the Appendix I include regressions of risk taking in the Dospert task that control for the risk attitudes derived in the lottery list elicitation task and the inheritance task. In general they cannot significantly explain the risk attitudes in the Dospert task. The investment and the gambling contexts resemble the scenarios of the lottery elicitation tasks more closely, and for these contexts the elicited risk preferences do have predictive value, they do not however succeed in reducing the gender gap in risk preferences.

These results indicate that females and males respond quite differently to different risks. For men the investment context stands out, as they are much more willing to take risk in this context and their risk attitude in this context is less correlated with risk taking in other contexts. For males higher risk taking in the LLT task and the inheritance task is predictive of higher risk taking in the investment context, whereas for females this is not the case. It appears that the involvement of monetary rewards has a distinct effect on males and not females.

2.4.4. Results: relating personal characteristics to gender differences in risk preferences

We consider several variables that have been proposed as drivers for the gender difference in risk preference. In the Appendix we compare the means across genders for these characteristics. We find that females score significantly higher on the

Table 2.6.: Correlations of Risk Attitude Elicitation Procedures by Gender

		D. soc.	D. rec.	D. gamble	D. health	D. ethic	D. invest	LLT pos	LLT neg	LLT mix	Inh. task	Riskav.
Dos rec.	M	0.21*	1									
	F	0.23**	1									
Dos gamble	M	-0.01	0.17	1								
	F	0.21**	0.19*	1								
Dos health	M	0.23**	0.42***	0.22**	1							
	F	0.29***	0.27***	0.21**	1							
Dos ethical	M	0.16	0.30***	0.26**	0.52***	1						
	F	0.20*	0.19*	0.32***	0.58***	1						
Dos Invest	M	0.27**	0.14	0.36***	0.01	0.06	1					
	F	0.30***	0.15	0.43***	0.16	0.19*	1					
LLT positive	M	0.16	-0.01	0.14	-0.11	0.01	0.21*	1				
	F	-0.11	0.20*	0.15	-0.03	0.03	0	1				
LLT negative	M	-0.05	0.15	0.06	-0.01	0.07	-0.20*	-0.28***	1			
	F	-0.01	0.04	0.12	0.12	0.04	-0.02	-0.14	1			
LLT mixed	M	0.02	0.04	0.1	-0.04	0.07	0.25**	0.5***	-0.37***	1		
	F	-0.09	0.02	-0.06	0.09	-0.02	-0.24**	0.5***	-0.22**	1		
Inherit task	M	0.13	0.1	0.03	-0.09	-0.05	0.41***	0.25**	-0.4***	0.26**	1	
	F	-0.03	0.04	-0.05	0.09	-0.09	0.11	0	0.22**	-0.08	1	
Riskaverse	M	0.19	0.28***	0.27**	0.17	0.11	0.16	-0.01	0.09	0.07	-0.02	1
	F	-0.14	0.31***	-0.12	0.01	-0.04	-0.22**	0.09	0.01	0.19*	0.01	1
Lossaverse	M	0.13	0.2*	0.15	-0.11	-0.17	0.11	0.16	0.09	0.14	0.27**	0.44***
	F	-0.05	0.17	0.15	0.14	0.05	-0.13	0.21*	0.07	0.24**	-0.07	0.61***

Pairwise correlations. Measures are transformed so that for each task a higher score means more risk aversion. Dospert measures are scales of the likelihood to participate in a certain risky activity.LLT score refers to the number of risk averse choices involving positive, negative, and mixed lotteries in the lottery list task. Inheritance task score refers to the category of risk aversion derived for this task derived in Table 4. *< 0.1, **< 0.05, ***< 0.01

anxiety task, score lower on the cognitive reflection test (consistent with Frederick, 2005) as well as the risk numeracy test. Significantly fewer females expect high future income. Consistent with the findings of (Jacobsen et al., 2014) and (Barber and Odean, 2001) they also report to be less optimistic, consider themselves to be less lucky and report lower confidence.

Tab. 2.7 examines the effect of these individual characteristics on the gender gap in risk aversion in the lottery list risk attitude elicitation task. Column 1 displays the risk preferences in the positive domain. Cognitive reflection and risk numeracy both significantly reduce risk aversion, whereas optimism increases it, the other individual characteristics have no significant effect on risk taking in the positive domain. Notably, inclusion of the individual characteristics does not reduce the absolute gender difference in the coefficient of risk aversion, the difference remains 0.22. Column 2 depicts the risk preference in the negative domain with b referring to the amount of risk taking. Risk numeracy and optimism now have no significant effect but anxiety (GAD) and cognitive reflection both significantly reduces risk taking in the negative domain, (the latter is consistent with the finding of Frederick, 2005). Those with higher expected future income are significantly more willing to take risk in the negative domain. Allowing for individual characteristics again does not reduce the absolute gender difference in risk preferences, it remains close to 0.23. Controlling for the individual characteristics this gender difference is now significant ($\chi^2(1) = 6.82, p < 0.01$.)

Table 2.7.: Risk Preferences in Lottery List Task, RDEU Specification

	Positive domain	Negative domain
a/b		
Male	-0.2190*	0.278***
	(0.123)	(0.106)
GAD	-0.002	0.011**
	(0.006)	(0.005)
CRT	-0.074***	-0.111***
	(0.022)	(0.022)
Risk Numeracy	-0.048*	-0.026
	(0.027)	(0.022)
High expected Income	-0.060	0.117**
	(0.049)	(0.057)
Optimist	0.211***	-0.011
	(0.058)	(0.089)
Lucky	-0.013	-0.067
	(0.047)	(0.064)
Confidence	-0.016	-0.007
	(0.014)	(0.012)
Constant	0.676***	0.366***
	(0.138)	(0.099)
w		
Male	-0.051	-0.144
	(0.335)	(0.264)
Constant	1.439***	1.815***
	(0.296)	(0.146)
Genderxtreatment effects	yes	yes

Maximum likelihood estimates of baseline DARA preference with a CRRA value function with coefficient of relative risk aversion aon the positive domain and risk lovingness bon the negative domain. w refers to the weighting of the probability of the low outcome (p=0.25) on the positive domain and the low outcome (p=0.75) on the positive domain. Clustered standard errors in brackets. ***<0.001; **<0.05; *<0.01

We next see whether individual characteristics can explain part of the gender gap in risk preferences in the Dospert task. Results are in Tab. 2.8. In column 1 we control for gender, context, and the interactions between gender and the context of the Dospert question. The initial gap in risk preferences is estimated as 0.29 and we estimate additional gender specific effects for several contexts. In column 2 the inclusion of the the individual characteristics does not help explain the overall gender gap in risk preferences or the gender specific context effects. None of the individual characteristics significantly predicts risk taking in the Dospert task. In column 3 the inclusion of expected outcomes and probabilities does eliminate the gender gap, except for the ethics context (males take less risk) and the gambling context (males take more risks). In the Appendix we control for individual characteristics in the large stake hypothetical inheritance task. None of the individual attributes has a significant effect on risk aversion in this task, and also for this risk attitude elicitation task the gender gap in risk preferences is not affected by the inclusion of these variables.

Table 2.8.: Effect of Individual Characteristics in Dospert Task

	(1)	(2)	(3)
Male	0.29**	0.25**	0.08
	(0.111)	(0.111)	(0.102)
Social context	1.13***	1.13***	0.29***
	(0.076)	(0.076)	(0.081)
Recreation context	0.04	0.04	-0.38***
	(0.101)	(0.101)	(0.086)
Gambling context	-0.96***	-0.96***	-1.27***
	(0.097)	(0.097)	(0.101)
Ethical context	-0.50***	-0.50***	-0.79***
	(0.065)	(0.065)	(0.062)
Investment context	0.40***	0.40***	-0.50***
	(0.113)	(0.114)	(0.104)
Social context	-0.38**	-0.38**	-0.17
x male	(0.120)	(0.120)	(0.109)
Recreation context	0.05	0.05	0.05
x male	(0.145)	(0.145)	(0.121)
Gambling context	0.20	0.20	0.22
x male	(0.159)	(0.159)	(0.144)
Ethics context	-0.33**	-0.33**	-0.23**
x male	(0.102)	(0.102)	(0.094)
Investment context	0.23	0.23	0.32**
x male	(0.177)	(0.177)	(0.157)
GAD		-0.01	-0.01
		(0.007)	(0.006)
CRT		0.00	-0.02
		(0.045)	(0.037)
Risk Numeracy		0.04	-0.00
		(0.052)	(0.040)
High expected income		-0.08	-0.06
		(0.100)	(0.087)
Optimist		-0.02	0.04
		(0.104)	(0.083)
Lucky		0.06	-0.01
		(0.081)	(0.070)
Confidence		0.02	0.01
		(0.023)	(0.019)
Value of high outcome			0.21***
			(0.018)
Value of low outcome			-0.18***
			(0.023)
Prob of high outcome			1.63***
			(0.092)
Constant	2.38***	2.27***	2.27***
	(0.069)	(0.179)	(0.180)
N	6400	6400	6400

Robust (clustered) standard errors in parentheses. ***<0.001; **<0.05; *<0.01

2.5. Discussion and Conclusion

Females are more risk averse than males. In this paper we replicated this gender difference in risk preference using a large range of risk attitude elicitation methods, and a large number of contexts. The results indicate that in a world of known outcomes and probabilities this gender difference is driven by risk aversion over outcomes and not by differences in probability weighting or differences in loss aversion, and is thus not clearly the result of increased behavioral bias. In environments where outcomes and probabilities are not common knowledge, females' relative pessimism increases the gender gap in risk preferences. Future research will need to establish whether females are overly pessimistic and whether there is a potential for policy interventions to address these biased expectations. We controlled for a wide variety of individual characteristics that have been argued to be behind the gender gap in risk preferences but were not able to significantly reduce the gender difference.

The results indicate that the gender gap is especially pronounced in risky decisions involving monetary contexts. In addition, the monetary context stands out as the gap in gender differences in risk aversion for risks in this type of context cannot be explained by the expected value associated with the risk. It appears that other factors are driving risk preferences in this context. This is surprising as we intuitively expect monetary risk taking to most closely resemble the rational decision framework. It is left for future research to find out what factors are responsible. For example, if males are more familiar with the financial context, this may -due to the familiarity bias- cause increased appetite for investing, regardless of the perceived outcomes and probabilities. Similarly, males may feel more proficient in the financial domain than females. Perceived proficiency is known to be correlated with higher risk taking.

3. Risk in the Background: How Men and Women Respond

3.1. Introduction

When considering risks in isolation, lab and field studies find that females have higher baseline risk aversion; they are generally less inclined to accept a risky prospect than males (e.g., Croson and Gneezy, 2009; Eckel and Grossman, 2008). However, decisions to take on risk typically do not occur in isolation; rather they occur in the presence of a risky environment. Decision makers will have had previous exposure to risks of which they have since learned the outcome, and may simultaneously face risks of which the outcome is not yet revealed. For example, an investment banker, deciding whether to purchase $300 million of stocks may moments before have made the deal of his lifetime. An engineer considering to undergo a risky surgery may at the same time be burdened by whispers of a looming reorganization at her company. Even when the fortunate realization of the investment banker's previous investment does not objectively alter the odds of the stock investment, this 'realized risk' may still impact his willingness to buy. Similarly, even though the potential layoff does not affect her medical outlook, the presence of this 'unrealized risk' may still influence the engineer's likelihood to opt for surgery. In order to have a true grasp of the gender differences in risk attitudes, we will therefore need to allow for these interactions with the decision environment.

In this paper I use laboratory experiments to test whether there are gender differences in how the presence of - independent, exogenous and uninsurable- risk, also termed 'background risk,' and its realizations affect subsequent risk attitudes. The design also allows me to establish whether this sensitivity is driven by the associated (potential) income. Decision models such as Expected Utility Theory or Prospect Theory allow for an effect of realized and unrealized background risk on subsequent risk attitudes through the income associated with their (potential) outcomes. However, realized and unrealized risk may also impact risk attitudes through non-income channels. The experience of winning after the realization of risk may for example give rise to altered emotional or hormonal states, or affect expectations on the odds of subsequent risky prospects. Similarly, the experience of the stochastic element associated with an unrealized risk may create a state of suspense that directly affects risk attitudes.

I find that an increase in income increases risk taking in females, but does not increase risk taking in males. Positive-outcome unrealized risk also tends to increase risk taking in females but not in males, and this effect appears to be mostly driven by the income associated with the potential outcomes of the lottery. The size of the increase in risk taking is significant, and eliminates the initial gender difference in risk attitude. The income effect appears transient, and females eventually return to baseline risk preferences. The experience of winning after the realization of risk increases risk taking in males but does not have this effect for females.

The sensitivity to the potential income associated with the unrealized risk for females is consistent with the commonly used isoelastic utility function. For this function increased sensitivity to income is implied by higher baseline risk aversion.

Because females have higher baseline risk aversion, the isoelastic utility function implies that a background risk with strictly positive potential outcomes will reduce their risk aversion more. The temporary nature of the effect of income is consistent with decision models that feature reference point dependence, including Prospect Theory. The observed winning effect for males is consistent with the observations by Wall Street trader turned neuroscientist John Coates, who reports how past profitable trades appeared to increase risk taking in male Wall Street traders to an irrational and dangerously high magnitude, an effect he attributes to winning-induced surges in testosterone Coates (2012).

The findings of this study support a more integrated view on risk attitude. This paper joins the ranks of other studies that stress the importance of the riskiness of the decision environment: Cameron and Shah (2013) find that the experience of an earthquake or flood reduces risk taking in lottery-based elicitation tasks by about 50%. Nagel and Malmendier (2011) find that the generation who experienced the Great Depression has a lower willingness to take financial risk. Provincial level GDP growth variability and variable business income have also been found to reduce risk taking (Guiso and Paiella, 2008; Heaton and Lucas, 2000). All these studies suggest the need to further explore the context in which risky decisions are made.

This paper proceeds as follows. Section II discusses the ways through which realized and unrealized risk can affect risk attitudes. In the conceptual framework of Section III, I derive the testable conditions that enable me to estimate gender-specific effects. Section IV describes the experimental design, Section V reports the experimental results and Section VI concludes.

3.2. Theory and Evidence

3.2.1. Sensitivity to realized risk: the income or the experience?

The realization of a risk can affect the risk attitude towards subsequent, unrelated, risks because of the *income* associated with the realization of the risk, and also through the associated *experience*. Income effects feature heavily in decision models such as (Cumulative) Prospect Theory (CPT) and Expected Utility Theory (EUT.) In true EUT, the assumptions of total wealth integration and decreasing absolute risk aversion imply increased risk taking after increased income. However, experimentalists have increasingly moved away from this model to explain risk taking in laboratory studies. Rabin and Thaler (2001) observe that the assumption of total wealth integration dictates that subjects should be risk neutral over the small stakes that are typically used in lab experiments, and this assumption is therefore irreconcilable with the facts. Many studies have since rejected the assumption of (full) wealth integration (Heinemann, 2008; Bardsley et al., 2006). The evidence of lack of wealth integration is consistent with the concept of 'narrow framing' where people make investment decisions without taking into account their total portfolio. A version of EUT, in which subjects partake in narrow framing and integrate only the earnings that fall within the frame (e.g., their earlier experimental earnings), does allow for income effects on the restricted domain that the laboratory experiment provides.[1] In the remainder of this paper whenever I discuss the income effects of 'EUT', I refer to this narrow framing version of EUT. Although the CPT model does not feature asset integration, because of the slope of the value function, in-

[1]The assumption of integration of experimental income is susceptible to similar criticism as the assumption of asset integration in EUT. Any theory of decision making under risk that relies on diminishing marginal utility of wealth to explain risk attitude, regardless if they specify full, partial or no asset integration, is susceptible to a version of the Rabin critique (Cox et al., 2007; Cox and Sadiraj, 2008).

come associated with the realization of risk may still temporarily affect risk taking if people fail to update their reference points.

These decision models do not allow for a direct effect from the experience of the realization of risk. When the outcome of the realized risk is fortunate and the decision maker 'wins,' the experience of winning may directly impact risk preferences. For example, winning may create certain emotions that in turn affect risk preferences. Good mood has been shown to increase optimism, and stock markets do better on sunny days (e.g., Wright and Bower, 1992; Hirshleifer and Shumway, 2003). Hormones are also affected by winning: Winning increases testosterone in men and increases their risk taking (Apicella et al., 2014). Winning may also affect beliefs. For example, it may give rise to hot-hand beliefs: After a string of successes, individuals have been shown to believe they are on a winning streak and give subjective probabilities of winning that exceed the objective probabilities (Sundali et al., 2012). In contrast, winning may give rise to the gambler's fallacy and reduce the perceived likelihood of winning a subsequent, independent, risk.

Although I am not aware of incentivized and controlled experiments that feature both genders, there is some research that suggests increased risk taking after past wins. Traders at the Taiwan Stock Exchange take above average risks after gains in prior trading outcomes (Wang et al., 2010). In a study by Thaler and Johnson (1990), participants were presented with a list of hypothetical statements in the form "you have won/lost X." Following a hypothetical win, participants were more likely to take a gamble. In their study with male subjects, Apicella et al. (2014) find that risk taking increases after winning the rock-paper-scissors game.

3.2.2. Sensitivity to unrealized risk: the income, or the anticipation?

An independent and uninsurable 'background risk' can affect the willingness to take on subsequent risk in two ways. Firstly, in income based decision models the potential income associated with the possible outcomes of background risks may affect risk taking. Within the EUT model, the assumption of asset integration ensures that the potential income associated with the possible outcomes of the background risk affects subsequent risk taking (Gollier and Pratt, 1996; Quiggin, 2003). For *undesirable* background risks, utility functions that satisfy proper risk aversion (Pratt and Zeckhauser, 1987), standard risk aversion (Kimball, 1993) and the more general risk vulnerability (Gollier and Pratt, 1996) all imply increased (absolute) risk aversion, and these conditions are met by a quite plausible set of assumptions. When it comes to *desirable* risks, the effect on risk attitude is less straightforward as there are two opposite forces. The increased riskiness may increase risk aversion for future risk in risk-vulnerable subjects, but the increased utility may reduce it (Gollier, 2001). Reference-point-dependent models such as CPT can also allow for an effect on risk attitude. If an individual's reference points are not updated after the introduction of background risk, the potential income associated with the background risk can affect risk taking through the effect of the induced gain- or loss-domain.

In a behavioral model, a second potential source of reactivity to unrealized risk is the sensitivity to the "risk component" of the background risk. The state of anticipation implies a presence of suspense that may change reference points or induce cognitive load and emotions; all of these have been shown to affect risk attitudes Loewenstein et al. (2001); Deck and Jahedi (2013). The model of Koszegi and Rabin (2006) allows for stochastic reference point distributions. It represents the notion that the risk component of unrealized risk affects tolerance to subsequent risks through the adjustment of the referent of what constitutes 'normal' risk. Sprenger (2010) shows that this model can result in an endowment effect for risks where the exposure to background risks makes a person more tolerant towards subsequent risks

There have been few studies that feature true experimental tests of the effect of background risk on risk taking, and none have studied gender differences. Results of the studies are conflicting. The low-stake lab-based experiments of Lusk and Coble (2008), and the low-stake field experiment with collector coins by List et al. (2007), suggest a small reduction in risk taking in the presence of negative-outcome background risk and mean-reserving spreads. On the other hand, research on narrow bracketing finds that when selecting multiple risks, subjects appear to make decisions as if the risks are isolated (e.g., Kahneman and Lovallo, 1993; Rabin and Weizsäcker, 2009). Samuelson (1963) finds that offering multiple risks increases risk taking, a phenomenon he termed the "fallacy of large numbers." Sprenger (2010) also reports an increase in risk tolerance after endowing people with risks.

3.3. Conceptual Framework

3.3.1. Determining income effects for men and women

To determine the income effects for men and women, we will first derive the income effect associated with obtaining a risk-free sum. Next we will derive the income associated with obtaining a risky sum.

3.3.1.1. Income effects associated with a risk-free sum

We define the absolute risk aversion at initial referent point w as ***baseline risk aversion***, depending on the relevant referent w can refer to wealth or initial income or something else, but for simplicity we will refer to w as initial income. Decreasing absolute risk aversion (AR) implies a negative income effect after an increase in income by x because it causes a reduction from baseline risk aversion, i.e., $AR(w) < AR(x,w)$.

Definition. The ***income effect*** of change in income x on absolute risk aversion at w is defined as $IC(x,w) = AR(x,w) - AR(w)$.

Letting income changes b and c denote 2 potential realizations of a lottery, with b>c>0. Letting p the probability associated with the occurrence of b we can represent the risk by the lottery $L^*(X) = (p; b, c)$. Decreasing absolute risk aversion implies a larger reduction in risk aversion if outcome b is realized than if outcome c is realized. Therefore, the income effect of b is more negative, i.e., $IC(b,w) < IC(c,w)$.

In EUT with narrow framing b and c refer to the earlier-obtained income within the narrow frame. In models with reference-point dependence, b and c refer to the distance as compared to the reference point w. This inequality may not hold if people more crudely respond to the gain (or loss) frame that the income induces, because in that case there may not be level effects of income on risk attitude.

The realization of risk may come with a larger income effect for females because they have higher baseline risk aversion (i.e., $AR^f(w) > AR^m(w)$. To see this, note that if at any wealth level w agent i is more risk averse than decision maker j, it holds that $u_i(w) = g \circ u_j(w)$ for some concave function g. It can be derived that $AR^i(w) = AR^j(w) - f(w)$ where $f(w) = \frac{g''(u_j(w)))}{g'(u_j(w))} u'_j(w)$ and $f(w) < 0$. Therefore, the income effects for agent i after the realization of the lottery can be written as:

$$IC^i(x,w) = (AR^j(x,w) - f(w+x)) - (AR^j(w) - f(w)) = IC^j(x,w) - (f(w+x) - f(w).)$$

For f increasing in w, the income effect after the assignment of a positive-outcome realized risk would reduce risk aversion more for agent i than it would for agent j. This condition can easily be verified for the commonly used power utility function $u(w) = \frac{w^{1-\rho}}{1-\rho}$. For this function, $RR = \rho$ is the coefficient of relative risk aversion. Absolute risk aversion is decreasing, with $AR = \frac{\rho}{w}$. If females have higher relative risk aversion than males (i.e., $RR^f = \rho^f = \rho^m + c$), they thus have higher baseline absolute risk aversion $AR^f = \frac{\rho^m + c}{w}$. Since $AR^f = AR^m - f(w)$ it must hold that

$f(w) = -\frac{c}{w}$ with $f' > 0$. Therefore, in this family of functions, income will increase risk taking more for agents with higher baseline risk aversion.[2]

3.3.1.2. Income effects associated with a risky sum

Similar to the effect of a certain change in income, the possible income associated with the potential outcomes of unrealized risk also can affect risk aversion. Allowing for a stochastic element in wealth in the form of random variable $\tilde{k}$, and making use of the indirect utility function (Kihlstrom et al., 1981), it holds for all w that $v(w) = Eu(w + \tilde{k})$. Since $v^{[n]}(w) = Eu^{[n]}(w + \tilde{k})$ (Gollier, 2001), it holds that:

$$AR(w, \tilde{k}) = -\frac{v''(w)}{v'(w)} = -\frac{Eu''(w + \tilde{k})}{Eu'(w + \tilde{k})}$$

Next let $\tilde{x}$ denote the random variable associated with the unrealized risk represented by lottery $L^*(X) = (p; b, c)$. Since $Ev(w + \tilde{x}) = p(v(w + b)) + (1 - p)v(w + c)$, it follows that:

$$AR(w, \tilde{k}, \tilde{x}) = -\frac{Ev''(w + \tilde{x})}{Ev'(w + \tilde{x})} = -\frac{pv''(w + b) + (1 - p)v''(w + c)}{pv'(w + b) + (1 - p)v'(w + c)}$$

This implies:

$$AR(w, \tilde{k}, \tilde{x}) = sAR(w, \tilde{k}, b) + (1 - s)AR(w, \tilde{k}, c) \tag{3.1}$$

[2] If the observed gender difference in risk attitude is driven by lower initial wealth levels/lower reference points for females ($w_f < w_m$,)5but they have otherwise identical DARA preferences, this could also explain the higher risk aversion of females. An increase in income by x then comes with a larger reduction in risk aversion for females as long as $IC' > 0$, which is satisfied for the power utility function, as for this function $IC' = \frac{\rho}{w^2} - \frac{\rho}{(w+x)^2} > 0$.

with $s = \frac{1}{1+\frac{1-p}{p}\frac{v'(c+w)}{v'(b+w)}}$. Note that $s>0.5$ for $c>b$. This can also be shown to hold for non-stochastic reference-point models such as CPT.[3]

The effect of the potential incomes of the background risk can thus be written as:

$$IC(w, \tilde{k}, \tilde{x}) = sIC(b.w) + (1-s)IC(c, w.)$$

For $b>c>0$ and with DARA preferences risk aversion would therefore also reduce in the presence of background risk.

The theoretical implications of the power utility function for gender differences in income effects are similar for unrealized risk as they were for realized risk. For stochastic income in the form of lottery $L^*(X)$, it holds that $IC^f(\tilde{x}, w) = s(IC^m(b, w) - f(w+b)) + (1-s)(IC^m(c, w) - f(w+c)$ and thus for f increasing in wealth, the income effect after the assignment of a positive-outcome unrealized risk would reduce risk aversion more for a female with higher baseline risk aversion than it would for a male with lower baseline risk aversion.

Assuming an income based decision model with narrow framing and a power utility function results in the following empirically testable predictions:

(1) An increase in income reduces aversion, i.e., $IC < 0$.

(2) *A background risk with strictly positive potential outcomes reduces aversion for subsequent risks, i.e.,* $IC < 0$.

(3) A background risk or realized risk with strictly positive outcomes will reduce female risk aversion for subsequent risks more, i.e., $IC^f < IC^m$.

[3]Within non-stochastic reference-point-based decision models such as CPT, there would be an effect of the potential incomes associated with unrealized risk on subsequent risk preferences if people fail to update their reference point. For a background risk with strictly positive outcomes that is not included in the reference point, condition (1) would still hold, although behavioral probability weighting would affect the weighting(s). Intuitively, because the outcomes of the risk are strictly positive, and risk preferences are still derived from the slope of the (concave) value function, the same results hold as before. Specifically, for a behavioral value function (denoted u*(x)), the indirect value function in the presence of unrealized risk can be set as $v(.) = Eu^*(\tilde{x}) = w(p)u^*(b) + (1-w(p))u^*(c)$.

3.3.2. Determining winning effects for men and women

Realized risks can affect risk attitude through non-income channels when people are affected by the experience of 'winning' or 'losing.' In order to define the winning effect associated with the realization of risk, we denote $\mathcal{L}$ the set of simple lotteries on the finite outcome set X. A lottery L in $\mathcal{L}$ is a function $L : X \to \triangle$ that has $L(x) \geq 0 \, \forall x \in X$ and $\sum_{x \in X} L(x) = 1$. For any x ∈ X, let δ_x denote the degenerate lottery that has $L(x) = 1$. Next, refer to $\mathcal{L}_x \supset (\mathcal{L} \backslash \delta_x)$ as the set of non-degenerate lotteries that have positive support on x. Then denote the set of outcome-generating processes for outcome x as $Q_x = \{\mathcal{L}_x, \delta_x\}$. When assuming that risk attitudes are not sensitive to the way in which the income comes about (i.e., no consequentialism), it holds that $AR(w, x, q_x) = AR(w, x)$, and thus $AR(w, x|\delta_x) = AR(w, x|L_x)$. If in contrast, people are sensitive to the experience that comes with the realization of risk, this condition will not hold. Letting x* be the highest possible outcome of lottery $L_{x*} \in \mathcal{L}_{x*}$, the ***winning*** of lottery L_{x*} implies the outcome x=x*.

Definition. For the lottery L_x the ***winning effect*** of winning x on absolute risk aversion, at wealth level *w,* is defined as $WN(w, x, L_x) = AR(w, x|L_x) - AR(w, x|\delta_x)$.[4]

The absolute risk aversion after winning x* in lottery L_{x*} can then be written as $AR(w, x*|L_{x*}) = AR(w) + IC(w, x*) + WN(w, x*, L_{x*})$. The assumption of non-consequentialism does not suggest a role for the experience of winning, resulting in the following testable condition:

[4] A reference point dependent model can imply a winning effect if the reference point of the subjects in the high fixed sum treatment has been updated (e.g. $AR(w, x|\delta_x) = AR(w)$). but the outcome of the background risk in the background risk treatment is not yet incorporated in the reference point. Letting w' denote the reference point in the presence of background risk we would have $AR(w, x|L_x) = AR(w') + IC(w*, (x + w - w')) = AR(w) + IC(w, (x + w - w'))$ and there will be relatively less risk aversion for the subjects in the background risk treatment that win the high outcome. Conditional on an income effect the surprise of winning will reduce risk aversion as it brings people in the gain domain.

(4) Both males and females are not sensitive to the effect of winning after a fortunate realization of a risk, i.e., $AR(w, x * | L_{x*}) = AR(w) + IC(w, x*)$.

3.3.3. Determining anticipation effects for men and women

If the presence of unrealized risk affects risk attitudes through channels other than income, 3.1 will not hold.

Definition. For random variable $\tilde{x}$ associated with lottery L_x, the ***anticipation effect*** on absolute risk aversion at wealth level w is defined as $ANT(w, x) = AR(w, \tilde{x}) - s(AR(w, b)) + (1 - s)(AR(w, c))$.

The anticipation effect can also be written as $ANT(w, \tilde{x}) = AR(w, \tilde{x}) - s(IC(w, b) + AR(w)) + (1 - s)((IC(w, c) + AR(w)$. If there are no income effects, this reduces to $ANT(w, \tilde{x}) = AR(w, \tilde{x}) - AR(w)$. As the conventional decision theories do not allow for an effect of anticipation, they make the following prediction:

(5) Females and males are not sensitive to an effect of anticipation, i.e., $AR(w, \tilde{x}) = s(AR(w, b)) + (1 - s)((AR(w, c))$.

3.4. Experimental Design

The experiments were run at the Harvard Decision Science Laboratory in the summer of 2013 with a total of 160 subjects. All subjects were students, 50% were male, about 2/3 were American, average age was 21-22 years, and a bit more than half were Caucasian. All subjects were identified by code numbers and remained anonymous. The experiments were programmed and conducted with the software Z-Tree Fischbacher (2007).

The experiment consists of four parts. The first part involves a filler task: All subjects started the experiment by completing a 45-minute questionnaire for which they received $15. This task was designed to prevent a house-money effect associated with the earning of $15. The second part of the experiment was designed to measure the effect of unrealized risk. Each subjects was either in a background-risk treatment, a fixed-sum treatment or a control treatment (see Tab. 3.1). The 75 subjects in the *Background Risk* treatment were provides with a dice and were informed that an experimenter would come to roll this dice later in the experiment and that they would earn $30 when the dice turned up 1, 2, or 3, and $2 otherwise. Subjects in the *High-Fixed Sum* treatment were notified that they received $30, and subjects in the *Low-Fixed Sum* treatment were notified that they received $2. Subjects in the *Control* treatment received nothing. Thereafter, subjects participated in a lottery-based risk attitude elicitation task (*Lottery List Task*). During this task, subjects in the background-risk treatment were reminded of the assigned risk through a heading bar at the top of their computer screens, and subjects in the high- and low-fixed-sum treatments were reminded of their assigned fixed sums. The elicitation method is similar to Abdellaoui et al. (2011) with the exception that I also include mixed-outcome and negative-outcome lotteries. Subjects could earn a maximum of $15 and lose a maximum of $15. The task comprises 15 lists of 7 decisions, each list being associated with one particular lottery. In each of the 7 decisions, the lottery is contrasted with a different fixed sum. The lowest fixed sum equals the lowest possible lottery outcome, and the highest fixed sum equals the highest possible outcome. After subjects make their decisions, one of the 15x7 decisions is randomly selected for payment. (See Fig. 3.1 for a screenshot of one of the lottery lists in the lottery list task.) This method is found to be more tractable and less sensitive to the bias that is associated with procedures that rely on comparing non-degenerate lotteries, (Abdellaoui et al., 2011; Bosch-Domènech and Silvestre, 2012).

Table 3.1.: Treatments for the Effect of Unrealized Risk

	High Sum (N)	Low Sum (N)	Control (N)
Unrealized Risk	Background Risk (75)	Background Risk (75)	
Assigned Sum	High Fixed Sum (29)	Low Fixed Sum (25)	Control (23)

Table 3.2.: Treatments for the Effect of Realized Risk

	High Sum (N)	Low Sum (N)	Control (N)
Realized Risk	High Realized Risk (36)	Low Realized Risk (39)	
Assigned Sum	High-Fixed Sum (29)	Low-Fixed Sum (25)	Control (23)

Part three of the experiment measured subjects' sensitivity to realized risk. Subjects in all treatments were informed about the selected decision in the Lottery List Task, the outcome of the associated lottery, and their payoffs. Then, in the Background-Risk treatments, an experimenter came by to roll the dice for each subject. Depending on the outcome of the dice roll, subjects received either the low outcome (*Low-Realized Risk*) or the high outcome (*High-Realized Risk*). (See Tab. 3.2.) Next, all subjects answered a standard incentivized risk elicitation task on the positive domain, similar to Holt and Laury (2002) but with quadrupled payoffs. This task is displayed in Fig. 3.2.[5] One of the 10 decisions was randomly selected for payment. In part 4, subjects filled out a short demographic questionnaire (most importantly for us indicating their gender) and received their payoffs. The total experiment took about 1.5 hours.

[5] I excluded 8 subjects from the analysis of the second part of the experiment because they had received incorrect information.

Figure 3.1.: Lottery List Decision Task

3.5. Results

3.5.1. The effect of unrealized risk

The decisions in the lottery list task in the first part of the experiment enable us to estimate the effect of unrealized risk. I apply the approach outlined in the conceptual framework on the number of safe choices in this task, since this number is a positive monotonic transformation of the coefficient of absolute risk aversion. Tab. 3.3 describes the number of safe choices per treatment by gender.

Table 3.3.: Number of Choices Favoring the Safe Option in the Lottery List Task

	Control	Background Risk	Fixed Sum	Fixed Sum & Background risk	Low-Fixed Sum	High-Fixed Sum
Female	57.77 (N=13)	50.31 (N=42)	52.59 (N=27)	51.20 (N=69)	51.43 (N=14)	53.85 (N=13)
Male	49.54 (N=13)	51.80 (N=35)	51.80 (N=30)	51.80 (N=65)	51.58 (N=12)	49.54 (N=18)

Figure 3.2.: HL Decision Task

Decision	Option A		Option B	
1	10% chance of $8	90% chance of $6.4	10% chance of $15.4	90% chance of $ 0.4
2	20% chance of $8	80% chance of $6.4	20% chance of $15.4	80% chance of $ 0.4
3	30% chance of $8	70% chance of $6.4	30% chance of $15.4	70% chance of $ 0.4
4	40% chance of $8	60% chance of $6.4	40% chance of $15.4	60% chance of $ 0.4
5	50% chance of $8	50% chance of $6.4	50% chance of $15.4	50% chance of $ 0.4
6	60% chance of $8	40% chance of $6.4	60% chance of $15.4	40% chance of $ 0.4
7	70% chance of $8	30% chance of $6.4	70% chance of $15.4	30% chance of $ 0.4
8	80% chance of $8	20% chance of $6.4	80% chance of $15.4	20% chance of $ 0.4
9	90% chance of $8	10% chance of $6.4	90% chance of $15.4	10% chance of $ 0.4
10	100% chance of $8	0% chance of $6.4	100% chance of $15.4	0% chance of $ 0.4

In the control treatment (Column 1), I replicate the often-reported finding that females have significantly higher baseline risk aversion than males. Females are about 16.6% more likely to select the safe option than males (p=0.02, Wilcoxon rank sum test). However, the introduction of background risk (Column 2) eliminates this common result; the difference is no longer significant (p=0.81, Wilcoxon rank sum test). The driver behind this result is that unrealized risk significantly increases risk taking by females. Compared to the control treatment, the average number of safe choices in the background-risk treatment for females is about 13% lower (p=0.04, Wilcoxon rank sum test). In contrast, the introduction of background risk does not have an economic or statistically significant effect on the risk preferences of males (p=0.60, Wilcoxon rank sum test).

I next explore the mechanisms behind this increase in risk taking for females and examine the respective roles of income and anticipation. An effect of the potential income associated with the background risk is conditional on an income effect of assigned non-stochastic income (3.1) When comparing the number of safe choices in the fixed-sum treatments (Column 3), I find that income also succeeds in eliminating the gender difference in risk aversion (Wilcoxon rank sum test, p=0.75). Compared to the control treatment (Column 1), risk taking of females is increased by about

9% in the fixed-sum treatments, and an effect of non-stochastic income is therefore directionally supported. A Wilcoxon rank-sum test, however, fails to reject the null hypothesis of equal distribution and central tendency (p=0.12). A Kolmogorov Smirnov test does find a difference in distribution (p=0.08). Comparing the decisions of females in the low-fixed-sum treatment (Column 5) with their choices in the high-fixed-sum treatment (Column 6), I do not find a statistically significant effect of the level of income (Wilcoxon rank sum test, p=0.40 and Kolmogorov Smirnov test, p=0.63).

Comparing females' choices in the fixed-sum treatments (Column 3) with their choices in the background-risk treatment (Column 2), there is a 4 percentage points higher increase in risk aversion in the background-risk treatments. This is not significant (Wilcoxon rank sum test, p=0.36), thus it cannot be rejected that the effect of background risk on females' risk attitude is fully driven by income effects. As expected by the earlier finding that unrealized risk does not affect males' risk preferences, I find no effect of non-stochastic income on male risk preferences (Wilcoxon rank sum test, p=0.50 and Kolmogorov Smirnov test, p=0.60).

Figure 3.3.: Distributions of Risk Preference in the Lottery List Task by Lottery Domain

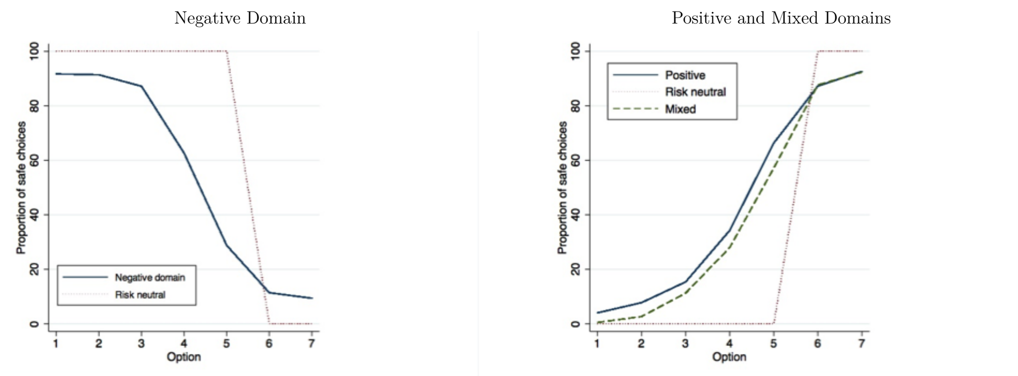


Fig. 3.3 reports the risk attitudes in the lottery list elicitation task in more detail and displays the decisions by lottery domain. The left panel of Fig. 3.3 displays the average percentage of choices for the safe option for the 5 negative outcome lotteries. p=0 value of the safe option increases over each decision, with decision 1 having a safe value that equals the highest (least negative) outcome of the lottery, and the safe value of decision 7 having the lowest (most negative) outcome. The dotted line in the figure depicts the risk-neutral decision for each option. Risk neutrality would imply choosing the risky choice in decisions 6 and 7 only. The actual choices for decisions 1 through 5 is suggestive of the risk-seeking behavior that is often observed in the loss domain. A large proportion of people prefer the risky lottery to the safe choice with higher expected value. For decisions 6 and 7, more than 90% of people select the gamble, which now has a higher expected value than the safe option. The remaining 10% of choices for the safe option may be the product of risk averse preferences for a portion of the subject pool, or the result of a stochastic choice process.

The panel on the right of Fig. 3.3 shows the distribution for the mixed- and the positive-domain lottery questions. The value of the fixed sum is now increasing over the 7 decisions, with the fixed sum of decision 7 equaling the highest lottery outcome. Risk neutrality would imply a preference for the safe option in decisions 6 and 7. The large proportion of choices for the safe option for options 1 through 5 in both types of lotteries is indicative of risk aversion. In the mixed domain, the preference for the safe choice appears slightly lower than in the positive domain. A possible reason for this reduction in risk aversion is that the safe choice for the first 4 decisions can now be a certain loss, which may induce a preference for risk taking in those questions. In questions 6 and 7, about 10 % keep selecting the gamble, even though it has lower expected value. This may be the result of a stochastic choice process or a very strong preference for risk for a small proportion of subjects.

To examine the effect of unrealized risk more closely and allow for a differential sensitivity to gain and loss frames, I proceed with a regression analysis of the effect of unrealized risk on the coefficient of relative risk aversion. This offers the additional advantage of allowing for the effect of individual characteristics and individual differences in the precision of estimates.

Table 3.4.: EUT Specification for the Lottery List Elicitation Task

	Positive Domain (α)				Negative Domain (β)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Background risk	0.10	-0.02	0.10	0.088*	-0.17**	0.09	-0.17**	-0.23**
	(0.065)	(0.070)	(0.065)	(0.049)	(0.076)	(0.090)	(0.076)	(0.070)
Low-fixed sum	0.14*	0.05	0.14*	0.11*	-0.11	0.01	-0.11	-0.11
	(0.083)	(0.105)	(0.082)	(0.066)	(0.093)	(0.106)	(0.093)	(0.073)
High-fixed sum	0.16*	-0.10	0.16**	0.15**	-0.18*	0.08	-0.18*	-0.14
	(0.067)	(0.079)	(0.067)	(0.063)	(0.107)	(0.075)	(0.107)	(0.088)
Male			0.18**	0.15**			-0.14*	-0.17**
			(0.082)	(0.072)			(0.083)	(0.076)
Background risk x Male			-0.12	-0.12			0.27**	0.37***
			(0.096)	(0.085)			(0.118)	(0.100)
Low fixed sum x Male			-0.19	-0.15			0.12	0.04
			(0.133)	(1.117)			(0.140)	(0.148)
High fixed sum x Male			-0.26**	-0.26**			0.26**	0.18
			(0.103)	(0.102)			(0.130)	(0.121)
Cognitive reflection test				0.06***				0.10***
				(0.016)				(0.022)
Age				-0.01				-0.01
				(0.008)				(0.012)
Constant	0.55***	0.74***	0.55*	0.77***	0.68***	0.55***	0.68***	0.66**
	(0.056)	(0.058)	(0.056)	(0.188)	(0.060)	(0.058)	(0.060)	(0.269)
Observations	2870	2730	5600	5600	2870	2730	5600	5600

Subject-clustered standard errors in parentheses. ***<0.001; **<0.05; *<0.01

Tab. 3.4 presents the regression results of an Expected Utility Model, assuming a CRRA power utility function of $u(x) = x^{\alpha}$ (corresponding with a CRRA coefficient of *1-a*.) The results in the first 4 columns use only the questions on the positive domain, and columns (5) through (8) report the negative domain lottery decisions. The results are robust to a Fechner specification of stochastic errors in the latent choice process (not shown). Column (1) reports this model for females only, and Column (2) for males only. Comparing the estimates of α (*Constant*) for females (*0.55*) and males (*0.74*) suggest increased risk aversion for females in the positive domain. For females, there are significant treatment effects for the low fixed sum and the high fixed sum; for males, there are not. These estimates of α are in line with the existing estimates in the literature (e.g., Harrison and Rutstroem (2008)).

In Columns (3) and (4), I allow for potential gender differences in the effects of income and background risk and include interactions with gender and the treatment indicators. In Column (4), I also allow for the effect of age and cognitive reflection, based on the score in the Cognitive Reflection Test developed by Frederick (2005) that measures both cognitive capacity and attributes such as impulsivity. Columns (3) and (4) show, as expected, that males have significantly lower baseline risk aversion. Females significantly increase risk taking in the low-fixed sum, the high-fixed sum and–after controlling for age and cognitive reflection–in the background risk treatment. The increase in risk taking by females in the fixed-sum and the background-risk treatments is considerable, as the effects negate most of the initial gender difference in risk attitude: Tests do not reject that females have the same risk preference as males in the low-sum treatment ($\chi^2(1) = 0.77, p = 0.38,$) or the background-risk treatment ($\chi^2(1) = 0.57, p = 0.45$) and they are significantly less risk averse in the high-fixed-sum treatment ($\chi^2(1) = 5.07, p = 0.02$). The effect of background risk on females' risk attitude can be fully explained by the income effect, as the equality between the three treatment indicators cannot be rejected ($\chi^2(2) = 1.46, p = 0.48$). Those who are older are more risk averse, but not significantly. Those who score better on the Cognitive Reflection Test are significantly less risk averse, a finding consistent with Frederick (2005), who reports more risk taking in the gain domain for those with high cognitive reflection scores.

Next I consider the negative domain lottery decisions in Columns (5)-(8). For the negative domain, I assume the power utility function $u(x) = -(-x)^{\beta}$ so that β <1 implies convexity in the negative domain. In Column (5), I estimate the model for females only, and in Column (6), for males only. The estimate of β for females (*0.68*) and males (*0.55*) suggests that both are risk seeking in the negative domain, but males take more risk. For females, the background risk and the high-fixed-sum treatments significantly increase risk taking, for males they do not.

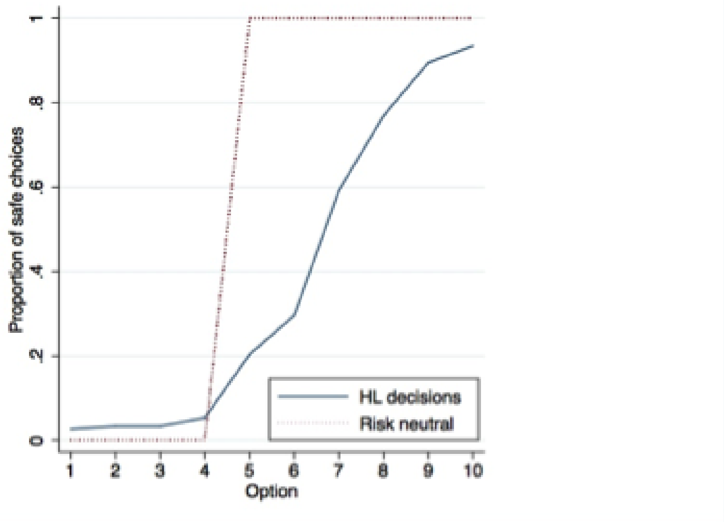
In Columns 7 and 8, I allow for gender interaction effects with the treatment indicators. In the baseline, males are significantly less risk averse. The background-risk and the high-fixed-sum treatments increase risk taking for females, the increase of risk taking in the low-fixed-sum treatment is not significant. Tests cannot reject that females and males have equal risk preferences in the low-fixed-sum treatment ($\chi^2(1) = 0.59, p = 0.74$) and the high-fixed-sum ($\chi^2(1) = 2.43, p = 0.12$) and females are significantly less risk averse than males in the background risk($\chi^2(1) = 5.56, p = 0.002$). The effect of the fixed sums is of a similar order of magnitude as the findings in the positive domain. The effect of unrealized risk appears to increase risk taking more in the negative domain than it did in the positive domain, but tests cannot reject equality of the treatment indicators ($\chi^2(2) = 0.59, p = 0.74$) and the effect of the background risk can thus still be fully explained by income. In Column (8), I allow for the effect of Cognitive Reflection and age. In contrast to the positive domain, Cognitive Reflection now is associated with a reduction in risk taking. This is consistent with Frederick (2005), who also finds that high cognitive reflection increases the likelihood that people accept a sure loss to avoid playing a lottery with lower expected value. The effect of background risk increases after the introduction of these variables, but tests can still not reject the equality across treatments ($\chi^2(2) = 2.10, p = 0.35$.)

In summary, unrealized risk reduces relative risk aversion for females, both in the negative and the positive domains. Income is also associated with an increase in risk taking for females and the strength of income and unrealized risk on risk taking cannot be distinguished. For men, there is no effect of unrealized risk or fixed income, except in the negative domain where receiving the high-fixed sum reduces their risk-seeking behavior.

3.5.2. The effect of realized risk

To study the effect of realized risk, I focus on the answers to the Holt and Laury task in the second part of the experiment. Fig. 3.4 plots the percentage of choices for the safe lottery for each of the 10 decisions presented. As outlined by the dotted line in the figure, risk neutrality would suggest a switching point after decision 4. The fact that a large number of subjects switch later on is indicative of risk aversion.

Figure 3.4.: Percentage of Safe Choices in Holt and Laury Task



Tab. 3.56 displays the number of safe choices across treatments by gender. On average, males choose the safe option 5.91 times, significantly less than the average 6.41 safe choices by females, (Wilcoxon rank sum test, $p = 0.05$). This is not driven

by the fixed-sum treatments; in line with the findings for the lottery list elicitation task, women take more risk in these treatments, although not significantly more (Wilcoxon rank sum test, p=0.78). When looking at the effect of winning after the realization of risk, I find that the difference across genders increases. In the control treatment, females take 15% less risk than males; in the high-realized risk treatment, females take over 33% less risk than males. In the low-background-risk treatment, where people receive $2 and "lose" the realized risk, there does not appear to be an effect on the gender gap in risk attitudes.[6]

Table 3.5.: Number of Choices Favoring the Safe Option in the Holt and Laury Task

	Control	High-Realized Risk	Low-Realized Risk	Fixed Sum	Low-Fixed Sum	High-Fixed Sum	Total
Female	6.40	6.82	6.20	6.23	6.43	6.00	6.42
	(N=10)	(N=22)	(N=20)	(N=26)	(N=14)	(N=12)	(N=104)
Male	5.77	5.21	5.68	6.46	6.82	6.24	5.91
	(N=13)	(N=14)	(N=19)	(N=28)	(N=11)	(N=17)	(N=102)

The first part of the experiment ruled out an income effect for males. In this part, I also do not find a significant income effect for males when comparing the risk attitudes in the fixed-sum treatment and the control treatment (Wilcoxon rank sum test, p=0.22). This implies that any potential increase in risk taking after the positive realization of risk can be attributed to the winning effect. The positive realization of risk increases risk taking by about 22% (Wilcoxon rank sum test, p=0.10). For females, the income effect is no longer significant when comparing the fixed-sum treatments and the control treatment (Wilcoxon rank sum test, p=0.77).

[6]Note that the observed risk attitudes in the Holt and Laury task are conditional on profits made in the earlier lottery list elicitation task. This is unlikely to complicate the measurement of the treatment effects as there are no significant differences in the obtained profits in this task across treatments (Kruskal Wallis Test, $\chi^2(4) = 5.40$, p=0.24). These profits are determined by randomly selecting one of 105 decisions for payment, the impact of risk attitude and thus the effect of the treatment effects in the first part on these profits are therefore negligible.

This points at a transient effect of income, consistent with the readjustments of reference points and the recency effects that are often observed in the literature (e.g., Erev and Haruvy (2013). The gains of $30 and $2 are old news once the second risk attitude method is implemented and, therefore, may not affect female's decision making as much. When comparing risk taking in the *High Realized Risk* treatment to risk taking in the fixed-sum treatments and the control treatment, I do not find an effect from winning on females (Wilcoxon rank sum test, p=0.39). This is unexpected because in a reference point dependent decision model, the income associated with winning $30, is $14 higher than the expected value of the lottery, although this is not as high as the $30 increase in the first part of the experiment, the induced gain frame would suggest an increase in risk taking. This suggests that there may be a counter factor that actually reduces their desire for risk.

This comparison across treatments obscures the income differences within treatments because of the profits obtained in the lottery list elicitation task in the experiment's first part. In addition, it does not allow for individual variation. I will, therefore, proceed with regression analyses. I estimate interval-censored models, using the interval-censored relative risk aversion (rr) as a dependent variable, assuming a CRRA power utility function of $u(x) = x^{(1-rr)}/(1-rr)$ (Tab. 3.6). The results are robust to other specifications, including a simple regression on the number of safe choices (not reported.) Of the original 152 subjects, a good amount (127) show the conventional behavior; they start off choosing the lotteries described under option A in the left of Fig. 3.2 and have a single switching point to the riskier lotteries described under option B. For 23 of the 25 subjects who deviated from this behavior, I could still calculate an interval by using the lower bound associated with the first risky choice as the lower bound, and using the upper bound associated with the last risky choice as the upper bound. Two subjects made a risky choice in the first row and a safe choice in the last; thus, I could not narrow the interval and excluded them from the analysis.

To allow for the treatment induced income effect, I include the indicators *Get30* and *Get2*. *Get30* reflects the subjects who are in the high-fixed-sum treatment *and* the subjects in the background-risk treatment who won \$30 after realization of the risk. *Get2* reflects the subjects who were in the \$2 fixed-sum treatment, and those who received \$2 after realization of the risk. The indicators *High-Realized Risk* and *Low-Realized Risk*, reflect the subjects in the background-risk treatment who only received \$30 or \$2 after the dice were rolled.

In Column (1) I estimate the model for females, and in Column 2 for males. High-realized risk increases risk taking for males and reduces risk taking for females, but the effect is not significant for either gender. Once I control for interactions between gender and treatment effects (Columns (3)-(6)), I find that males significantly reduce their risk aversion after winning and receiving the high-realized risk ($\chi^2(1) = 2.70, p = 0.10$) and this effect is many magnitudes stronger than the effect that being male has on baseline risk attitude ($\chi^2(1) = 2.88, p = 0.09$). Females in contrast, do not increase risk taking after winning.

In Column (4), I allow for the effect of profits in the lottery list elicitation task. I allow for gender-specific income effects of the earnings in the lottery list elicitation task, and find that higher profits in this task cause females to significantly increase their risk attitude; for males there is no such effect. In column (5) I allow for effects of age and cognitive reflection and find that older subjects have higher risk aversion.

To sum up, winning after the realization of risk increases risk taking in males, but does not increase risk taking in females. Risk preferences of females are no longer significantly affected by earnings they made earlier in the experiment, but are affected by the more recent earnings that they made in the previous lottery list

Table 3.6.: EUT Specification for the Lottery List Elicitation Task

	(1)	(2)	(3)	(4)	(5)
rr					
Get30	-0.22	0.15	-0.15	-0.11	-0.15
	(0.318)	(0.173)	(0.271)	(0.257)	(0.255)
Get2	-0.01	0.30*	0.15	0.12	0.16
	(0.262)	(0.170)	(0.152)	(0.149)	(0.146)
High-realized risk	0.31	-0.34	0.31	0.29	0.32
	(0.265)	(0.241)	(0.260)	(0.253)	(0.241)
Low-realized risk	-0.07	-0.32**	-0.19	-0.16	-0.18
	(0.186)	(0.150)	(0.122)	(0.124)	(0.121)
Male			-0.11	-0.15	-0.14
			(0.101)	(0.103)	(0.102)
Male x Get30			0.25	0.21	0.28
			(0.294)	(0.284)	(0.285)
Male x High-realized risk			-0.66*	-0.64*	-0.71**
			(0.357)	(0.353)	(0.358)
LLT Profits				-0.01**	-0.01**
				(0.006)	(0.006)
LLT Profits x male				0.01	0.01
				(0.008)	(0.008)
Age					0.05**
					(0.021)
CRT					-0.01
					(0.046)
Constant	0.72***	0.49***	0.65***	0.69***	-0.34
	(0.208)	(0.106)	(0.133)	(0.127)	(0.485)
lnsigma					
Constant	-0.57***	-0.73***	-0.64***	-0.66***	-0.67***
	(0.159)	(0.164)	(0.113)	(0.113)	(0.115)
Observations	760	740	1500	1500	1500

Subject-clustered standard errors in parentheses. ***<0.001; **<0.05; *<0.01

elicitation task.

3.6. Discussion and Conclusions

This paper shows that there are stark gender differences in the way men and women respond to a risky environment. I find that a positive outcome background risk temporarily increases risk taking in females but not in males. Although baseline risk aversion for females exceeds that of males, the introduction of background risk eliminates this difference. A potential effect of anticipation cannot be completely discarded, but the results suggest that this is driven by the effect of income, to which males are not sensitive. In contrast, once the outcome of the background risk is revealed and subjects 'win' the risk and receive the high outcome, risk taking increases in males but not in females.

Increased sensitivity to income for females is consistent with the predictions of utility functions such as the isoelastic utility function. For these functions, the higher baseline risk aversion of females implies higher sensitivity to previous earnings. This result is applicable to EUT models with narrow framing, and –as long as reference points have not been updated to take these earnings into account– reference-point dependent models such as CPT. However, the complete lack of response to income by males suggests that they are not affected by their previous experimental earnings. Their behavior is consistent with a decision model such as Expected Utility over Income (EUI). This model stipulates a scenario of even 'narrower' framing in which people only consider the potential earnings of a prospect, and are not affected by *any* previous earnings.

The increase of risk taking in males after the experience of winning the lottery is in line with studies on all-male subjects that find an increase in risk taking after the experience of a win (Apicella et al., 2014). The reason that females do not increase risk taking after a win is potentially driven by a different emotional or hormonal response (testosterone has been suggested to be a major driver of the winner's effect). Winning may also affect beliefs on the likelihood of winning subsequent risks differently for females; they may be less prone to hot-hand beliefs, or be more prone to the gambler's fallacy and believe that their luck ran out after the recent win.

It is left to future research to replicate these findings, to study the effects of negative-outcome risks, and to more closely investigate the mechanisms behind these gender differences. It is possible that the genders are by nature differently disposed towards risk, but the differences may also be socially constructed, or be a consequence of special circumstances that vary across gender. For example, females may have lower expected incomes, or lower reference points of income; this may be a driver in the gender differences in baseline risk attitude as well as their increased sensitivity to income.

The finding that the exposure to past and current risks affects decision-making supports a more dynamic view of risk taking, and can be a motivation for policy interventions to mitigate this threat to optimal judgment and decision-making. The winning effect and the income effect may also be used to induce changes in risk taking. In cases where increases in risk taking are not desirable, such as in a financial-trading setting, withholding the outcomes of previously acquired risks may improve decision-making in males. In instances where increases in risk taking by females is desirable, such as in a micro-finance setting, repeated stimuli may induce the desired effect.

4. When Performance Trumps Gender Bias: Joint Versus Separate Evaluation [1]

4.1. Introduction

Gender-based discrimination in hiring, promotion, and job assignments is difficult to overcome (e.g., (Neumark et al., 1996; Rich and Riach, 2002). In addition to conscious taste-based or statistical discrimination (Becker, 1976), gender biases are automatically activated as soon as evaluators learn the sex of a person. Biases lead to unintentional and implicit discrimination that is not based on a rational assessment of the usefulness of sex in predicting future performance (e.g.,Banaji and Greenwald, 1995 ,Bertrand et al., 2005). For example, science faculty rated a male candidate who applied for a laboratory manager position as significantly more competent and hireable than an otherwise identical female candidate, and this differential evaluation was moderated by the faculty's pre-existing bias against women (Moss-Racusin et al., 2012).

The most effective mechanisms to date to decrease the impact of such biases are

[1] This is joint work with Iris Bohnet and Max Bazerman and is currently under review for publication

blind evaluation procedures in which musicians audition behind a curtain. These methods have proven to substantially decrease gender discrimination in the selection of musicians for orchestras, but they are not very practical for most other evaluation and selection problems (Rouse and Goldin, 2000). Other attempts at overcoming gender biases include diversity training, which, however, seems to have had little impact (Dobbin et al., 2007). Gender quotas on search and evaluation committees have had mixed results, given that stereotypes tend to affect both male and female evaluators (Moss-Racusin et al., 2012; Bagues and Esteve-Volart, 2010). Quotas—e.g., for political bodies, corporate boards or senior management—are effective in increasing the fraction of members from underrepresented groups and with enough exposure to counter-stereotypic evidence, have been shown to affect gender stereotypes but have had mixed effects on performance (Miller and Matsa, 2013; Duflo et al., 2009, 2012).

Generally, there seems to be less room for discrimination, the more specific and equally available evaluation and selection criteria are (Reskin and Bielby, 2005). Building on this literature, this paper suggests a new intervention aimed at overcoming biased assessments: an “evaluation nudge,” in which people are evaluated jointly rather than separately regarding their future performance[2]. We expect cognitive shortcuts, such as group stereotypes, to have less of an impact when multiple candidates are presented simultaneously and evaluated comparatively than when evaluators look at one person at a time.

Both, joint and separate evaluation procedures are common for hiring and promotion decisions. Based on a recent survey of senior business executives in US companies with more than 1,000 employees (Penn, Schoen Berland Associates, inc., 2012),

[2] A nudge is any aspect of choice design that is based on psychological insights into how our minds work and that alters people’s behavior in a predictable way without restricting the freedom of individual choice. For nudges more generally, see Thaler and Sunstein, 2009.

in 30 percent of all promotion decisions, only one candidate was considered. For hiring decisions, we rely on the literature on sequential vs. non-sequential searches, building on Stigler (1961). In sequential search, a firm screens each applicant upon arrival and offers the job to the first applicant whose productivity exceeds a certain threshold. In non-sequential searches, a firm pools a number of applicants, screens them and offers the job to the best person in the pool.The former search strategy resembles separate and the latter joint evaluation. Recruitment strategies vary with firm and job characteristics but overall, about half of the hiring procedures studied seem to correspond to sequential (separate evaluation) and half to non-sequential (joint evaluation) searches (Ommeren and Russo, 2013; Oyer and Schaefer, 2011). Unfortunately, neither the promotion nor the hiring literature has examined the gender impacts of the different hiring and promotion strategies.

We employ laboratory experiments to examine whether evaluating candidates jointly rather than separately leads to individual performance playing a more important role than group stereotypes. In our experiments, we have subjects assume the role of either evaluators or candidates. Evaluators assess the likely future performance of candidates either in separate or joint evaluation of their performance. Specifically, they are informed of candidates' past performance and their sex (plus a number of filler characteristics) and are asked to decide whether given candidates are suitable for given jobs, either evaluating them separately or jointly, in one of two sex-typed tasks, a math or a verbal task. Most studies that measure explicit gender attitudes find that females are believed to be worse at math and better at verbal tasks than males(Perie et al., 2004; Price, 2011). Implicit association tests (IATs) measuring people's implicit attitudes report math and verbal skills to be associated with maleness and femaleness respectively (Greenwald and Banaji, 2002; Plante et al., 2009). The evidence on actual performance differences between the genders is mixed and varies by country and population, sometimes finding support for a gender gap in the expected direction, sometimes finding no gender differences, and, in recent years, finding a reversal of the gender gap in mathematics in several countries (Guiso et al., 2008). Despite the mixed evidence, we expect gendered beliefs to be sticky and these tasks to create stereotype-advantaged and stereotype-disadvantaged groups, with men being stereotype-advantaged in the math task and women in the verbal task. In addition, we expect that members of these groups will be affected by these biases even when at the individual level, conditional on the information available on the individual, gender is not informative and should not impact the evaluation. Although in an organizational context, additional complexities come into play, our framework relates to hiring, assignment, and promotion decisions where evaluators evaluate a given candidate's suitability for a specific job.

Our work is inspired by findings in psychology suggesting that people make more rational choices when examining products jointly rather than separately. We model a potential change in candidate assessments depending on the evaluation mode by introducing a simple behavioral model of information processing in which evaluators overweight the importance of the characteristics of the group that the candidate belongs to. In joint evaluation, more potentially counter-stereotypical data points are available at a time than in separate evaluation, thus providing evaluators with more information to update their stereotypical beliefs. The difference in the amount of available information could lead evaluators to choose a lower-performing stereotypical person in separate evaluation but a higher-performing counter-stereotypical person in joint evaluation.

In our experiment, gender was not correlated with task performance. Still, gender stereotypes had a strong and significant impact on evaluators' candidate assessments. Evaluators were significantly more likely to focus on group stereotypes in separate than in joint evaluation and on the past performance of the individual in joint than in separate evaluation, making joint evaluation the profit-maximizing evaluation procedure.

Our experimental findings have implications for the design of hiring and promotions procedures. Organizations may seek to overcome biases in hiring, job assignment and promotion because they want to maximize economic returns. They may worry about the inaccuracy of stereotypes in predicting future productivity, or they may hold gender equality as a goal in itself. Introducing joint rather than separate evaluation procedures may enable them to nudge evaluators toward taking individual performance information into account rather than gender stereotypes.

Our paper is organized as follows: Part 2 offers a conceptual framework, Part 3 describes the experimental design, Part 4 reports our experimental results and Part 5 concludes.

4.2. Conceptual Framework

Our evaluation nudge builds on the observation in behavioral decision research that people make more rational decisions in joint than in separate evaluation modes. These evaluation procedures may provide different amounts of data that allow evaluators to update their (possibly biased) beliefs on group characteristics to various degrees.

A simple Bayesian model of information processing may illustrate this. We assume that evaluators are informed of candidate(s)' individual past performance in a given task, their sex and the average past performance of the pool of candidates. Based on the information received, evaluators have to decide whether to "hire" the candidate (s) presented to them for future performance in the task or go back to the pool and be allocated a candidate at random. Evaluators are paid based on their candidates' future performance and thus have an incentive to select who they believe to be most productive, based on the candidate's future expected performance.

Evaluators either evaluate one candidate at a time (separate evaluation) or two candidates at a time (joint evaluation). In both conditions, evaluators hire one candidate only, either by selecting one of the candidates presented or by going back to the pool and being allocated a random candidate.

A "behavioral" Bayesian model of information processing (that allows evaluators to take irrelevant group characteristics into account) can explain an increase in the likelihood that evaluators choose higher-performing candidates in joint as compared to separate evaluation. Evaluating more than one person at a time implies having more data points available on the candidate's relative performance to update prior biased beliefs. If the new information is counter-stereotypical, it could theoretically shift beliefs enough for the evaluator to choose a counter-stereotypical person for a given job in joint but not in separate evaluation. We present this approach more formally in Appendix A. It yields the following empirically testable hypothesis:

Candidates are more likely to be selected based on their performance when evaluated jointly and more likely to be selected based on their gender when evaluated separately.

Our model is inspired by earlier work in psychology suggesting that evaluation modes affect the quality of decisions by making evaluators switch from more intuitive decision-making in separate evaluation, to more accurate choices in joint evaluation. This often is attributed to the System 1/System 2 distinction where people are assumed to have two distinct modes of thinking that are variously activated under certain conditions: the intuitive and automatic System 1 and the reflective and reasoned System 2 (Stanovich and West, 2000; Kahneman, 2011). Specifically, it has been suggested that the lack of comparison information available in separate evaluation leads people to invoke intuitively available internal referents (Kahneman and Miller's norm theory), focus on the attributes that can be most easily calibrated (Hsee's evaluability hypothesis), and rely more on emotional desires than on reasoned analysis (Bazerman, Tenbrunsel, and Wade-Benzoni's want/should proposition).

Loewenstein and Bazerman (1992)provided the original demonstration of preference reversals between joint and separate evaluation. In a two-party negotiation, they had study participants evaluate two possible negotiation outcomes—an even split of a smaller pie and a disadvantageous uneven split of a larger pie that still made both parties better off—either one at a time or jointly. When presented separately, most people preferred the equal split; when presented jointly, most preferred the money-maximizing alternative. Later studies on joint versus separate preference reversals found that brand name was more important than product features and price when people evaluated products separately rather than jointly (Nowlis and Simonson, 1997); people were willing to pay more to protect animal species when evaluating separately and to invest in human health when evaluating the two causes jointly Ritov et al. (1993); and people were willing to pay more for a small portion of ice cream in a tiny, over-filled container when evaluating separately but for a large portion of ice cream in an under-filled huge container when evaluating the two serving options jointly Hsee et al. (1999).

4.3. Experimental Design

Our experiment was conducted in the Harvard Decision Science Laboratory. We had 180 subjects participate as "candidates" in a math or a verbal task, and 328 subjects assumed the role of "evaluators," selecting one of the candidate for future performance in the task. All were American college students. We employed equal numbers of male and female evaluators. All our participants were identified by code numbers and remained anonymous to each other and to the experimenter. We employed a 2x2x2x2 experimental main (between-subject) design in which the key treatment condition of interest was the evaluation mode, joint or separate. In addition, we varied the individual candidates' past performance levels and their gender. Finally, candidates participated in either a math or a verbal task, with men being the stereotype-advantaged group in the math task and women the stereotype-advantaged group in the verbal task. The experiment was programmed and conducted in two stages, using Z-Tree software (Fischbacher (2007) Fischbacher 2007, sample instructions are included in Appendix B).

In stage 1, candidates participated in either a verbal or a math task and were paid based on their performance. In stage 2, evaluators were informed of candidates' past performance and their gender and then were asked to select a candidate for future performance in the same task. In stage 1, the candidates participating in the verbal task engaged in a word-search puzzle. They were given a list of 20 words and were instructed to mark as many of the words as they could find in three minutes in a matrix containing letters (Bohnet and Saidi (2011)). Most letters appeared in random order, but some formed words, and participants could search horizontally, vertically, and diagonally. On average, the 100 candidates participating in this task found 10 words (SD=3.81) in the first round and 12 words (SD=4.56) in the second round.

The math task involved correctly adding as many sets of five two-digit numbers as possible (Niederle and Vesterlund, 2007; Niederle et al., 2013). On average, the 80 candidates who participated in this task solved 10 problems correctly (SD=3.09) in the first round and 10 problems (SD=3.35) in the second round. After completing their task, participants filled out a short demographic questionnaire (most importantly for us, indicating their gender). Candidates then were paid based on their performance and were not informed of Stage 2 of the experiment.

In stage 2, evaluators in both the verbal and the math tasks were asked to choose a candidate, knowing that they would be paid based on that candidate's Round 2-performance. They could either choose the candidate presented to them, or go back to the pool and accept a randomly selected person. They had the candidate's Round 1-performance and his or her gender available as a basis for their decision, and were informed that on average, evaluators in the pool had provided 10 correct answers (as was the case for both tasks). To make the gender-attribute less salient, without creating any additional demographic variation, we took advantage of the demographic similarity of our candidates and provided evaluators with truthful filler information on their candidates' characteristics. In addition to learning a person's sex and past performance, evaluators were also informed that he or she was a student, American, and from the greater Boston area.

The candidates presented to the evaluators were either average or slightly below-average performers, having provided either 10 or 9 correct answers in the first round. We chose first-round performance scores at and below the mean performance level of the pool to make sure that our results were not driven exclusively by evaluators' risk (or loss) aversion. Presenting rather precise performance indicators compared to most performance measures in the field and using fewer possible criteria than typical in practice provides a conservative test for the impact of gender stereotypes. Heuristics likely play a more important role in situations where performance cannot be objectively measured (Stainback et al., 2010) and where multiple criteria for evaluation are available as they allow evaluators to focus on specific criteria only to justify their biased decisions (Norton et al., 2004). In our design, it seems difficult to justify neglecting individual performance information collected for the same task in the previous round.

In the separate-evaluation condition, evaluators were presented with either a male or a female candidate who was either an average- or below-average performer. We randomly selected four candidates of the required gender-performance combinations from our pool: Male-10, Female-10, Male-9, and Female-9, with identical filler characteristics. In the joint-evaluation condition, evaluators were presented with a male and a female candidate simultaneously, drawing from the same candidates used in the separate-evaluation condition. The candidates differed on both gender and past performance, leading to two possible combinations: Male-10/Female-9 and Male-9/Female-10.

After the experiment was completed, evaluators participated in an incentivized risk-attitude assessment task (Holt and Laury, 2002) and completed a short questionnaire that collected basic demographic information. Evaluators were paid based on their decision, i.e. either the chosen candidate's second-round performance or the randomly allocated candidate's second-round performance. They received $1 for every correct answer that the candidate had provided. Evaluator earnings varied between $17.8 and $34.75, which included a $10 show-up fee, experimental earnings, and the payment for the risk-attitude assessment task.

In addition to our main experiment, we ran a small control experiment in which we informed evaluators about candidates' present rather than past performance. Specifically, evaluators were informed of an candidate's second-round performance and then had to decide whether or not to select this candidate and be paid based on the candidate's performance in the second round or go back to the pool and accept a randomly allocated candidate. This experiment was designed to distinguish belief-based from taste-based discrimination. While in our main experiment, both motives could lead to gender-based decisions, in the control experiment, only taste-based discrimination was possible. We replicated the separate-evaluation conditions, in which we expected gender to be most prevalent, and used average performers, the group we were most concerned about being discriminated against. For separate evaluation, 23 evaluators participated in the male math condition, 27 in the female math condition, 33 in the male verbal condition, and 27 in the female verbal condition. Other than giving evaluators information about candidates' present rather than past performance, the control study was run identically to our main experiment. After participants had made their decisions, learned their outcomes, and given us their demographic information, they presented their code number and were given a sealed envelope containing their earnings.

4.4. Results

We first present candidates' performance in the two tasks, then examine what role gender and individual performance played in the two different evaluation modes, and finally examine alternative explanations.

4.4.1. Candidates' performance

We first examine whether or not having gender-stereotypical beliefs was accurate in our context. There were no significant gender differences in performance on either task, although directionally, the small differences we did observe accord with stereotypical assumptions[3]. Thus, ex-post, statistical discrimination was unwar-

[3]In the math task, performance levels were as follows: Round 1, men: Mean=10.63, SD=3.41; women: Mean=10.33, SD= 2.78; p=0.67. Round 2, men: Mean=10.63, SD=3.57; women: Mean=9.95, SD =3.13; p=0.37. In the verbal task, performance levels were as follows: Round 1, men: Mean=9.82, SD=4.05; women: Mean=10.98, SD=3.49; p=0.13. Round 2, men: Mean=12.46, SD=4.27; women: Mean=12.08, SD=4.87; p=0.68. There are no significant differences in variance across the genders, and the distributions in performance are not significantly different according to Kolmogorov-Smirnov tests

ranted. In addition, information on group characteristics in our experiment was always combined with individual performance information. Conditional on this performance information, stereotypes were completely irrelevant for predicting future performance.

Tab. 4.1reports the regression results of individual past (first-round) performance and gender on future (second-round) performance for both tasks. Columns 1 and 3 show that first-round performance was highly correlated with second-round performance, while the gender of the candidate was irrelevant for second-round performance in both tasks. In Columns 2 and 4, we control for potential gender differences in the relationship between first- and second-round performance and include an interaction term between the two variables. For example, the strong first-round performance of a candidate from a stereotype-disadvantaged group could be due to luck and thus be less predictive of future performance than the same performance by a member of a stereotype-advantaged group (and vice versa for low performance).

Columns 2 and 4 suggest that first-round performance was equally predictive of future performance for both genders.[4]

4.4.2. Evaluators' choices

We start by aggregating across both evaluation modes and both performance levels. In the math task (N=138), the likelihood that the stereotype-disadvantaged candidate, i.e., the woman, was chosen was 0.4, and the likelihood that the stereotype-

[4]In addition to controlling for the gender-specific randomness of performance across rounds, we also examined the possibility of gender-specific learning across rounds. On average and across both genders, little learning between rounds took place in the math task, while candidates in the verbal task performed significantly better in the second than in the first round, with men finding 2.64 and women 1.1 words more on average in the second than in the first round. However, the gender difference in learning was not significant, including in GLS regressions on performance in both rounds. Similar to the above results, average performance across both rounds was similarly correlated with the first-round performance of men and women in both tasks.

Table 4.1.: The Effect of Past Performance and Stereotypes on Future (Second-Round) Performance

	Math Task		Verbal Task	
	(1)	(2)	(3)	(4)
First-round Performance	0.849***	0.797***	0.708***	0.813**
	(0.08)	(0.15)	(0.10)	(0.15)
Male Candidate	0.420	-0.481	1.201	3.118
	(0.46)	(1.91)	(0.77)	(2.42)
First-round Performance xmale		0.086		-0.183
		(0.17)		(0.20)
Constant	1.189	1.723	4.311*	3.158
	(0.97)	(1.66)	(1.35)	(1.95)
N	80	80	100	100
R^2	0.6217	0.6232	0.3423	0.3478

Notes: Each specification is an OLS regression. Robust standard errors in brackets. The dependent variable isthe number of correctly added sequences in round 2 for the math task and the number of words found in round 2 for the word task. *** Significant at the 1 percent level. ** Significant at the 5 percent level.* Significant at the 10 percent level.

advantaged man was chosen was 0.46. In the verbal task (N=145), the likelihood that the stereotype-disadvantaged man was chosen across conditions was 0.39, while the likelihood that the stereotype-advantaged woman was chosen across conditions was 0.5. Thus, evaluators had a slight preference for men in math tasks and for women in verbal tasks, but these differences are not significant.

Looking at the two evaluation modes separately, we find that these differences were entirely driven by the stereotype-advantaged group being preferred in separate evaluation: Across both tasks and when evaluated separately (N=202), the likelihood that a candidate from the stereotype-advantaged group was chosen was 0.66, and the likelihood that someone from the stereotype-disadvantaged group was chosen was 0.49 ($\chi^2(1) = 5.45$, $p < .05$). In joint evaluation (N=252), stereotypes did not matter at all: 32 percent of the evaluators chose a candidate from the advantaged group and 31 percent chose a candidate from the disadvantaged group. (The remainder of the evaluators, 37 percent, decided to go back to the pool.)

Generally, the likelihood that a given candidate was chosen was higher in separate than in joint evaluation. We attribute this to the number of options available in separate versus joint evaluation. If evaluators had chosen randomly, a given candidate would have been chosen by 50 percent of the evaluators in separate evaluation and by only 33 percent in joint evaluation. Thus, compared to random selection, the stereotype-advantaged candidates were significantly more likely to be chosen than what a random process would have predicted in separate ($\chi^2(1)$=9.18, $p<.01$) but not in joint evaluation ($\chi^2(1)$ =.16, p=.69). The likelihood that stereotype-disadvantaged candidates were chosen did not differ from chance in either mechanism (for separate: $\chi^2(1)$=.04, p=.8445; for joint: $\chi^2(1)$=0.59, p=.4424).

Generally, our results hold for both the verbal and the math task. Tab. 4.2 shows our results for each evaluation mode, task, gender and performance level. In separate evaluation, the gender gaps in the likelihood of being selected are apparent, with the stereotype-advantaged group being favored in both the math and the verbal tasks. In joint evaluation, a performance gap emerged, with the higher-performing candidates being more likely to be selected than lower performers. Performance does not seem to matter in separate evaluation in the math task (but, in addition to gender, is relevant in the verbal task), and gender does not seem to matter in either task in joint evaluation.

To examine the significance of these results, we ran regressions. Tab. 4.3 shows that gender only affected decisions in separate evaluation (Column 1), and performance only affected decisions in joint evaluation (Column 2). Members of the stereotype-advantaged group were significantly more likely to be chosen in the sep-

Table 4.2.: Likelihood of Candidate Selection in Separate and Joint Evaluation

	Separate Evaluation		Joint Evaluation	
	Male	Female	Male	Female
Math Task				
Higher Performer	0.66	0.44	0.52	0.57
	(N=29)	(N=32)	(N=31)	(N=35)
Lower Performer	0.65	0.53	0.03	0.06
	(N=26)	(N=30)	(N=35)	(N=31)
Verbal Task				
Higher Performer	0.64	0.81	0.52	0.55
	(N=22)	(N=21)	(N=31)	(N=29)
Lower Performer	0.35	0.50	0.07	0.16
	(N=20)	(N=22)	(N=29)	(N=31)

arate evaluation mode but not in the joint evaluation mode. In contrast, higher-performing candidates were only favored in joint but not in separate evaluation. Columns 3 and 4 include controls for the risk attitudes and the gender of the evaluator. Male and more risk- tolerant evaluators were less likely than female and more risk-averse evaluators to select a given candidate than to go for the random option. Both of these results accord with the intuition that individuals liking risk more should be more likely to participate in the lottery of random candidate assignment than people liking risk less.

4.4.3. Alternative explanation: taste-based discrimination

To test whether choices of evaluators were indeed based on biased expectations of future performance rather than on a preference for men for math and women for verbal tasks (taste-based discrimination), we ran a control experiment with an additional 110 subjects where evaluators were informed of their candidate's current performance and asked whether they wanted to hire that candidate for the current round or be allocated a random person from the pool. In our control experiments, we replicated the separate-evaluation condition for higher-performing candidates to

Table 4.3.: The Effect of past performance and Stereotypes on Candidate Selection, Marginal Effects at Mean

	(1) Separate	(2) Joint	(3) Separate plus controls	(4) Joint plus controls
First-round performance	0.099	0.462**	0.117	0.472***
	(0.07)	(0.06)	(0.07)	(0.06)
Stereotype-Advantage	0.165**	0.009	0.164**	0.008
	(0.07)	(0.07)	(0.07)	(0.07)
Math	-0.009	-0.043	0.018	-0.040
	(0.07)	(0.05)	(0.07)	(0.05)
Risk Tolerance			-0.059***	-0.002
			(0.02)	(0.01)
Male Evaluator			-0.099	-0.199***
			(0.07)	(0.05)
Decision Outcomes	202	252	202	252
R^2	0.0271	0.2201	0.0664	0.2579

Notes: Each specification is a Probit regression, marginal effects reported in percentage points. The dependent variable in the separate treatment is the selection of a given candidate. In the joint treatment we score two outcomes for each individual: namely, whether the employer selected the higher (1) or the lower (2) performer: This implies a total of 252 outcomes. Robust standard errors are in brackets and adjusted for clustering at the employer level. Risk tolerance is measured by the numberof risky choices made in a lottery (identical to Holt and Laury (2002). *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

examine whether the focus on group characteristics in our first set of experiments was driven by stereotypical beliefs about group performance in the two tasks or by an (implicit) distaste of female candidates for math tasks and male candidates for verbal tasks.

We did not find any evidence for taste-based discrimination in our experiment. Across the two tasks, the likelihood that a member of the stereotype-advantaged group was chosen was 0.46, and the likelihood that a member of the stereotype-disadvantaged group was chosen was 0.43. Specifically, instead of going back to the pool, in the math task (N=50), 35 percent of the evaluators chose the male candidate and 41 percent chose the female candidate; in the verbal task (N=60), 55 percent chose the male and 56 percent the female candidate. None of these differences are significant; women and men were just as likely to be chosen for both tasks.

4.5. Conclusions

This paper examines whether an "evaluation nudge," namely evaluating candidates jointly rather than separately, can overcome gender-biased assessments of job candidates that prefer men for male-typed tasks and women for female-typed tasks, even if gender is not predictive of future performance and more reliable individual performance measures are available. We find that when evaluators are tasked with choosing a candidate for future performance in a math or a verbal task, a joint-evaluation mode helps them focus on individual performance, irrespective of candidates' gender and evaluator bias: evaluators were significantly more likely to choose the higher- rather than the lower-performing candidate in this mode. In contrast, in separate evaluation, evaluators were heavily influenced by a candidate's gender, even though gender was not predictive of future performance and individual past performance was: they were significantly more likely to choose men for the math task and women for the verbal task.

Joint evaluation may affect choices by providing additional data that evaluators can use to update their stereotypical beliefs about a group to which a candidate belongs. By definition, an evaluator has more data points available in joint than in separate evaluation. If these data points provide counter-stereotypical information, they may shift an evaluator's beliefs about the group enough to make him or her choose counter-stereotypically.

Our work is in line with extensive work in behavioral decision making suggesting that people may evaluate products differently in joint than in separate evaluation. This research attributed differences in decision outcomes to a switch in judgment modes from a more intuitive mode based on heuristics in separate evaluation to a more reasoned mode when comparing alternatives in joint evaluation (Gino et al., 2011; Kassam et al., 2009; Bazerman and Moore, 2012).

Our findings have implications for organizations that want to decrease the likelihood that hiring, promotion, and job-assignment decisions will be based on irrelevant criteria triggered by stereotypes. Joint evaluation is common for many hiring decisions but rare for job assignments and for promotion decisions. Organizations concerned about discrimination in this later phase might want to review how, for example, career-relevant jobs are assigned or how promotion decisions are made. According to the Corporate Gender Gap Report 2010(Zahidi and Ibarra, 2010), in most countries, fewer than 10 percent of career-relevant jobs are held by women. In many academic fields, including economics, controlling for performance, women are less likely to be granted tenure than men (Ginther and Kahn, 2004; Kahn and Ginther, 2006).

Organizations can move from separate-evaluation to joint-evaluation procedures to promote more accurate decision-making and maximize performance. In addition to being a profit-maximizing decision procedure, joint evaluation is also a fair mechanism, as it encourages judgments based on people's performance rather than their demographic characteristics. Companies concerned about discrimination might choose to review how job candidates are evaluated, how jobs are assigned and promotion decisions made. Our work suggests that organizations can nudge evaluators toward taking individual performance information rather than gender stereotypes into account.

A. Appendix chapter 2

A.1. Experimental Instructions

A.1.1. Dospert Questionnaire

Question 1

For each of the following statements, please indicate your likelihood of engaging in each activity or behavior. Provide a rating from 1 to 5, using the following scale: 1 (very unlikely) 5 (very likely)

1. Admitting that your tastes are different from those of your friends. (S)
2. Going camping in the wilderness, beyond the civilization of a campground. (R)
3. Betting a day's income at the horse races. (G)
4. Buying an illegal drug for your own use. (H)
5. Cheating on an exam. (E)
6. Chasing a tornado or hurricane by car to take dramatic photos. (R)
7. Investing 10% of your annual income in a moderate growth mutual fund. (I)
8. Consuming five or more servings of alcohol in a single evening. (H)

9. Cheating by a significant amount on your income tax return. (E)
10. Disagreeing with your father on a major issue. (S)
11. Betting a day's income at a high stake poker game. (G)
12. Having an affair with a married man or woman. (E)
13. Forging somebody's signature. (E)
14. Passing off somebody else's work as your own. (E)
15. Going on a vacation in a third-world country without prearranged travel and hotel accommodations. (R)
16. Arguing with a friend about an issue on which he or she has a very different opinion. (S)
17. Going down a ski run that is beyond your ability or closed. (R)
18. Investing 5% of your annual income in a very speculative stock. (I)
19. Approaching your boss to ask for a raise. (S)
20. Illegally copying a piece of software. (E)
21. Going whitewater rafting during rapid water flows in the spring. (S)
22. Betting a day's income on the outcome of a sporting event (e.g. baseball, soccer, or football). (G)
23. Telling a friend if his or her significant other has made a pass at you. (S)
24. Investing 5% of your annual income in a conservative stock. (I)
25. Shoplifting a small item (e.g. a lipstick or a pen). (E)

26. Wearing provocative or unconventional clothes on occasion. (S)

27. Engaging in unprotected sex. (H)

28. Stealing an additional TV cable connection off the one you pay for. (E)

29. Not wearing a seatbelt when being a passenger in the front seat. (H)

30. Investing 10% of your annual income in government bonds (treasury bills). (I)

31. Periodically engaging in a dangerous sport (e.g. mountain climbing or sky diving). (S)

32. Not wearing a helmet when riding a motorcycle. (H)

33. Gambling a week's income at a casino. (G)

34. Taking a job that you enjoy over one that is prestigious but less enjoyable. (S)

35. Defending an unpopular issue that you believe in at a social occasion. (S)

36. Exposing yourself to the sun without using sunscreen. (H)

37. Trying out bungee jumping at least once. (R)

38. Piloting your own small plane, if you could. (R)

39. Walking home alone at night in a somewhat unsafe area of town. (H)

40. Regularly eating high cholesterol foods. (H)

Question 2

People often see some risk in situations that contain uncertainty about what the outcome or consequences will be and for which there is the possibility of negative consequences. However, riskiness is a very personal and intuitive notion, and we are interested in your gut level assessment of how risky each situation or behavior is. For each of the following statements, please indicate how risky you perceive each situation to be. Provide a rating from Not at all Risky to Extremely Risky, using the following scale: 1 (very unlikely) 5 (very likely)

Question 3

The actions described in each of these statements can lead to a good outcome. For each of the following statements, please indicate the benefits you would obtain from each situation in case of a good outcome Provide a rating from 1 to 5, using the following scale: 1 (no high benefit) 5 (very high benefit)

Question 4

The actions described in each of these statements can lead to a bad outcome. For each of the following statements, please indicate the costs you would obtain from each situation in case of a bad outcome. Provide a rating from 1 to 5, using the following scale: 1 (not costly) 5 (very costly).

Question 5

For each of the following statements, please indicate the probability you will obtain the good outcome. Provide a rating from 0 to 1, using the following scale: 0 (zero probability), 0.1 (10 % probability), 1 (certainty.)

A.1.2. Generalized Anxiety Disorder Test (GAD)

1. Do you experience excessive worry? Yes No
2. Is your worry excessive in intensity, frequency, or amount of distress it causes? Yes No
3. Do you find it difficult to control your worry (or stop worrying) once it starts? Yes No

4. Do you worry excessively and uncontrollably about minor things such as being late for an appointment, minor repairs, homework, etc.? Yes No

5. Please list the most frequent topics about which you worry excessively and uncontrollably: a. d. b. e. c. f.

6. During the last six months~ have you been bothered by excessive and uncontrollable worries more days than not? Yes No

   IF YES,CONTINUE. IF NO, SKIP REMAINING QUESTIONS.

7. During the past six months, have you often been bothered by any of the following symptoms? Place a check next to each symptom that you have had more days than not:

   ___ Restlessness or feeling keyed up or on edge

   ___ Difficulty falling/staying asleep or restless/unsatisfying sleep

   ___ Difficulty concentrating or mind going blank

   ___ Irritability

   ___ Being easily fatigued

   ___ Muscle tension

8. How much do worry and physical symptoms interfere with your life, work, social activities, family, etc.?

   Circle one number: 012345678, where 0 Is none, 2 is mildly, 4 is moderately, 6 Is severely and and 8 is very severely 9.

9. How much are you bothered by worry and physical symptoms (how much distress does it cause you)?

Circle one number: 012345678, where 0 Is no distress, 2 is mild distress, 4 is moderate distress, 6 Is severe distress and 8 is very severe distress

A.1.3. Cognitive Reflection and Risk Proficiency

Cognitive Reflection Test

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

Risk Proficiency Test

1. Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)

   a) 5 out of 50 throws

   b) 25 out of 50 throws

   c) 30 out of 50 throws

   d) None of the above

2. Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent

a) 10%

b) 25%

c) 40%

d) None of the above

3. Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, about how many times would the die show the number 6?

A.1.4. Lottery Based Risk Attitude Elicitation Tasks

Figure A.1.: Lottery List Decision Task

The assigned risk yields $2 with 50% and $30 with 50% probability

LOTTERY LIST QUESTIONS

GAMBLE 2

This gamble gives you a 75% chance of gaining $ 15.00 and a 25% chance of gaining $ 3.33 instead.

Option A			Option B (in $)	
Take the gamble	OR	gain	3.33	option A option B
Take the gamble	OR	gain	5.28	option A option B
Take the gamble	OR	gain	7.22	option A option B
Take the gamble	OR	gain	9.17	option A option B
Take the gamble	OR	gain	11.11	option A option B
Take the gamble	OR	gain	13.06	option A option B
Take the gamble	OR	gain	15.00	option A option B

OK

Hypothetical Gamble Questions

1. Suppose that a distant relative left you a share in a private business worth one million dollars. You are immediately faced with a choice — whether to cash out now and take the one million dollars, or to wait until the company goes public in one month, which would give you a 50–50 chance of doubling your money to two million dollars and a 50–50 chance of losing one-third of it, leaving you 667 thousand dollars. Would you cash out immediately or wait until after the company goes public?

   a) Cash out

   b) Wait

   c) Do not know

2. Suppose that waiting a month, until after the company goes public, would result in a 50–50 chance that the money would be doubled to two million dollars and a 50– 50 chance that it would be reduced by half, to 500 thousand dollars. Would you cash out immediately and take the one million dollars, or wait until the company goes public?

   a) Cash out

   b) Wait

   c) Do not know

3. Suppose the chances were 50–50 that waiting would double your money to two million dollars and 50–50 that it would reduce it by seventy-five percent, to 250 thousand dollars. Would you cash out immediately and take the one million dollars, or wait until after the company goes public?

   a) Cash out

b) Wait

c) Do not know

4. Suppose that waiting a month, until after the company goes public, would result in a 50–50 chance that the money would be doubled to two million dollars and a 50– 50 chance that it would be reduced by twenty percent, to 800 thousand dollars. Would you cash out immediately and take the one million dollars, or wait until after the company goes public?

   a) Cash out

   b) Wait

   c) Do not know

5. Suppose the chances were 50–50 that waiting would double your money to two million dollars and 50–50 that it would reduce it by ten percent, to 900 thousand dollars. Would you cash out immediately and take the one million dollars, or wait until after the company goes public?

   a) Cash out

   b) Wait

   c) Do not know

Holt and Laury Task

Figure A.2.: HL Decision Task

Decision	Option A		Option B	
1	10% chance of $8	90% chance of $6.4	10% chance of $15.4	90% chance of $ 0.4
2	20% chance of $8	80% chance of $6.4	20% chance of $15.4	80% chance of $ 0.4
3	30% chance of $8	70% chance of $6.4	30% chance of $15.4	70% chance of $ 0.4
4	40% chance of $8	60% chance of $6.4	40% chance of $15.4	60% chance of $ 0.4
5	50% chance of $8	50% chance of $6.4	50% chance of $15.4	50% chance of $ 0.4
6	60% chance of $8	40% chance of $6.4	60% chance of $15.4	40% chance of $ 0.4
7	70% chance of $8	30% chance of $6.4	70% chance of $15.4	30% chance of $ 0.4
8	80% chance of $8	20% chance of $6.4	80% chance of $15.4	20% chance of $ 0.4
9	90% chance of $8	10% chance of $6.4	90% chance of $15.4	10% chance of $ 0.4
10	100% chance of $8	0% chance of $6.4	100% chance of $15.4	0% chance of $ 0.4

A.2. Additional Experimental Results

A.2.1. Relating the risk attitude elicitation tasks

Table A.1.: Gender differences in the Inheritance Task

	Male	Female
Baseline Category	2.308***	3.077***
	(0.259)	(0.305)
Treatment controls	Yes	Yes
Gender*treatment controls	Yes	Yes

Estimates of baseline category in linear regression onre-categorized data. Robust standard. errors in brackets. Males are lessrisk averse (p=0.06.)*** Significance at the 1 percent level.

Table A.2.: Average Self-reported Risk Preferences

	Male	Female
Risk aversion	3.351	3.756
	(1.078)	(1.291)
Loss aversion	3.621	4.090
	(1.224)	(1.291)

Males are significantly less risk averse (Wilcoxon rank sum test, p = 0.04) and significantly less loss averse (Wilcoxonrank sum test, p = 0.03.) Standard deviations in brackets.

Table A.3.: Relation Lottery Tasks and Risk Taking in Dospert Task

	(1)	(2)	(3)	(4)	(5)
Male	0.29**	0.27**	0.67	0.35	0.08
	(0.111)	(0.112)	(0.421)	(0.434)	(0.098)
Social context	1.13***	1.13***	1.13***	0.29***	0.28***
	(0.076)	(0.076)	(0.076)	(0.082)	(0.081)
Recreation context	0.04	0.04	0.04	-0.38***	-0.38***
	(0.101)	(0.101)	(0.101)	(0.087)	(0.087)
Gambling context	-0.96***	-0.96***	-0.96***	-1.28***	-1.28***
	(0.097)	(0.097)	(0.097)	(0.101)	(0.102)
Ethical context	-0.50***	-0.50***	-0.50***	-0.79***	-0.79***
	(0.065)	(0.065)	(0.065)	(0.062)	(0.062)
Investment context	0.40***	0.40***	0.40***	-0.50***	-0.51***
	(0.113)	(0.113)	(0.114)	(0.104)	(0.104)
Social context	-0.38**	-0.38**	-0.38**	-0.17	-0.17
x male	(0.120)	(0.120)	(0.120)	(0.110)	(0.110)
Recreation context	0.05	0.05	0.05	0.05	0.05
x male	(0.145)	(0.145)	(0.145)	(0.121)	(0.121)
Gambling context	0.20	0.20	0.20	0.22	0.22
x male	(0.159)	(0.159)	(0.159)	(0.144)	(0.144)
Ethics context	-0.33**	-0.33**	-0.33**	-0.23**	-0.23**
x male	(0.102)	(0.102)	(0.102)	(0.094)	(0.094)
Investment context	0.23	0.23	0.23	0.32**	0.32**
x male	(0.177)	(0.177)	(0.177)	(0.157)	(0.157)
LLT positive		-0.01	-0.01	-0.01**	
domain		(0.007)	(0.009)	(0.007)	
LLT negative		-0.01	-0.01	-0.00	
domain		(0.007)	(0.009)	(0.008)	
LLT mixed		0.00	0.01	0.01	
domain		(0.007)	(0.008)	(0.007)	
Inheritance task		-0.03	0.00	0.00	
		(0.034)	(0.046)	(0.037)	
LLT positive			0.01	0.02*	
domain x male			(0.014)	(0.012)	
LLT negative			-0.01	-0.01	
domain x male			(0.013)	(0.013)	
LLT mixed			-0.02	-0.02	
domain x male			(0.015)	(0.013)	
Inheritance task			-0.06	-0.04	
x male			(0.077)	(0.062)	
Probability of high outcome				0.21***	0.21***
				(0.018)	(0.018)
Value of low outcome				-0.18***	-0.18***
				(0.022)	(0.022)
Prob of high outcome				1.63***	1.62***
				(0.091)	(0.093)
Constant	2.38***	2.70***	2.56***	2.32***	2.24***
	(0.069)	(0.212)	(0.244)	(0.261)	(0.130)
N	6400	6400	6400	6400	6400

Dependent Variable: Likelihood to take risk. Robust (clustered) standard errors in brackets. ***<0.001; **<0.05 *<0.01

A.2.2. Relating personal characteristics

Table A.4.: Gender differences in background characteristics

Characteristics	Female	Male
GAD	5.15***	3.77***
	(0.06)	(0.04)
CRT	1.80***	2.05***
	(0.01)	(0.01)
Risk numeracy	1.73***	2.17***
	(0.01)	(0.01)
High exp. income	0.67***	0.87***
	(0.01)	(0.00)
Optimist	0.82***	0.92***
	(0.00)	(0.00)
Confidence	5.06***	6.17***
	(0.02)	(0.02)
Lucky	0.66***	0.79***
	(0.01)	(0.00)

Standard errors of means in parentheses * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table A.5.: The effect of personal characterstics on risk taking in the inheritance task

	Inheritance task
Male	-0.713*
	(0.404)
GAD	0.035
	(0.026)
CRT	-0.039
	(0.123)
Risk Numeracy	-0.195
	(0.125)
High expected Income	-0.056
	(0.282)
Optimist	0.226
	(0.323)
Lucky	0.174
	(0.242)
Confidence	-0.061
	(0.074)
Constant	3.382***
	(0.674)

Robust (clustered) standard errors in brackets. ***<0.001; **<0.05 *<0.01

A.2.3. Lottery-based risk attitudes and the Dospert task by context

Table A.6.: Lottery Tasks and Dospert by context, Part 1

	Social context					Gamble context					Health context				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Male	-0.09	-0.11	0.49	0.56	-0.11	0.49***	0.48**	1.21	0.47	0.34**	0.29**	0.29**	-0.90	-1.01	0.04
	(0.082)	(0.083)	(0.501)	(0.408)	(0.075)	(0.140)	(0.146)	(0.849)	(0.762)	(0.129)	(0.111)	(0.114)	(0.720)	(0.809)	(0.097)
LLT positive		-0.00	0.01	-0.00			-0.03*	-0.04	-0.04*			0.01	0.01	0.02	
domain		(0.008)	(0.010)	(0.009)			(0.016)	(0.022)	(0.018)			(0.012)	(0.016)	(0.013)	
LLT negative		0.00	0.00	0.01			-0.02	-0.02	-0.02*			-0.01	-0.02	-0.01	
domain		(0.008)	(0.013)	(0.009)			(0.013)	(0.014)	(0.012)			(0.010)	(0.013)	(0.011)	
LLT mixed		0.01	0.00	0.00			0.01	0.02	0.02			-0.01	-0.02	-0.02	
domain		(0.008)	(0.010)	(0.009)			(0.014)	(0.019)	(0.016)			(0.011)	(0.014)	(0.011)	
Inheritance task		-0.02	0.01	0.02			0.01	0.06	0.01			-0.01	-0.04	-0.04	
		(0.037)	(0.058)	(0.052)			(0.057)	(0.058)	(0.061)			(0.048)	(0.056)	(0.057)	
LLT positive			-0.03*	-0.02				0.01	0.03				0.01	0.02	
domain x male			(0.015)	(0.014)				(0.031)	(0.027)				(0.023)	(0.020)	
LLT negative			-0.00	-0.01				-0.00	0.01				0.03	0.02	
domain x male			(0.017)	(0.013)				(0.024)	(0.022)				(0.022)	(0.024)	
LLT mixed			0.01	0.00				-0.04	-0.04				0.02	0.01	
domain x male			(0.017)	(0.015)				(0.028)	(0.025)				(0.025)	(0.022)	
Inheritance task			-0.06	-0.05				-0.09	-0.06				0.10	0.10	
x male			(0.078)	(0.070)				(0.129)	(0.122)				(0.108)	(0.096)	
Value of high outcome				0.05	0.04				0.04	0.06*				0.41***	0.41***
				(0.029)	(0.029)				(0.038)	(0.035)				(0.047)	(0.047)
Value of low outcome				-0.14***	-0.14***				-0.11**	-0.12**				-0.24***	-0.24***
				(0.035)	(0.036)				(0.052)	(0.053)				(0.045)	(0.048)
Prob of high outcome				1.41***	1.41***				1.71***	1.66***				1.39***	1.35***
				(0.159)	(0.163)				(0.318)	(0.324)				(0.160)	(0.164)
Constant	3.50***	3.44***	3.21***	2.70***	3.01***	1.41***	2.18***	1.92***	1.95***	1.23***	2.37***	2.51***	2.95***	2.41***	2.23***
	(0.057)	(0.249)	(0.319)	(0.274)	(0.163)	(0.085)	(0.434)	(0.490)	(0.480)	(0.270)	(0.069)	(0.345)	(0.443)	(0.435)	(0.243)
N	1280	1280	1280	1280	1280	640	640	640	640	640	1280	1280	1280	1280	1280

Dependent Variable: Likelihood to take risk. Robust (clustered) standard errors in brackets. ***<0.001; **<0.05 *<0.01

Table A.7.: Lottery Tasks and Dospert by context, Part 2

	Ethical context					Recreation context					Investment context				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Male	-0.04	-0.01	0.34	-0.00	-0.14*	0.34**	0.30**	1.22	1.08	0.08	0.52***	0.38**	2.30**	1.49**	0.44***
	(0.101)	(0.104)	(0.586)	(0.569)	(0.082)	(0.141)	(0.144)	(0.815)	(0.823)	(0.108)	(0.146)	(0.152)	(0.886)	(0.717)	(0.133)
LLT positive		-0.00	-0.01	-0.00			-0.02	-0.04**	-0.04***			-0.02	-0.02	-0.03*	
domain		(0.009)	(0.013)	(0.009)			(0.014)	(0.016)	(0.011)			(0.015)	(0.018)	(0.017)	
LLT negative		-0.01	-0.01	-0.01			-0.02	-0.01	0.00			0.02	0.02	-0.00	
domain		(0.009)	(0.012)	(0.011)			(0.013)	(0.018)	(0.017)			(0.014)	(0.022)	(0.016)	
LLT mixed		-0.00	0.01	0.01			-0.00	0.01	0.02*			0.02	0.05**	0.03	
domain		(0.010)	(0.014)	(0.011)			(0.015)	(0.016)	(0.014)			(0.016)	(0.020)	(0.018)	
Inheritance task		0.04	0.06	0.04			-0.05	-0.01	0.03			-0.20**	-0.08	-0.05	
		(0.049)	(0.072)	(0.058)			(0.070)	(0.093)	(0.067)			(0.060)	(0.083)	(0.072)	
LLT positive			0.01	0.03*				0.04	0.04*				0.02	0.02	
domain x male			(0.017)	(0.014)				(0.027)	(0.020)				(0.028)	(0.025)	
LLT negative			-0.00	-0.00				-0.04	-0.03				-0.02	-0.00	
domain x male			(0.018)	(0.017)				(0.024)	(0.022)				(0.027)	(0.021)	
LLT mixed			-0.02	-0.03				-0.03	-0.04*				-0.07**	-0.06**	
domain x male			(0.021)	(0.017)				(0.029)	(0.023)				(0.029)	(0.028)	
Inheritance task			-0.04	-0.01				-0.14	-0.10				-0.23*	-0.17	
x male			(0.106)	(0.084)				(0.149)	(0.102)				(0.124)	(0.109)	
Value of high outcome				0.18***	0.17***				0.36***	0.36***				0.21***	0.22***
				(0.029)	(0.030)				(0.044)	(0.045)				(0.045)	(0.045)
Value of low outcome				-0.23***	-0.23***				-0.17***	-0.18***				-0.03	-0.06
				(0.034)	(0.033)				(0.038)	(0.039)				(0.049)	(0.055)
Prob of high outcome				1.40***	1.35***				1.88***	1.87***				2.07***	2.10***
				(0.156)	(0.155)				(0.185)	(0.191)				(0.220)	(0.231)
Constant	1.87***	2.02***	1.89***	1.75***	1.85***	2.41***	3.33***	3.03***	1.32*	1.29***	2.77***	2.99***	2.30***	1.18**	1.01**
	(0.073)	(0.306)	(0.389)	(0.376)	(0.163)	(0.094)	(0.459)	(0.517)	(0.736)	(0.240)	(0.102)	(0.512)	(0.602)	(0.567)	(0.335)
N	1280	1280	1280	1280	1280	1280	1280	1280	1280	1280	640	640	640	640	640

Dependent Variable: Likelihood to take risk. Robust (clustered) standard errors in brackets. ***<0.001; **<0.05 *<0.01

B. Appendix chapter 4

B.1. Information-based Model

We show that under certain conditions, evaluators will choose the higher-performing candidate in joint but not in separate evaluation. In joint evaluation, they may observe more counter-stereotypical candidates (e.g., women performing strongly and men poorly in math) than in separate evaluation. The evaluator is either in a joint or separate evaluation condition and has to choose a candidate for future performance. The evaluator is informed of selected candidate(s)' past performance and gender (type), as well as the total size (and average performance of the candidate pool. Each candidate i has a type $g \in (m, f)$, and type m is believed to have higher expected performance. We define x_{ig} as the observed past performance of candidate i with type g, and let y_{ig} denote this candidate's (un(known) future performance. The total size of the candidate pool is known to be $N = 2j$ with $x_{im} \in (x_1 ... x_j)$ and $x_{if} \in (x_{j+1} ... x_{2j})$, reflecting a 50/50 distribution of types. The average previous performance across types, i.e $\bar{x} = \frac{1}{N} \sum_{g=f}^{m} \bar{x}_g$,is known. We denote $\bar{x}_g$ as the (unknown) average previous performance by gender. Evaluators know $x_g \sim N(\bar{x}_g, \sigma^2)$, with known variance ($\sigma^2$).

In joint evaluation, evaluators observe the performance of two candidates (i.e., they observe $\mathbf{x} = (x_{if} \ x_{im})$). Evaluators can choose between these candidates and the random option with known expected performance of $\bar{x}$.

We denote evaluators' prior expectation of $\bar{y_g}$ as μ_g. Evaluators have no taste for discrimination but hold stereotypical beliefs. Specifically, $\mu_m > \mu_f$, and because of the 50/50 distribution and known $\bar{x}$, we can write $\mu_m = \bar{x} + h$,and $\mu_f = \bar{x} - h$ with $h > 0$ (assuming no learning of the task for candidates over time). Evaluators' prior beliefs about each type's average future productivity look as follows: $\theta_g \sim N(\mu_g, \nu^2)$, assuming equal, positive, and known variances across the genders. Because of symmetry, we have $(|\theta_g - \bar{x}|) \sim N(h, \nu^2)$.

We assume that evaluators are risk-neutral expected-value maximizers and that the expected future performance (y_{ig}) is a linear combination of the observed previous performance of the candidate and the (updated) belief about the candidate's performance based on gender; i.e., $E(y_{ig}|x) = \alpha x_{ig} + (1-\alpha)\mu_g(.|\mathbf{x})$ with $0 < \alpha < 1$.

After observing x, evaluators update beliefs on according to Bayes rule, to the posterior distribution of $\mu_g(.|\mathbf{x})$ over θ_g. An evaluator in separate evaluation confronted with a below-average male candidate (i.e. with $x_m = b$) will update beliefs about mean performance for males in the pool to: $\mu_m(.|\mathbf{x}) = (\bar{x} + h|x_{im})) = b + \frac{\sigma^2}{\sigma^2+\nu^2}(h + (\bar{x} + b))$.

If faced with an average female candidate (i.e. with $x_f = \mu$), an evaluator will update beliefs about the mean performance of females in the pool to: $\mu_f(.|\mathbf{x}) = (\bar{x} - h|x_{if})) = \bar{x} - \frac{\sigma^2}{\sigma^2+\nu^2}h$.

In joint evaluation, evaluators have two data points available; they use both the male and the female candidates' past performance to update their prior of h. In the counter-stereotypical situation where an evaluator is confronted with a lower-performing male ($x_m = b$) and a higher-performing female candidate ($x_f = \mu$), this results in updated beliefs of mean performance for males in the pool of: $\mu_m(.|\mathbf{x}) = (\bar{x} + h|x_{im}, x_{if})) = \frac{(\bar{x}+h)\sigma^2+\nu^2(\bar{x}+b)}{\sigma^2+2\nu^2}$.

It results in an updated mean for females in the pool of:$\mu_f(.|\mathbf{x}) = (\bar{x}-h|x_{im}, x_{if})) =$

$\frac{(\bar{x}-h)\sigma^2+\nu^2((2\bar{x}-b)+\bar{x})}{\sigma^2+2\nu^2}$.

The evaluator will compare the updated expected future performance $E(y_{ig}|\mathbf{x})$ of the candidates with the expected value of the random option ($\bar{x}$) and choose the option with the highest expected performance. To have a situation where evaluators prefer a high-performing female in a joint evaluation setting over a low-performing male and the random option, but prefer the random option over the high-performing female in the separate evaluation setting, the following conditions need to simultaneously hold:

1. The expected future performance of a higher-performing female candidate dominates the random option.

2. The expected future performance of a higher-performing female candidate dominates the lower-performing male option in joint evaluation, and

3. The expected future performance of a higher-performing female candidate is lower than the expected value of the random option in the separate treatment.

If $\frac{h\sigma^2}{\nu^2} < (\bar{x} - b)$ these conditions hold:

1. $(1-\alpha) * \bar{x} + \alpha * \frac{(\bar{x}-h)\sigma^2+\nu^2((2\bar{x}-b)+\bar{x})}{\sigma^2+2\nu^2} > \bar{x}$.

2. $(1-\alpha) * \bar{x} + \alpha * \frac{(\bar{x}-h)\sigma^2+\nu^2((2\bar{x}-b)+\bar{x})}{\sigma^2+2\nu^2} > (1-\alpha) * b + \alpha * \frac{(\bar{x}+h)\sigma^2+\nu^2(\bar{x}+b)}{\sigma^2+2\nu^2}$.

3. $\bar{x} > (1-\alpha) * \bar{x} + \alpha * (\bar{x} - \frac{h\sigma^2}{\sigma^2+\nu^2})$

Thus, whenever there is sufficient variance of the expected difference between male and female performance, there is enough counter-stereotypical evidence, and evaluators are not too biased, the joint-separate reversal can be observed. For the parameters used in the experiment, the condition reduces to $h\sigma^2 < \nu^2$.

B.2. Experimental Instructions

B.2.1. Instructions stage 1

All treatments were programmed in Z-tree. (Fischbacher 2000) We include as an example our instructions for the math task (inspired by Niederle and Vesterlund 2007), instructions for the word task were similar and are available upon request.

Welcome!

In this experiment you are asked to correctly solve as many math problems as possible. In each problem, you are asked to sum up five two-digit numbers.

For each correct answer you will receive 25 cents. There will be three rounds; each round consists of 15 problems. You have five minutes available for each round.

Before we begin with the experiment there will be a practice round where you can get used to the task.

At the end of the experiment, you will receive an overview of the number of correct answers and of your total payoff. An example of this task is given in the figure below.

After performing the task, participants filled out a questionnaire collecting demographic information.

B.2.2. Instructions Stage 2

These were the instructions for the joint treatment with the math task and a higher performing female candidate and a low performing male. Instructions for the other treatments were similar and are available upon request.

Welcome!

You are participating in a study in which you will earn some money. The amount will depend on a choice that you will have to make below. At the end of the study, your earnings (1 point = $ 1) will be added to a show-up fee, and you will be paid in cash.

Figure B.1.: Lottery List Decision Task

The assigned risk yields $2 with 50% and $30 with 50% probability

LOTTERY LIST QUESTIONS

GAMBLE 2

This gamble gives you a 75% chance of gaining $ 15.00 and a 25% chance of gaining $ 3.33 instead.

Option A			Option B (in $)	
Take the gamble	OR	gain	3.33	option A ◉ ○ option B
Take the gamble	OR	gain	5.28	option A ◉ ○ option B
Take the gamble	OR	gain	7.22	option A ◉ ○ option B
Take the gamble	OR	gain	9.17	option A ◉ ○ option B
Take the gamble	OR	gain	11.11	option A ○ ◉ option B
Take the gamble	OR	gain	13.06	option A ○ ◉ option B
Take the gamble	OR	gain	15.00	option A ○ ◉ option B

OK

Your Choice.

Another group of study participants has participated in two rounds of a task before this session. You will receive information on two of the participants, person A and person B and on how well they performed in Round 1. You will then have to decide whether you want to be paid according to the Round 2 performance of person A, person B or of a randomly selected person from the pool of participants.

Information on Task

In a previous study, participants were shown rows of five two-digit numbers. Participants had to add up the numbers of each row. Participants were asked and incentivized to add up as many rows as possible as possible. They had five minutes available for each round of the task. While the task was otherwise identical, they saw different sequences containing different numbers in Rounds 1 and 2.

Their point score was calculated as follows: For every correctly added sequence they received one point. Sequences that were not correctly added received no points.

To have a better understanding of the task, please click on this button to see a sample task (SAMPLE TASK)

Information on Average Round-1 Performance of all Study Participants

On average participants scored 10 points in Round 1.

Information on Persons.

You will be paid according to the Round 2-performance of one of the two study participants described below, Person A or Person B, or a study participant drawn at random from all the people who participated in the study. We had 40 male and 40 female students participate, recruited by the Harvard Decision Science Laboratory.

Person A	**Person B**
Student	Student
American	American
Female	Male
Caucasian	Caucasian
Performance Indicator: In Round 1 the person scored 10 points in 5 minutes	Performance Indicator: In Round 1 the person scored 9 points in 5 minutes

Procedure to Determine your Earnings.

Once you have decided whether you want to be paid based on the performance of person A, person B or a randomly selected person and have completed a short questionnaire, we will inform you of their point score and your payoffs.

If you chose to be paid according to the performance of one of the persons described above, you will receive $1 x that person's point score for Round 2.

If you chose to be paid according to the performance of a random person, you will receive $1 x the random person's point score for Round 2.

For example if your chosen person scores 2 points in round 2, you will receive $2.

If you have any questions, please press the help button now. Once we have addressed all questions, we will move to the main question of this study.

Main question: Do you want to be paid based on the Round 2-performance of one of the persons described above, or do you want to be paid based on the Round 2-performance of a person drawn at random from all the people who participated in the study? (Please check one box)

NOTE: THE AVERAGE SCORE IN ROUND 1 WAS 10 POINTS

Person A	**Person B**	**Random Draw**
Student American Female Caucasian Performance Indicator: In Round 1 the person scored 10 points in 5 minutes	Student American Male Caucasian Performance Indicator: In Round 1 the person scored 9 points in 5 minutes	

Note: after the main question of the experiment participants were notified of the score of the randomly selected candidate, the score of person A, and the score of person B.

Additional Decision.

Please choose Option A or Option B in all ten paired lottery-choice decisions below (select your preferred option in each row). One of the pairs will be chosen at random, the lottery will be conducted and you will be paid according to the outcome of your preferred choice.

For example, if PAIR 1 (first row) is randomly chosen, and your preferred option is A, we will conduct a lottery where the chance of winning $2 is 1/10 (1 blue ball in an urn containing 10 balls) and the chance of winning $1.6 is 9/10 (9 green balls in the urn). If the blue ball is picked, you will receive \$2. If the green ball is picked, you will receive \$1.6.

Figure B.2.: HL Decision Task

Decision	Option A		Option B	
1	10% chance of \$8	90% chance of \$6.4	10% chance of \$15.4	90% chance of \$ 0.4
2	20% chance of \$8	80% chance of \$6.4	20% chance of \$15.4	80% chance of \$ 0.4
3	30% chance of \$8	70% chance of \$6.4	30% chance of \$15.4	70% chance of \$ 0.4
4	40% chance of \$8	60% chance of \$6.4	40% chance of \$15.4	60% chance of \$ 0.4
5	50% chance of \$8	50% chance of \$6.4	50% chance of \$15.4	50% chance of \$ 0.4
6	60% chance of \$8	40% chance of \$6.4	60% chance of \$15.4	40% chance of \$ 0.4
7	70% chance of \$8	30% chance of \$6.4	70% chance of \$15.4	30% chance of \$ 0.4
8	80% chance of \$8	20% chance of \$6.4	80% chance of \$15.4	20% chance of \$ 0.4
9	90% chance of \$8	10% chance of \$6.4	90% chance of \$15.4	10% chance of \$ 0.4
10	100% chance of \$8	0% chance of \$6.4	100% chance of \$15.4	0% chance of \$ 0.4

Bibliography

Abdellaoui, M., Driouchi, A., and L'Haridon, O. (2011). Risk aversion elicitation: reconciling tractability and bias minimization. *Theory and Decision*, 71(1):63–80.

Agarwal, S., He, J., Sing, T. F., and Zhang, J. (2013). Gender and Household Financial Decision: Evidence from Personal Bankruptcy. *SSRN Electronic Journal*.

Anderson, L. R. and Mellor, J. M. (2009). Are risk preferences stable? Comparing an experimental measure with a validated survey-based measure. *Journal of Risk and Uncertainty*, 39(2):137–160.

Apicella, C., Dreber, A., and Mollerstrom, J. (2014). Salivary testosterone change following monetary wins and losses predicts future financial risk-taking. *Psychoneuroendocrinology*, (39):58–64.

Bagues, M. F. and Esteve-Volart, B. (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *The Review of Economic Studies*, 77(4):1301–1328.

Banaji, M. and Greenwald, A. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, 68(2):181–198.

Barber, B. M. and Odean, T. (2001). Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. *The Quarterly Journal of Economics*, 116(1):261–292.

Bardsley, N., Huang, M., and Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, 96(4):1069–1090.

Bazerman, M. H. and Moore, D. A. (2012). *Judgment in Managerial Decision Making, 8th Edition*. Wiley Global Education.

Becker, G. S. (1976). *The Economic Approach to Human Behavior*. University of Chicago Press.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2):94–98.

Blais, A. R., Weber, E. U., and Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4):263–290.

Bohnet, I. and Saidi, F. (2011). Informational Differences and Performance: Experimental Evidence. *Working Paper*.

Bonin, H., Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, 14(6):926–937.

Bosch-Domènech, A. and Silvestre, J. (2012). Measuring risk aversion with lists: a new bias. *Theory and Decision*, 75(4):465–496.

Brachinger, H. W., Schubert, R., Brown, M., and Gysler, M. (2000). Gender specific attitudes towards risk and ambiguity: an experimental investigation. *Working Paper*.

Cameron, L. and Shah, M. (2013). Risk-Taking Behavior in the Wake of Natural Disasters.

Cárdenas, J. C., Dreber, A., von Essen, E., and Ranehill, E. (2012). Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. *Journal of Economic Behavior and Organization*, 83(1):11–23.

Cashman-McGrath, L., Newman, M. G., Constantino, M. J., Zuellig, A. R., Kachin, K. E., Przeworski, A., and Erickson, T. (2002). Preliminary reliability and validity of the generalized anxiety disorder questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. *Behavior Therapy*, 33(2):215–233.

Coates, J. (2012). The Hour Between Dog and Wolf: How Risk-Taking Transforms Us, Body and Mind.

Cokely, E. T., Galesic, M., and Schulz, E. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7(1):25–47.

Cox, J. C. and Sadiraj, V. (2008). Risky Decisions in the Large and in the Small: Theory and Experiment. *Risk Aversion in Experiments*, 12:9–40.

Cox, J. C., Sadiraj, V., Vogt, B., and Dasgupta, U. (2007). Is There A Plausible Theory for Risky Decisions? *Experimental Economics Center Working Paper*, (5).

Croson, R. and Gneezy, U. (2009). Gender differences in risk preferences. 47(2):448–474.

Deck, C. and Jahedi, S. (2013). The Effect of Cognitive Load on Decision Making. *Working Paper*.

Deck, C. and Schlesinger, H. (2012). Consistency of higher order risk preferences. *CESifo Working Paper*, (4047).

Dobbin, F., Kalev, A., and Kelly, E. (2007). Diversity management in corporate America. *Contexts*, 6(4):21–28.

Duflo, E., Beaman, L., Topalova, P., Chattopadhyay, R., and Pande, R. (2009). Powerful Women: Does Exposure Reduce Bias? *The Quarterly Journal of Economics*, 124(4):1497–1540.

Duflo, E., Beaman, L., Topalova, P., and Pande, R. (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science*, 335(6068):582–586.

Ebert, S. and Wiesen, D. (2010). Joint measurement of risk aversion, prudence and temperance. (20).

Eckel, C. C. and Grossman, P. J. (2008). Men, Women and Risk Aversion: Experimental Evidence. pages 1061–1073. Elsevier.

Erev, I. and Haruvy, E. (2013). Learning and the economics of small decisions. In *The Handbook of Experimental Economics*.

Fehr Duda, H., Schubert, R., and de Gennaro, M. (2006). Gender, Financial Risk, and Probability Weights. *Theory and Decision*, 60(2-3):283–313.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, (10):171–178.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.

Gino, F., Schweitzer, M. E., Mead, N. L., and Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational behavior and human decision processes*, 115:191–203.

Ginther, D. K. and Kahn, S. (2004). Women in economics: moving up or falling off the academic career ladder? *Journal of Economic perspectives*, 18(3):193–214.

Gollier, C. (2001). *The economics of risk and time*. MIT Press, Cambridge, Mass.

Gollier, C. and Pratt, J. (1996). Risk Vulnerability and the Tempering Effect of Background Risk. *Econometrica*, 64(5):1109.

Greenwald, A. and Banaji, M. (2002). Math = male, me = female, therefore math $\neq$ me. *Journal of Personality and Social Psychology*, 83(1):44–59.

Guiso, L., Monte, F., and Sapienza, P. (2008). Diversity: Culture, gender, and math. *Science*, 320(5880):1164–1165.

Guiso, L. and Paiella, M. (2008). Risk Aversion, Wealth, and Background Risk. *Journal of the European Economic Association*, 6(6):1109–1150.

Gysler, M., Kruse, J. B., and Schubert, R. (2002). Ambiguity and gender differences in financial decision making: an experimental examination of competence and confidence effects.

Harris, C., Jenkins, M., and Glaser, D. (2006). Gender differences in risk assessment: Why do women take fewer risks than men. *Judgment and Decision Making*, 1(1):48–63.

Harrison, G. W. and Rutstroem, E. E. (2008). Risk Aversion in the Laboratory. *Research in Experimental Economics*, 12:41–196.

Heaton, J. and Lucas, D. (2000). Portfolio choice in the presence of background risk. *Economic Journal*, 110(460):1–26.

Heinemann, F. (2008). Measuring Risk Aversion and the Wealth Effect. *Risk Aversion in Experiments*, 12:293–313.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance*, 58(3):1009–1032.

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

Hsee, C. K., Blount, S., Loewenstein, G. F., and Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological bulletin*, 125(5):576–590.

Jacobsen, B., Lee, J., and Marquering, W. (2014). Gender Differences in Optimism and Asset Allocation. *Journal of Economic Behavior & Organization*.

Kahn, S. and Ginther, D. K. (2006). Does Science Promote Women? Evidence from Academia 1973-2001. *NBER Working Paper*.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Kahneman, D. and Lovallo, D. (1993). Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. 39(1):17–31.

Kahneman, D. and Tversky, A. (2000). *Choices, values, and frames.* Cambridge University Press, Cambridge.

Kassam, K. S., Paharia, N., Greene, J. D., and Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational behavior and human decision processes*, 109(2):134–141.

Kihlstrom, R. E., Romer, D., and Williams, S. (1981). Risk-Aversion with Random Initial Wealth. *Econometrica*, 49(4):911–920.

Kimball, M. (1993). Standard Risk Aversion. *Econometrica*, 61(3):589–611.

Koszegi, B. and Rabin, M. (2006). A Model of Reference -Dependent Preferences. *Quarterly Journal of Economics*, 121(4):1133–1165.

List, J. A., Harrison, G. W., and Towe, C. (2007). Naturally occurring preferences and exogenous laboratory experiments: A case study of risk aversion. *Econometrica*, 75(2):433–458.

Loewenstein, G. F. and Bazerman, M. H. (1992). Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives.

Loewenstein, G. F., Weber, E. U., and Hsee, C. K. (2001). Risk as feelings. *Psychological bulletin*, 127(2):267–286.

Lusk, J. L. and Coble, K. H. (2008). Risk aversion in the presence of background risk: Evidence from an economic experiment. pages 315–340. Research in Experimental Economics.

Miller, A. R. and Matsa, D. A. (2013). A Female Style in Corporate Leadership? Evidence from Quotas. *American Economic Journal: Applied Economics*, 5(3):136–169.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., and Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41):16474–16479.

Nagel, S. and Malmendier, U. (2011). Depression Babies: do Macroeconomic Experiences Affect Risk Taking? *Quarterly Journal of Economics*, 126(1):373–416.

Neumark, D., Van Nort, K. D., and Bank, R. (1996). Sex Discrimination in Restaurant Hiring: An Audit Study. *The Quarterly Journal of Economics*, 111(3):915–941.

Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science*, 59(1):1–16.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–11–1.

Nolen, P. and Booth, A. L. (2012). Gender differences in risk behaviour: does nurture matter? *The economic journal*, 122(558):56–78.

Norton, M. I., Vandello, J. A., and Darley, J. M. (2004). Casuistry and Social Category Bias. *Journal of Personality and Social Psychology*, 87(6):817–831.

Noussair, C. N. and Trautmann, S. (2011). Higher Order Risk Attitudes, Demographics, and Financial Decisions.

Nowlis, S. M. and Simonson, I. (1997). Attribute-Task Compatibility as a Determinant of Consumer Preference Reversals. *Journal of Marketing Research*, 34(2):205.

Ommeren, J. v. and Russo, G. (2013). Firm Recruitment Behaviour: Sequential or Non-sequential Search? *Oxford Bulletin of Economics and Statistics*.

Oyer, P. and Schaefer, S. (2011). Personnel economics: hiring and incentives. *Handbook of Labor Economics*.

Penn, Schoen Berland Associates, inc. (2012). T*he Capstone Project*.

Perie, M., Moran, R., and Lutkus, A. D. (2004). NAEP 2004 Trends in Academic Progress: Three Decades of Student Performance in Reading and Mathematics. Technical report, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC:.

Plante, I., Theoret, M., and Favreau, O. E. (2009). Student gender stereotypes: Contrasting the perceived maleness and femaleness of mathematics and language. *Educational Psychology*, 29(4):385–405.

Powell, M. and Ansic, D. (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology*, 18(6):605–628.

Pratt, J. and Zeckhauser, R. J. (1987). Proper Risk-Aversion. *Econometrica*, 55(1):143–154.

Price, C. R. (2011). Gender, Competition, and Managerial Decisions. *Management Science*, 58(1):114–122.

Quiggin, J. (2003). Background risk in generalized expected utility theory. *Economic Theory*, 22(3):607–611.

Rabin, M. and Thaler, R. H. (2001). Anomalies: risk aversion. *Journal of Economic perspectives*, 15(1):219–232.

Rabin, M. and Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *The American Economic Review*, 99(4):1508–1543.

Reskin, B. F. and Bielby, D. D. (2005). A Sociological Perspective on Gender and Career Outcomes. *Journal of Economic perspectives*, 19(1):71–86.

Reuben, E., Wiswall, M., and Zafar, B. (2013). Preferences and Biases in Educational Choices and Labor Market Expectations: Shrinking the Black Box of Gender. *IZA DP*, (7579).

Rich, J. and Riach, P. (2002). Field Experiments of Discrimination in the Market Place. *The economic journal*, 112(483):480–518.

Ritov, I., Kahneman, D., and Jacowitz, K. E. (1993). Stated willingness to pay for public goods: A psychological perspective.

Rouse, C. and Goldin, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, 90(4):715–741.

Samuelson, P. A. (1963). Risk and Uncertainty - a Fallacy of Large Numbers. *Scientia*, 98(612):108–113.

Sprenger, C. (2010). An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points. *Working Paper*, pages 1–46.

Stainback, K., Tomaskovic-Devey, D., and Skaggs, S. (2010). Organizational approaches to inequality: Inertia, relative power, and environments. *the annual review of sociology*, (33):473–501.

Stanovich, K. E. and West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5):645–665.

Stigler, G. J. (1961). The economics of information. *The journal of political economy.*

Sundali, J. A., Safford, A. H., and Croson, R. (2012). The impact of near-miss events on betting behavior: An examination of casino rapid roulette play. 7(6):768–778.

Thaler, R. H. and Johnson, E. J. (1990). Gambling with the House Money and Trying to Break Even - the Effects of Prior Outcomes on Risky Choice. *Management Science*, 36(6):643–660.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge.* Improving Decisions About Health, Wealth, and Happiness. Penguin.

van Geen, A. (2014). Risk in the Background: How Men and Women Respond. *Working Paper*.

Wang, M., Liu, Y.-J., Tsai, C.-L., and Zhu, N. (2010). Prior Consequences and Subsequent Risk Taking: New Field Evidence from the Taiwan Futures Exchange. *Management Science*, 56(4):606–620.

Wright, W. F. and Bower, G. H. (1992). Mood Effects on Subjective-Probability Assessment. *Organizational behavior and human decision processes*, 52(2):276–291.

Zahidi, S. and Ibarra, H. (2010). The corporate gender gap report 2010. *World Economic Forum Geneva*.