I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16138

The Gender Reference Point Gap

Nathan Kettlewell
Jonathan Levy
Agnieszka Tymula
Xueting Wang

MAY 2023

IZA – Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16138

The Gender Reference Point Gap

Nathan Kettlewell
University of Technology Sydney and IZA

Jonathan Levy
University of Sydney

Agnieszka Tymula
University of Sydney

Xueting Wang
Royal Melbourne Institute of Technology

MAY 2023

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Gender Reference Point Gap*

Studies have frequently found that women are more risk averse than men. In this paper, we depart from usual practice in economics that treats risk attitude as a primitive, and instead adopt a neuroeconomic approach where risk attitude is determined by the reference point which can be easily estimated using standard econometric methods. We then evaluate whether there is a gender difference in the reference point, explaining the gender difference in risk aversion observed using traditional approaches. In our study, women make riskier choices less frequently than men. Compared to men, we find that women on average have a significantly lower reference point. By acknowledging the reference point as a potential source of gender inequality, we can begin a new discussion on how to address this important issue.

JEL Classification: C90, D87, D91, J16

Keywords: reference point, risk attitude, neuroeconomics, gender, inequality, experiment

Corresponding author:
Nathan Kettlewell
University of Technology Sydney
Economics Discipline Group
15 Broadway
Ultimo NSW 2007
Australia

E-mail: nathan.kettlewell@uts.edu.au

* We are especially thankful to Australian twins who gave up their time to take part in our study. Participant recruitment was facilitated by Twins Research Australia, a national resource supported by the National Health & Medical Research Council (NHMRC) and administered by the University of Melbourne. This work was supported by the University of Technology Sydney Chancellor's Post-Doctoral Fellowship for Kettlewell, a UTS Business School small research grant, and the Australian Research Council's Centre of Excellence for Children and Families over the Life Course (Project ID CE200100025). We are grateful for valuable feedback from the participants at the 2022 Society for Neuroeconomics Annual Meeting, the 2022 Women in Experimental Finance seminar, the 2022 Gender, Norms, and Economics Workshop, the 2023 Western Economic Association International meeting, and the neuroeconomics seminar at the University of Zurich.

1. Introduction

Studies have frequently found that women are more risk averse than men (Agnew et al., 2008; Borghans et al., 2009; Charness & Gneezy, 2012; Eckel & Grossman, 2008; Shurchkov & Eckel, 2018). This gender gap may contribute to numerous gender-based inequities, such as the underrepresentation of females in entrepreneurship, managerial positions, investing, as well as the gender pay gap.

Although the gender difference in risk attitude has been replicated in many studies (this is not always the case, e.g., Filippin & Crosetto, 2016; Nelson, 2015), economists have struggled with whether and how to use this finding to restore gender balance in economic decision-making. The main reason for this is that traditional economic models take risk attitudes as static primitives which makes it impossible to establish the underlying mechanism that leads to gender differences in risk attitudes. Conceptually, in these models, people maximize their utility given their preferences and forcing them to change their decisions should result in lower utility. As such, we are no closer to devising a solution to address this risk attitude-based gender inequality.

In this paper, we take a different approach to risk attitude. Instead of treating it as a primitive, we employ a neuroeconomic model in which risk attitude is determined by the payoff expectation (reference point) (Glimcher & Tymula, 2023). We then evaluate whether there is a gender difference in the estimated dollar amount that serves as a reference point, explaining the gender difference in risk aversion observed using traditional approaches. The advantage of our approach is that it sheds light on the origins of the gender differences in risk attitudes which in turn enables us to suggest policies that would reduce or eliminate these differences. In the neuroeconomic model, a person's reference point is malleable and affected by their financial history (Glimcher & Tymula, 2023; Rangel & Clithero, 2012; Tymula & Plassmann, 2016). Different histories of financial outcomes across genders can lead to different reference points and different risk attitudes. This implies that the existing gender pay gap may be self-reinforcing—lowering women's reference points could decrease their tolerance to risk which in turn reduces their expected financial outcomes.

We use a dataset of 853 participants (18-67 years old; Mean = 41.71, Std. Dev. = 14.22) to provide the first evidence on gender differences in reference points estimated from behavior.

We use a canonical neuroeconomic model (divisive normalization) in which the reference point is derived from the neurobiological capacity constraints of the nervous system (Glimcher & Tymula, 2023; Louie et al., 2013; Webb et al., 2020), and estimate the reference point for men and women using choices across multiple binary lottery tasks. We find that the reference point for men is more than double that of women. We estimate models of Expected Utility and Prospect Theory and show that a model that accounts for a gender difference in the reference point fits the data better than a model that accounts for gender differences in utility curvature and probability weighting. Our results suggest that income mediates the gender difference in the reference point but does not eliminate it. We also replicate the gender gap in reference point using data from a very similar task (Baillon et al., 2020). As reference points determine risk tolerance, which arguably could in turn determine financial outcomes (Budria et al., 2013; Shaw, 1996), the gender difference in reference points can adversely affect women. This insight is important for guiding both whether and how policymakers should intervene to eliminate these differences.

We contribute to two strands of literature. First, we contribute to the emerging literature on gender differences in expectations. Gender differences in expectations have been identified in a variety of domains, including differences in expectations about inflation (D'Acunto et al., 2021), labour supply (Grewenig et al., 2020), and salary (Briel et al., 2022; Cortés et al., 2022; Fernandes et al., 2021; Reuben et al., 2017). In these studies expectations are stated (not incentivized) and may be influenced by unmeasured differences in the information received by participants. Our study differs from prior research in this space in two ways. First, the task used to elicit reference points, conceptually related to payoff expectations, is incentivized. Second, we elicit reference points in a controlled setting where objective payoffs and probabilities are identical between subjects and there is no information asymmetry. Heterogeneity in stated beliefs in areas like wage expectations can be driven by both different satisficing outcomes, as well as other factors, like private information about ability or beliefs about discrimination. By holding other factors constant, we provide evidence on whether the gender difference in expectations is due to a real gap in the reference point.

Second, we contribute to the empirical literature on the estimation of reference points. Although there has been research prior to ours which has examined reference-dependence theoretically (e.g., Bell, 1985; Koszegi & Rabin, 2006; Loomes & Sugden, 1986), and predictions of such models empirically (e.g., Abeler et al., 2011; Allen et al., 2017; Bartling et al., 2015; Baucells et al., 2011; Card & Dahl, 2011; Crawford & Meng, 2011; Gill & Prowse, 2012; Heffetz & List, 2014; Lien & Zheng, 2015; Rosato & Tymula, 2019; Wenner, 2015), very few studies estimate reference points. To our knowledge the only studies that do so are Baillon et al. (2020), Rees-Jones & Wang (2022), and Terzi et al. (2016). In each of these studies a set of reference point rules is proposed and the extent to which participants employ these reference point rules is estimated. These reference point rules are forward looking meaning that when a person makes a decision, they evaluate its potential outcomes by comparing them to some or all of the other possible outcomes that they could have received had they made a different decision, or their luck changed. The limitation of proposing that individuals employ such forward-looking reference point rules when making decisions is that we implicitly assume that reference points are not dependent on historical outcomes. However, there has been ample research in finance (Andrikogiannopoulou & Papakonstantinou, 2020; Barberis et al., 2001), behavioral economics (Imas, 2016; Malmendier & Nagel, 2011; Post et al., 2008), and neuroeconomics (Glimcher & Tymula, 2023; Guo & Tymula, 2021) which suggests that reference points are at least partially affected by historical outcomes. Unlike the previous research on reference point estimation (Baillon et al., 2020; Rees-Jones & Wang, 2022; Terzi et al., 2016), we estimate the reference point as a dollar amount rather than a simple rule. The major difference is that instead of evaluating how the reference point changes from trial to trial, we propose that the reference point has a more stable component and that risk attitudes emerge from the reference point rather than being assumed (Rayo & Becker, 2007; Robson et al., 2022; Woodford, 2012). Using this approach, we estimate that women have a lower reference point than men, a difference not documented in prior research. In our framework, the reference point is a continuous latent variable and there is no discontinuity in the utility function at the reference point. This makes it easy to estimate with the popular risk elicitation tasks and maximum likelihood estimation routines used widely in behavioral economics.

The rest of the paper is organized as follows. In section 2, we outline the methods used for data collection and the empirical approach. In section 3 we report results along with several checks for robustness. Section 4 concludes.

2. Methods

2.1. Ethics statement

Our protocols and procedures were approved by the University of Technology Sydney Human Research Ethics Committee (application number ETH21-6527) and by Twins Research Australia.

2.2. Participants and procedures

We use data from the second wave of the Australian Twins Economic Preferences Survey.[1] Our sample comprises 853 participants[2] (18-67 years old; Mean = 41.71, Std. Dev. = 14.22, 708 female) recruited from Twins Research Australia, the largest twin registry in Australia. Like most datasets that experimentally elicit preferences, our sample is not representative. However, compared to many related studies, which rely on student samples, our sample has two notable advantages. First, it includes participants with a range of ages, education levels, and other socioeconomic and demographic characteristics that are more representative of the general population. Second, our sample is large, which allows us to precisely estimate gender differences, even though we have a smaller number of males in our sample.[3] Table 1 presents the comparison of male and female participants on key demographic and socioeconomic characteristics. Compared to females, males were significantly less likely to have children, were significantly more likely to be educated at the university level, had significantly higher weekly incomes, judged themselves as significantly wealthier, and were significantly less likely to have a long-term health condition. We control for these differences in our analysis.

[1] Kettlewell & Tymula (2021) describe the recruitment and study design for the first wave of the survey in detail. The second wave included 657 participants from wave 1 as well as an additional 196 new participants.
[2] Our sample includes 9 participants who completed the survey but did not provide bank details, so could not be paid. We retain these participants since they are so few, and we have no reason to think they misrepresented their preferences (the fact they completed the whole survey indicates a strong intrinsic motivation). Our results do not change when we exclude these participants from the analysis.
[3] This gender difference is largely a feature of the composition of Twins Research Australia’s twin registry (see Kettlewell & Tymula, 2021).

Table 1: Descriptive statistics

Variable	Mean Female	Mean Male	Difference	P-value
Age	41.816	41.172	0.644	0.620
Lives in a city	0.651	0.724	-0.073	0.101
Married/defacto	0.641	0.641	-0.000	1.000
Household size	2.805	2.738	0.067	0.595
Has children	0.487	0.407	0.080	0.083
University degree	0.569	0.683	-0.114	0.012
Employed	0.802	0.828	-0.025	0.564
Retired	0.075	0.097	-0.022	0.396
Income (weekly)	1248.387	1702.427	-454.04	0.000
Wealth (self-reported)[4]	4.107	4.269	-0.162	0.019
Long-term health condition	0.220	0.131	0.089	0.017

Notes: Calculated from non-missing values from a full sample of 145 males and 708 females. See Table A.1 for variable definitions and observation counts.

The survey was programmed in Qualtrics and conducted online between November 2021 and April 2022. Participants completed four tasks and a demographic and socioeconomic questionnaire. One of the four tasks was randomly selected for payment and the payment was determined by the decision made by participants in this task and chance. To emphasize the importance of participants' decisions, they did not receive a fixed payment for participation.

2.3. Lottery choice task

We analyze data from the lottery choice task that included 46 decision scenarios designed to estimate participants' reference point (Baillon et al., 2020).[5] Each decision was between two options: lottery A and lottery B. Each lottery had between one and four possible payoffs at various probability levels. All payoffs were in Australian dollars and were weakly positive to make sure that participants would not lose money. Between participants, we randomized the order of the decision scenarios, and the side on which lotteries appeared. The selection of decision scenarios ensured the complete coverage of the outcome and probability space and a balanced pairing of prospects with different numbers of outcomes.

[4] Participants reported how prosperous they felt on a scale ranging from 'very poor' (=1) to 'prosperous' (=5).
[5] To reduce the length of the survey, we used 46 decision scenarios instead of the original 70. We also adjusted the payoffs to Australian dollars. All 46 decision scenarios are listed in Table A.2.

Figure 1: Presentation of the choices in the survey

Figure 1 shows how the lottery options were displayed. Prospects were presented as vertical bars with as many parts as there were different payoffs. The size of each part corresponded with the probability of the payoff. The intensity of the color of each part increased with the size of the payoff. The payoffs were presented in decreasing order (the lowest at the top and the highest at the bottom). Participants clicked on a bullet located next to a lottery to indicate their preference. Decision scenarios 45 and 46 (see Table A.2) had one lottery stochastically dominating the other.

If the lottery choice task was selected for payment, one of the 46 decisions within this task was randomly selected for payment. An expected value maximizer would in expectation earn $19.37 (Std. Dev. = 6.09) for this task.[6]

2.4. Empirical approach

We estimate the reference point using the recently proposed modeling approach in Expected Subjective Value Theory (ESVT) (Glimcher & Tymula, 2023). ESVT is based on the neuroscientific understanding about how value signals are efficiently encoded in the brain. The

[6] An expected value maximizer would earn $18.38, $19, and $10.50 for the other three tasks, respectively.

main intuition behind ESVT is that the utility function adapts to the payoff expectation to efficiently encode value (Bucher & Brandenburger, 2022; Steverson et al., 2019). Because the brain does not have unlimited resources (action potentials) to encode the utility of payoffs, it adjusts dynamically so that the subjective value function[7] is most sensitive to the payoff ranges that the brain is expecting to encounter. In this vein, the model is very similar to range normalization models (Kontek & Lewandowski, 2018; Padoa-Schioppa & Rustichini, 2014). However, it has a unique advantage over range normalization models, in that it allows for reference point estimation. ESVT implements behaviors captured by Prospect Theory (PT), offering new interpretations for risk taking, reflection in risk attitudes, probability weighting, the endowment effect, and the Allais paradox. For our purposes, the biggest benefit of ESVT is that it does not assume a discontinuity at the reference point (like PT or other loss-aversion based reference point models) which allows us to estimate it using the standard maximum likelihood procedure and is in line with neural evidence that biological systems rarely, if ever, employ discontinuous functions.

Under ESVT the utility of a payoff $x \in \mathbb{R}_+$ is given by:

$$u(x) = \frac{x^\alpha}{x^\alpha + M^\alpha} \tag{1}$$

where M is the payoff expectation (reference point) and $\alpha > 0$ is a free parameter called *predisposition*.

The utility function takes values between 0 and 1 ($u \in [0, 1]$) consistent with the idea that decision makers are bounded in the range of subjective values that they can biophysically assign to payoffs (Rayo & Becker, 2007; Robson et al., 2022; Woodford, 2012).

We assume that the expected utility of a lottery that pays x with probability p is calculated as:

$$v(p, x) = E[u(x), p] = w(p)u(x). \tag{2}$$

We therefore diverge from Glimcher & Tymula (2023) by allowing for a probability weighting function.

[7] Neuroeconomists use the term “subjective value” to distinguish it from utility to capture that the former is usually thought of as cardinal and the latter ordinal. Throughout the paper we use the term utility as is the norm in economics.

Both M and α determine the shape of the utility function and therefore the risk attitude. As the reference point (M) increases, the utility function shifts rightward along the horizontal axis. This feature prevails in many reference-dependent models. The shift in the utility function is such that at the reference point the function always takes the same value, $u(x = M) = 0.5$. This property is consistent with mounting evidence that people are indeed reference dependent when making decisions. Predisposition (α) affects the curvature of the utility function. When predisposition is low the utility function is concave for all x and thus the decision maker is always risk averse. As α increases, the utility function becomes increasingly S-shaped as seen in PT—for small values of x utility is convex (risk seeking) and then as x increases changes to concave (risk averse). Unlike in PT, in ESVT the inflection point does not have to occur at the reference point. Instead, the inflection point depends on both the reference point (M) and predisposition (α). If the utility function inflects, then as in PT, for individuals with a higher reference point (M), the utility inflects at a higher x meaning that they are risk seeking for a larger range of x values.

We fit decisions of our participants with a logistic choice function, where the probability of choosing lottery A is:

$$P(A) = \frac{1}{1 + e^{-Z}} \tag{3}$$

where $Z = \frac{v(A)-v(B)}{\mu}$, and μ captures noise. The log-likelihood function is then given by:

$$LL(\boldsymbol{\theta}) = \sum_{n=1}^{N} \sum_{i=1}^{I} y_{ni} \ln\big(P_{ni}(A_i)\big) + (1 - y_{ni}) \ln\big(1 - P_{ni}(A_i)\big) \tag{4}$$

where N is the number of participants, I is the number of trials, $y_{ni} = 1(0)$ is an indicator function denoting the choice of lottery $A(B)$ for participant n in trial i, and $\boldsymbol{\theta}$ is the vector of behavioral parameters to be estimated.

To capture gender effects, we introduce a dummy variable $Male$ which takes the value 1 if the participant is male, and 0 otherwise. In our analysis we control for age, and for all descriptive variables listed in Table 1 in which women and men differ (income, wealth, whether the participant obtained a university level education, whether the participant has children, and whether the participant has a long-term health condition).

For each parameter θ_n in our model, we specify:

$$\theta_n = \theta_0 + \theta_{Male} \times Male_n + X_n'\beta \quad (5)$$

where X is a vector of controls and β are the associated coefficients. For each parameter θ, we report the point estimate for the maximum likelihood estimation. The standard errors are clustered at the individual level.[8]

For comparison and robustness checks we use the one-parameter ($w(p) = e^{-(-\ln(p))^\gamma}$) and two-parameter ($w(p) = e^{-\delta(-\ln(p))^\gamma}$) probability weighting functions proposed by Prelec (1998). In these functions, $\gamma \in [0,1]$ governs the shape of the weighting function, with smaller γ corresponding to a more inverse S-shaped function (i.e., greater over (under)-weighting of low (high) probability outcomes), while δ determines the level of elevation (where the function crosses the 45-degree line). The two-parameter probability weighting function is more flexible and thus we present these results in the main paper and the results using one-parameter probability weighting in Appendix A.

To allow for easy comparisons with existing literature on risk attitudes, we also estimate a power utility function ($u(x) = x^r$) with and without Prelec probability weighting. We label models with probability weighting as PT models and the model without probability weighting as Expected Utility (EU). Further details on the models estimated are in Appendix C.

3. Results

3.1. Preliminary results

A simple way to determine whether there is a gender difference in risk attitude in our sample is to compare whether women choose riskier lotteries less or more frequently than men. We determine which lottery is riskier in each trial using the coefficient of variation (Weber, 2010). The coefficient of variation for a lottery is its standard deviation divided by its expected value. Lotteries with higher coefficients of variation are considered riskier. We find that women choose riskier lotteries less frequently than men (42.7% versus 46.5%, t-test: p-value = 0.0087)—see Figure 2A.[9] Their expected earnings, calculated based on their decisions in the

[8] The main results do not change when we cluster standard errors at the sibling level.
[9] All reported p-values are from two-sided tests.

lottery choice task, are 0.46 standard deviations lower than men's ($18.88 versus $18.96, t-test: p-value < 0.001)—see Figure 2B.[10]

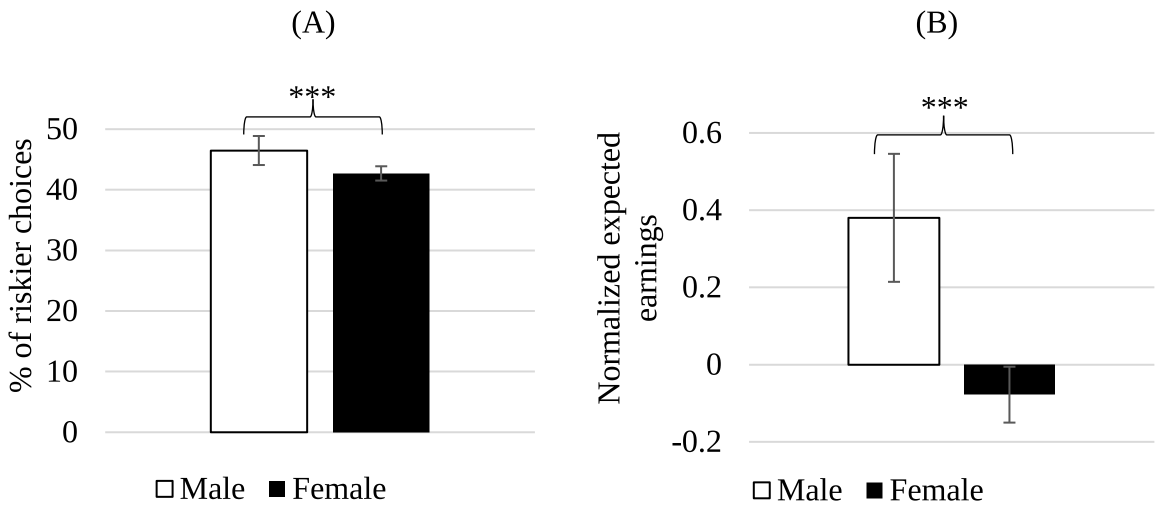

Figure 2: Percentage of riskier choices and normalized expected earnings by gender

We confirm that women have a more concave utility function (i.e., are more risk averse) than men when modeling decisions using EU (Table A.3) and PT with one-parameter (Table A.5) and two-parameter (Table A.7) Prelec probability weighting functions—see the significant and positive coefficient estimate on r_{Male} in all models. As the literature has suggested (Fehr-Duda et al., 2006; Filippin & Crosetto, 2016), we observe statistically significant differences across genders in probability weighting (Table A.7). We find that men have significantly lower γ (no significant gender difference in δ). The coefficient estimates for γ_{Male} are between -0.033 and -0.029. This gender difference seems small as evidenced by the similarity in the estimated probability weighting functions drawn separately for each gender based on Table A.7 estimates in Figure B.1. Such differences in probability weighting could nevertheless have an impact on earnings if they are paired with substantial payoffs. In our study though, the differences in probability weighting are not the key driver of differences across the genders. We conducted an in-sample prediction exercise and found that model predictions are more accurate when allowing for men and women to differ in utility curvature than when allowing for men and women to differ in probability weighting. Moreover, the gender difference in utility curvature persists even when probability weighting is incorporated. To summarize, these findings suggest

[10] In Figure 2B, the normalized expected earnings for participant i are equal to $(E_i - E_P)/\sigma_P$ where E_i are the expected earnings for participant i, E_P are the average expected earnings in the sample population, and σ_P is the standard deviation of the expected earnings in the sample population.

that the observed gender differences in risk attitudes are not primarily due to probability weighting. Our results do not change qualitatively when we employ different probability weighting functions[11] or exclude participants who violated first-order stochastic dominance (Table A.8).

3.2. Reference point heterogeneity

We now estimate the ESVT model (Glimcher & Tymula, 2023). This allows us to understand whether in the neuroeconomic framework the observed differences in willingness to take risk across genders should be attributed to differences in the reference point or predisposition as both, in principle, affect risk attitude. We begin by estimating the ESVT model without controls (first column in Table 2).

Table 2: ESVT model estimates

	(1)	(2)
M_{Male}	15.119*	34.999***
	(8.184)	(13.555)
M	13.002***	12.620***
	(1.524)	(4.743)
Controls	No	Yes
α_{Male}	-0.093	-0.246***
	(0.087)	(0.068)
α	1.065***	0.983***
	(0.054)	(0.151)
Controls	No	Yes
μ_{Male}	-0.026***	-0.033***
	(0.006)	(0.005)
μ	0.053***	0.052***
	(0.004)	(0.004)
Obs.	39238	37352
Clusters	853	812
AIC	50910.979	48337.850
BIC	50962.443	48491.357

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[11] In addition to estimating PT with $w(p)$ as in Prelec (1998) we also estimated prospect theory models with $w(p)$ as in Goldstein & Einhorn (1987) and Tversky & Kahneman (1992). The results from these alternative estimations were consistent and the model with two-parameter Prelec probability weighting had the lowest BIC score.

The coefficient estimate for M_{Male} is significant and positive, implying women have lower reference points than men. Using the estimates in the first column of Table 2 we plot the utility functions for males and females in Figure 3.

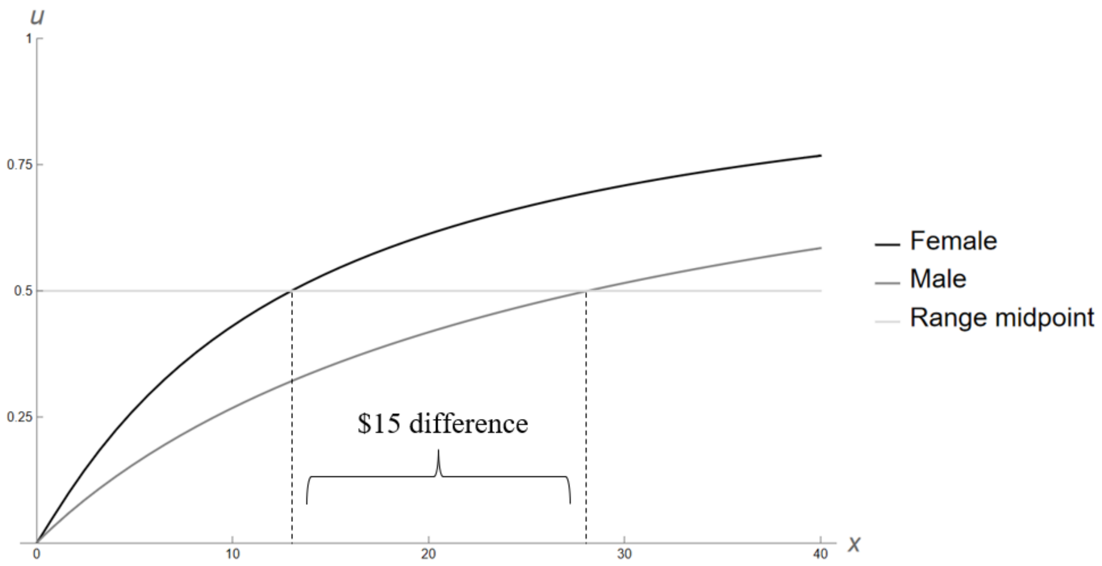

Figure 3: Utility functions for males and females plotted using parameter estimates from column (1) in Table 2

When estimated as a continuous variable, the reference point for women is on average $15 lower than the reference point for men. This is a substantial difference, considering that in our task, the maximum payoff was $43. This, of course, could happen simply because men and women in our sample differ on variables that determine the reference point but are not related to gender. Or it could be that we underestimate the difference in reference points by not controlling for education or income levels. To ensure that we capture the true gender effect and compare men and women who are similar in demographic and socioeconomic characteristics, we introduce controls to the model—age, income, wealth, whether the participant obtained a university level education, whether the participant has children, and whether the participant has a long-term health condition (all variables in which we observe gender differences, see Table 1). Not only does the gender difference prevail after adding controls, but it becomes larger (Table 2 (2)).

In the cross-sectional analysis so far, we assumed that the reference point is fixed within individuals and estimated it on the aggregate level. Research in neuroscience and neuroeconomics (Frydman & Jin, 2021; Glimcher & Tymula, 2023; Padoa-Schioppa & Rustichini, 2014; Rangel & Clithero, 2012; Tymula & Plassmann, 2016) tells us that the reference point is malleable and adaptable and that participants generally adjust their utility functions to either minimize the number of mistakes or maximize payoffs. In particular, based on Glimcher & Tymula (2023), we would hypothesize that lower-income individuals adjust to have lower reference points. While we cannot establish causality, we can check whether such correlations exist in our data. Furthermore, we can check whether the gender difference in the reference point exists above and beyond the potential impact of income and wealth on the reference point. This is particularly important because in our sample women have lower incomes and are less wealthy than men.

Recall, in the second column of Table 2 the gender difference in the reference point is not only robust but increases significantly when we include controls. In Table 3 we present the values of all coefficients of the model presented in the second column of Table 2.

Table 3: Determinants of reference points and predisposition

	Reference point (M)	Predisposition (α)
Male	34.999***	-0.246***
	(13.555)	(0.068)
Age	0.057	-0.005***
	(0.071)	(0.002)
University degree	-1.246	0.201***
	(1.490)	(0.048)
Wealth	-1.074	0.045
	(0.965)	(0.028)
Income (weekly)	0.002**	-0.000
	(0.001)	(0.000)
Has children	-0.186	0.004
	(2.199)	(0.053)
Long-term health condition	1.077	0.020
	(1.551)	(0.068)
Constant	12.620***	0.983***
	(4.743)	(0.151)
Obs.	37352	
Clusters	812	
AIC	48337.850	
BIC	48491.357	

Notes: These results are from one estimation. The estimates were split across two columns for display purposes. The first column provides parameter estimates with respect to the reference point (M) and the second column provides parameter estimates with respect to predisposition (α). Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In the first column of Table 3 we can see that in line with our hypothesis the coefficient of

Income is positive and statistically significant. This indicates that a \$100 decrease in weekly income leads to a \$0.20 decrease in the reference point. This supports the narrative that the existing gender pay gap may be self-reinforcing—lowering women's reference points could decrease their tolerance to risk which in turn reduces their expected financial outcomes. Income does seem to be a mitigating factor, however, the gender difference in reference points persists as indicated by the significantly positive coefficient estimate of *Male* in the first column of Table 3.[12]

To summarize, our findings suggest that the observed gender differences in risk attitudes could be due to differences in reference points across genders.

3.3. Robustness checks

In this subsection, we test the robustness of our results. We explore whether our main results change when we exclude stochastic dominance violators from our sample and incorporate different probability weighting specifications. We also check whether ESVT better describes chooser's behavior than Prospect Theory or Expected Utility Theory. Finally, we try to replicate our findings using publicly available data from Baillon et al. (2020) whose task we adopt.

3.3.1 First-order stochastic dominance check

Two trials in our task had one lottery stochastically dominating the other. Out of the 853 participants we recruited, 646 participants (118 males and 528 females) never chose the dominated lottery, 188 participants (25 males and 163 females) chose the dominated lottery once, and 19 participants (2 males and 17 females) chose the dominated lottery twice. We do observe a gender difference in decisions in the two trials where one lottery stochastically dominates the other. Women violate first-order stochastic dominance more frequently than men (25.4% versus 18.6%, Fisher's exact test: p-value = 0.089). However, the results pertaining to

[12] In column (1) of Table 2, the coefficient of α_{Male} is not statistically significant, indicating no gender difference in predisposition. As we include controls the coefficient becomes negative and statistically significant. In the second column of Table 3 we can see that the coefficients of *Age* and *University* degree are significantly negative and significantly positive, respectively. The noise parameter μ was significantly different across genders. Indicating that decisions were noisier for women than men.

the structural estimations in subsection 3.2 do not change when we exclude participants that violated stochastic dominance—see Table A.9.

3.3.2 Reference point or probability weighting?

The overweighting of low probabilities and the underweighting of high probabilities is a feature of human behavior which has been frequently discussed. ESVT (the model we use to estimate the reference point) does not explicitly employ probability weighting. However, without the inclusion of probability weighting it can achieve effects captured by probability weighting much like some other models of behavior which do not include probability weighting (e.g., Kontek & Lewandowski, 2018; Schneider & Day, 2018). Yet one might argue that incorporating probability weighting into ESVT may impact our findings. We find that the results in subsection 3.2 do not change qualitatively when we incorporate probability weighting. This is true both when we use a one-parameter probability weighting function (Table A.10) and a two-parameter probability weighting function (Table A.11).

3.3.3 Predictive power of ESVT versus PT and EU

We compare the BIC scores across the six models listed in Appendix C. ESVT generally outperforms (has lower BIC scores than) the EU and PT models. This is illustrated in Figure B.2 where we plot the BIC scores across all models based on the estimates in tables 2, A.3, A.5, A.7, A.10, and A.11. This implies that ESVT provides a better fit of the data than EU and PT. ESVT with a two-parameter probability weighting function provides the best fit of the data across the models.

Next, we determine whether allowing for reference point differences across genders enhances model fit in a standard Prospect Theory framework. We do so by using the male and female reference point estimates presented in the first column of Table 2[13] and estimating a standard behavioral economics model with a reference point and comparing it to EU. To be precise, we estimate the following PT model where $v(x) = pu(x)$ and

$$u(x) = \begin{cases} (x - M)^{r^g} & \text{if } \; x \geq M \\ -(M - x)^{r^l} & \text{otherwise,} \end{cases} \tag{6}$$

[13] The reference point used for males and females was 28.121 and 13.002, respectively.

and r^g and r^l represent utility curvature in the gain and loss domains, respectively. To make the comparison fair, this model is without loss aversion and probability weighting but allows for different utility curvature in the gain and loss domains (relative to the reference point). The BIC scores for this model are lower than the corresponding BIC scores for EU—see Table A.15 and Figure B.4. Therefore, we conclude that allowing for reference point differences across genders improves the predictive power of the model. Furthermore, the parameter estimates from the prospect theory model indicate that both men and women were more risk averse for payoffs below the reference point than above it (Wald test: p-values < 0.01).

3.3.4 Replication study

To check whether the results in this paper extend beyond our dataset, we repeated the structural estimation process for the models outlined in Appendix C using data from Baillon et al. (2020), selected due to task similarity.[14] Their sample contains 139 students and employees from the Technical University of Moldova (17-47 years old; Mean = 22.57, Std. Dev. = 4.66). Unlike in our sample, the majority (66%) of participants are male. Consistent with the results from our dataset, in their dataset the reference point for women is on average 38 Lei (59.3%) lower than the reference point for men (an expected value maximizer earned approximately 260 Lei in expectation)—see tables A.12, A.13, and A.14. Furthermore, the BIC scores indicate that ESVT with a one-parameter Prelec probability weighting function fits the data best. Figure B.3 displays the BIC scores across the six models.

4. Conclusion

In this paper we capture risk attitudes differently. Instead of treating them as primitives, we use a neuroeconomic approach where risk attitudes are determined by the reference point. We then evaluate whether there is a gender difference in the reference point, explaining the gender difference in risk aversion observed using traditional approaches. We find that women make riskier choices less frequently than men. We also find that women on average have a significantly lower reference point. We have shown that this result is not only robust to

[14] Note, the only demographics included in the Baillon et al. (2020) dataset were age and gender. Hence, we only controlled for these variables in our analysis.

controlling for socioeconomic and demographic variables but that the gender gap in reference point gets bigger when we control for gender differences in these variables. We have also replicated our result in an independent sample. Our results suggest that gender differences in reference points may be the reason why we observe gender differences in risk attitudes.

There have been many studies in behavioral economics that identified the gender gap in risk attitudes (Borghans et al., 2009; Charness & Gneezy, 2012; Eckel & Grossman, 2008), tested its limits (Filippin & Crosetto, 2016)[15], and stressed the importance of its implications (Agnew et al., 2008; Shurchkov & Eckel, 2018). However, little progress has been made in applying this finding because of the specific meaning of risk preferences in economics. As they are treated as the primitives in the economic models of choice, economists would argue that we should not enforce more or less risk upon expected utility maximizers because this will decrease their utility. Evolving literature in neuroeconomics (Glimcher & Tymula, 2023; Rangel & Clithero, 2012; Tymula & Plassmann, 2016) has begun to point out that risk preferences may not be fixed features of the choosers. Instead, this literature argues that they emerge from historical payoffs. This happens because the brain efficiently allocates neural resources to encode payoff values that it expects we are most likely to encounter. In such a setting, the interpretation of risk attitudes is different—they are malleable and determined by our payoff history. Under this setting, any economic inequalities that lead to lower payoffs would decrease the reference point. With a lower reference point, people exhibit greater risk aversion and thus expect lower payoffs, which in turn begets economic inequality further. To us, this suggests that as long as there is economic discrimination, the disadvantaged groups will take less risk and end up making decisions that reinforce the cycle of economic disadvantage. Policies that address economic inequality by equalizing payoff expectations should be particularly effective at breaking this self-reinforcing cycle. We find support for this idea in our paper as we estimate that the reference point increases in income. Policy changes that equalize pay and improve transparency about salaries could be an effective way to level the playing field. Indeed, Recalde & Vesterlund (2022) in their literature review conclude that transparency of pay reduces the gender differences in negotiations.

[15] Filippin & Crosetto (2016) argue that women prefer safe options and the gender gap in risk attitudes disappears if you remove the safe option. Our results provide further evidence of the gender gap in risk attitudes in a lottery choice task without safe options.

In this paper, we focused on gender differences in risk attitudes and reference points. Nevertheless, our conclusions can be extended to any group that experiences economic inequality. The insight that risk preferences are not fixed but are shaped by historical outcomes is perhaps not an entirely new concept (Imas, 2016; Malmendier & Nagel, 2011; Post et al., 2008) but it has not yet been applied to improve our understanding of economic disadvantage. Haushofer & Fehr (2014) argued that economic inequality creates a self-perpetuating loop—poverty increases levels of stress which in turn increases impatience and risk aversion leading to decisions that result in lower payoffs in expectation. Here, using the example of gender we provided a new suggestion on how poverty could reinforce itself—through a lower reference point. We have also provided the first empirical example of how a recent model from neuroeconomics can be applied to easily estimate reference points from choices. This opens the door to more research on reference points that relates to economic inequality and other topics where reference-dependence plays a role.

References

Abeler, J., Falk, A., Goette, L., & Huffman, D. (2011). Reference Points and Effort Provision. *American Economic Review*, *101*(2), 470–492. https://doi.org/10.1257/aer.101.2.470

Agnew, J. R., Anderson, L. R., Gerlach, J. R., & Szykman, L. R. (2008). Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults. *American Economic Review*, *98*(2), 418–422. https://doi.org/10.1257/aer.98.2.418

Allen, E. J., Dechow, P. M., Pope, D. G., & Wu, G. (2017). Reference-Dependent Preferences: Evidence from Marathon Runners. *Management Science*, *63*(6), 1657–1672. https://doi.org/10.1287/mnsc.2015.2417

Andrikogiannopoulou, A., & Papakonstantinou, F. (2020). History-Dependent Risk Preferences: Evidence from Individual Choices and Implications for the Disposition Effect. *The Review of Financial Studies*, *33*(8), 3674–3718. https://doi.org/10.1093/rfs/hhz127

Baillon, A., Bleichrodt, H., & Spinu, V. (2020). Searching for the Reference Point. *Management Science, 66*(1), 93–112. https://doi.org/10.1287/mnsc.2018.3224

Barberis, N., Huang, M., & Santos, T. (2001). Prospect Theory and Asset Prices. *The Quarterly Journal of Economics*, *116*(1), 1–53. https://doi.org/10.1162/003355301556310

Bartling, B., Brandes, L., & Schunk, D. (2015). Expectations as Reference Points: Field Evidence from Professional Soccer. *Management Science, 61*(11), 2646–2661. https://doi.org/10.1287/mnsc.2014.2048

Baucells, M., Weber, M., & Welfens, F. (2011). Reference-Point Formation and Updating. *Management Science, 57*(3), 506–519. https://doi.org/10.1287/mnsc.1100.1286

Bell, D. E. (1985). Disappointment in Decision Making Under Uncertainty. *Operations Research*, *33*(1), 1–27. https://doi.org/10.1287/opre.33.1.1

Borghans, L., Golsteyn, B. H. H., Heckman, J. J., & Meijers, H. (2009). Gender Differences in Risk Aversion and Ambiguity Aversion. *Journal of the European Economic Association*, *7*(2–3), 649–658. https://doi.org/10.1162/JEEA.2009.7.2-3.649

Briel, S., Osikominu, A., Pfeifer, G., Reutter, M., & Satlukal, S. (2022). Gender differences in wage expectations: the role of biased beliefs. *Empirical Economics*, *62*(1), 187–212. https://doi.org/10.1007/s00181-021-02044-0

Bucher, S. F., & Brandenburger, A. M. (2022). Divisive normalization is an efficient code for multivariate Pareto-distributed environments. *Proceedings of the National Academy of Sciences*, *119*(40). https://doi.org/10.1073/pnas.2120581119

Budria, S., Diaz-Serrano, L., Ferrer-i-Carbonell, A., & Hartog, J. (2013). Risk attitude and wage growth: replicating Shaw (1996). *Empirical Economics*, *44*(2), 981–1004. https://doi.org/10.1007/s00181-012-0549-5

Card, D., & Dahl, G. B. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior*. *The Quarterly Journal of Economics*, *126*(1), 103–143. https://doi.org/10.1093/qje/qjr001

Charness, G., & Gneezy, U. (2012). Strong Evidence for Gender Differences in Risk Taking. *Journal of Economic Behavior & Organization*, *83*(1), 50–58. https://doi.org/10.1016/j.jebo.2011.06.007

Cortés, P., Pan, J., Reuben, E., Pilossoph, L., & Zafar, B. (2022). Gender Differences in Job Search and the Earnings Gap: Evidence from the Field and Lab. *The Quarterly Journal of Economics*.

Crawford, V. P., & Meng, J. (2011). New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and

Income. *American Economic Review*, *101*(5), 1912–1932. https://doi.org/10.1257/aer.101.5.1912
D'Acunto, F., Malmendier, U., Ospina, J., & Weber, M. (2021). Exposure to Grocery Prices and Inflation Expectations. *Journal of Political Economy*, *129*(5), 1615–1639. https://doi.org/10.1086/713192
Eckel, C. C., & Grossman, P. J. (2008). *Chapter 113 Men, Women and Risk Aversion: Experimental Evidence* (pp. 1061–1073). https://doi.org/10.1016/S1574-0722(07)00113-8
Fehr-Duda, H., de Gennaro, M., & Schubert, R. (2006). Gender, Financial Risk, and Probability Weights. *Theory and Decision*, *60*(2–3), 283–313. https://doi.org/10.1007/s11238-005-4590-0
Fernandes, A., Huber, M., & Vaccaro, G. (2021). Gender differences in wage expectations. *PLOS ONE*, *16*(6), e0250892. https://doi.org/10.1371/journal.pone.0250892
Filippin, A., & Crosetto, P. (2016). A Reconsideration of Gender Differences in Risk Attitudes. *Management Science*, *62*(11), 3138–3160. https://doi.org/10.1287/mnsc.2015.2294
Frydman, C., & Jin, L. J. (2021). Efficient Coding and Risky Choice. *The Quarterly Journal of Economics*, *137*(1), 161–213. https://doi.org/10.1093/qje/qjab031
Gill, D., & Prowse, V. (2012). A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review*, *102*(1), 469–503. https://doi.org/10.1257/aer.102.1.469
Glimcher, P. W., & Tymula, A. A. (2023). Expected subjective value theory (ESVT): A representation of decision under risk and certainty. *Journal of Economic Behavior & Organization*, *207*, 110–128. https://doi.org/10.1016/j.jebo.2022.12.013
Goldstein, W. M., & Einhorn, H. J. (1987). Expression theory and the preference reversal phenomena. *Psychological Review*, *94*(2), 236–254. https://doi.org/10.1037/0033-295X.94.2.236
Grewenig, E., Lergetporer, P., & Werner, K. (2020). Gender Norms and Labor-Supply Expectations: Experimental Evidence from Adolescents. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3711887
Guo, J., & Tymula, A. (2021). Waterfall illusion in risky choice – exposure to outcome-irrelevant gambles affects subsequent valuation of risky gambles. *European Economic Review*, *139*, 103889. https://doi.org/10.1016/j.euroecorev.2021.103889
Haushofer, J., & Fehr, E. (2014). On the psychology of poverty. *Science*, *344*(6186), 862–867. https://doi.org/10.1126/science.1232491
Heffetz, O., & List, J. A. (2014). IS THE ENDOWMENT EFFECT AN EXPECTATIONS EFFECT? *Journal of the European Economic Association*, *12*(5), 1396–1422. https://doi.org/10.1111/jeea.12084
Imas, A. (2016). The Realization Effect: Risk-Taking After Realized Versus Paper Losses. *American Economic Review*, *106*(8), 2086–2109. https://doi.org/10.1257/aer.20140386
Kettlewell, N., & Tymula, A. (2021). The Australian Twins Economic Preferences Survey. *Twin Research and Human Genetics*, *24*(6), 365–370. https://doi.org/10.1017/thg.2021.49
Kontek, K., & Lewandowski, M. (2018). Range-Dependent Utility. *Management Science*, *64*(6), 2812–2832. https://doi.org/10.1287/mnsc.2017.2744
Koszegi, B., & Rabin, M. (2006). A Model of Reference-Dependent Preferences. *The Quarterly Journal of Economics*, *121*(4), 1133–1165. https://doi.org/10.1093/qje/121.4.1133

Lien, J. W., & Zheng, J. (2015). Deciding When to Quit: Reference-Dependence over Slot Machine Outcomes. *American Economic Review*, *105*(5), 366–370. https://doi.org/10.1257/aer.p20151036

Loomes, G., & Sugden, R. (1986). Disappointment and Dynamic Consistency in Choice under Uncertainty. *The Review of Economic Studies*, *53*(2), 271. https://doi.org/10.2307/2297651

Louie, K., Khaw, M. W., & Glimcher, P. W. (2013). Normalization is a general neural mechanism for context-dependent decision making. *Proceedings of the National Academy of Sciences*, *110*(15), 6139–6144. https://doi.org/10.1073/pnas.1217854110

Malmendier, U., & Nagel, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking?*. *The Quarterly Journal of Economics*, *126*(1), 373–416. https://doi.org/10.1093/qje/qjq004

Nelson, J. A. (2015). ARE WOMEN REALLY MORE RISK-AVERSE THAN MEN? A RE-ANALYSIS OF THE LITERATURE USING EXPANDED METHODS. *Journal of Economic Surveys*, *29*(3), 566–585. https://doi.org/10.1111/joes.12069

Padoa-Schioppa, C., & Rustichini, A. (2014). Rational Attention and Adaptive Coding: A Puzzle and a Solution. *American Economic Review*, *104*(5), 507–513. https://doi.org/10.1257/aer.104.5.507

Post, T., van den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show. *American Economic Review*, *98*(1), 38–71. https://doi.org/10.1257/aer.98.1.38

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, *66*(3), 497. https://doi.org/10.2307/2998573

Rangel, A., & Clithero, J. A. (2012). Value normalization in decision making: theory and evidence. *Current Opinion in Neurobiology*, *22*(6), 970–981. https://doi.org/10.1016/j.conb.2012.07.011

Rayo, L., & Becker, G. S. (2007). Evolutionary Efficiency and Happiness. *Journal of Political Economy*, *115*(2), 302–337. https://doi.org/10.1086/516737

Recalde, M. P., & Vesterlund, L. (2022). Gender Differences in Negotiation and Policy for Equalizing Outcomes. In *Bargaining* (pp. 455–475). Springer International Publishing. https://doi.org/10.1007/978-3-030-76666-5_21

Rees-Jones, A., & Wang, A. (2022). *An Approach to Testing Reference Points*. https://doi.org/10.3386/w30773

Reuben, E., Wiswall, M., & Zafar, B. (2017). Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. *The Economic Journal*, *127*(604), 2153–2186. https://doi.org/10.1111/ecoj.12350

Robson, A. J., Whitehead, L. A., & Robalino, N. (2022). Adaptive Cardinal Utility. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4036252

Rosato, A., & Tymula, A. A. (2019). Loss aversion and competition in Vickrey auctions: Money ain't no good. *Games and Economic Behavior*, *115*, 188–208. https://doi.org/10.1016/j.geb.2019.02.014

Schneider, M., & Day, R. (2018). Target-Adjusted Utility Functions and Expected-Utility Paradoxes. *Management Science*, *64*(1), 271–287. https://doi.org/10.1287/mnsc.2016.2588

Shaw, K. L. (1996). An Empirical Analysis of Risk Aversion and Income Growth. *Journal of Labor Economics*, *14*(4), 626–653. https://doi.org/10.1086/209825

Shurchkov, O., & Eckel, C. C. (2018). Gender differences in behavioral traits and labor market outcomes. In *The Oxford Handbook of Women and the Economy* (pp. 480–512). Oxford University Press.

Steverson, K., Brandenburger, A., & Glimcher, P. (2019). Choice-theoretic foundations of the divisive normalization model. *Journal of Economic Behavior & Organization*, *164*, 148–165. https://doi.org/10.1016/j.jebo.2019.05.026

Terzi, A., Koedijk, K., Noussair, C. N., & Pownall, R. (2016). Reference Point Heterogeneity. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01347

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. https://doi.org/10.1007/BF00122574

Tymula, A., & Plassmann, H. (2016). Context-dependency in valuation. *Current Opinion in Neurobiology*, *40*, 59–65. https://doi.org/10.1016/j.conb.2016.06.015

Webb, R., Glimcher, P. W., & Louie, K. (2020). Divisive normalization does influence decisions with multiple alternatives. *Nature Human Behaviour*, *4*(11), 1118–1120. https://doi.org/10.1038/s41562-020-00941-5

Weber, E. U. (2010). On the coefficient of variation as a predictor of risk sensitivity: Behavioral and neural evidence for the relative encoding of outcome variability. *Journal of Mathematical Psychology*, *54*(4), 395–399. https://doi.org/10.1016/j.jmp.2010.03.003

Wenner, L. M. (2015). Expected prices as reference points—Theory and experiments. *European Economic Review*, *75*, 60–79. https://doi.org/10.1016/j.euroecorev.2015.01.001

Woodford, M. (2012). Prospect Theory as Efficient Perceptual Distortion. *American Economic Review*, *102*(3), 41–46. https://doi.org/10.1257/aer.102.3.41

Appendix

A. Tables

Table A.1: Variable definitions

Variable	Definition	Female obs.	Male obs.
Age	Age at last birthday	707	145
Lives in a city	= 1 if currently live in a major city (Sydney, Melbourne, Brisbane, Adelaide, Perth, Canberra)	708	145
Married/defacto	= 1 if married or in a defacto relationship	708	145
Household size	How many people live in your household	707	145
Has children	= 1 if they have any children	708	145
University degree	= 1 if highest level of education obtained is a university degree	708	145
Employed	= 1 if worked any time in the last 7 days or if had a job but did not work in the last 7 days due to holidays, sickness or any other reason	708	145
Retired	= 1 if currently retired from the workforce	708	145
Income (weekly)	Average usual weekly own income in the last month using midpoint value for the following categories: \$1-\$149, \$150-\$299, \$300-\$399, \$400-\$499, \$500-\$649, \$650 \$799, \$800-\$999, \$1,000 \$1,249, \$1,250-\$1,499, \$1,500-\$1,749, \$1,750-\$1,999, \$2,000-\$2,999, \$3,000 or more (coded as \$3000). Negative or nil coded as missing.	670	143
Wealth	Given your current needs and financial responsibility, would you say that you and your family are: = 1 if Poor, = 2 if Just getting along, = 3 if Comfortable, = 4 if Very comfortable, = 5 if Prosperous.	708	145

Long-term health condition	= 1 if has a long-term health condition, impairment or disability that has lasted more than 6 months	708	145

Table A.2: Lottery decision scenarios

Scenario number	x_1^a	x_2^a	x_3^a	x_4^a	p_1^a	p_2^a	p_3^a	x_1^b	x_2^b	x_3^b	x_4^b	p_1^b	p_2^b	p_3^b
1	17	20	29	0	0.4	0.5	0.1	11	14	26	0	0.1	0.4	0.5
2	7	16	29	0	0.6	0.15	0.25	3	12	20	24	0.3	0.1	0.05
3	7	18	0	0	0.85	0.15	0	1	10	15	0	0.1	0.7	0.2
4	16	23	35	0	0.35	0.55	0.1	12	19	27	31	0.05	0.55	0.1
5	17	38	0	0	0.85	0.15	0	6	22	33	0	0.05	0.7	0.25
6	9	20	32	0	0.15	0.8	0.05	3	15	26	37	0.1	0.35	0.45
7	17	32	0	0	0.55	0.45	0	7	27	36	0	0.25	0.7	0.05
8	7	16	0	0	0.2	0.8	0	4	10	21	0	0.2	0.5	0.3
9	19	32	0	0	0.8	0.2	0	11	28	37	0	0.35	0.45	0.2
10	17	32	0	0	0.45	0.55	0	12	22	41	0	0.05	0.7	0.25
11	7	17	0	0	0.8	0.2	0	1	10	14	20	0.1	0.4	0.45
12	7	14	0	0	0.3	0.7	0	4	9	12	19	0.25	0.3	0.05
13	15	24	0	0	0.75	0.25	0	9	18	21	27	0.35	0.05	0.45
14	10	18	0	0	0.55	0.45	0	2	6	14	27	0.05	0.05	0.85
15	9	15	0	0	0.8	0.2	0	4	7	13	17	0.2	0.1	0.6
16	17	28	0	0	0.4	0.6	0	7	12	23	39	0.05	0.1	0.6
17	17	26	0	0	0.7	0.3	0	12	15	23	29	0.3	0.2	0.1
18	19	27	0	0	0.7	0.3	0	11	23	31	35	0.05	0.85	0.05
19	14	23	0	0	0.6	0.4	0	4	18	27	32	0.15	0.7	0.1
20	4	13	0	0	0.45	0.55	0	1	7	16	19	0.4	0.05	0.5
21	11	28	0	0	0.6	0.4	0	0	5	28	0	0.25	0.05	0.7
22	19	42	0	0	0.75	0.25	0	12	27	42	0	0.15	0.7	0.15
23	9	18	0	0	0.75	0.25	0	6	12	18	0	0.15	0.7	0.15
24	10	16	21	0	0.7	0.05	0.25	5	8	21	0	0.4	0.1	0.5
25	7	34	0	0	0.3	0.7	0	7	25	43	0	0.1	0.8	0.1
26	9	21	0	0	0.2	0.8	0	9	25	0	0	0.45	0.55	0
27	4	19	0	0	0.3	0.7	0	4	14	25	0	0.05	0.85	0.1
28	8	14	21	0	0.1	0.05	0.85	8	24	27	0	0.5	0.3	0.2
29	7	0	0	0	1	0	0	1	14	0	0	0.45	0.55	0
30	17	0	0	0	1	0	0	0	22	0	0	0.25	0.75	0
31	21	0	0	0	1	0	0	6	28	36	0	0.35	0.45	0.2
32	15	0	0	0	1	0	0	8	11	23	0	0.45	0.05	0.5
33	15	0	0	0	1	0	0	7	11	18	22	0.3	0.05	0.5
34	29	0	0	0	1	0	0	13	34	0	0	0.25	0.75	0

35	22	31	0	0	0.55	0.45	0	14	18	26	39	0.05	0.05	0.85
36	16	26	0	0	0.45	0.55	0	13	20	29	32	0.4	0.05	0.5
37	19	0	0	0	1	0	0	13	26	0	0	0.45	0.55	0
38	27	0	0	0	1	0	0	20	24	35	0	0.45	0.05	0.5
39	16	32	0	0	0.3	0.7	0	16	26	37	0	0.05	0.85	0.1
40	23	40	0	0	0.6	0.4	0	12	17	40	0	0.25	0.05	0.7
41	19	27	0	0	0.7	0.3	0	11	23	31	35	0.05	0.85	0.05
42	7	34	0	0	0.3	0.7	0	7	25	43	0	0.1	0.8	0.1
43	17	32	0	0	0.45	0.55	0	12	22	41	0	0.05	0.7	0.25
44	10	18	0	0	0.55	0.45	0	2	6	14	27	0.05	0.05	0.85
45	8	15	17	0	0.5	0.4	0.1	6	7	15	0	0.1	0.4	0.5
46	8	13	15	0	0.1	0.4	0.5	6	12	14	0	0.25	0.25	0.5

Notes: Table A.2 describes the 46 choices between lotteries $A = (p_1^a, x_1^a; p_2^a, x_2^a; p_3^a, x_3^a; 1 - p_1^a - p_2^a - p_3^a, x_4^a)$ and $B = (p_1^b, x_1^b; p_2^b, x_2^b; p_3^b, x_3^b; 1 - p_1^b - p_2^b - p_3^b, x_4^b)$ used in the survey.

Table A.3: EU model estimates

	(1)	(2)
r_{Male}	0.124***	0.128***
	(0.031)	(0.032)
r	0.414***	0.359***
	(0.014)	(0.044)
Controls	No	Yes
μ_{Male}	0.040	0.056
	(0.036)	(0.037)
μ	0.250***	0.243***
	(0.015)	(0.015)
Obs.	39238	37352
Clusters	853	812
AIC	51310.934	48757.487
BIC	51345.244	48842.769

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: EU model estimates for those who did not violate stochastic dominance

	(1)	(2)
r_{Male}	0.116***	0.119***
	(0.033)	(0.034)
r	0.409***	0.338***
	(0.015)	(0.048)
Controls	No	Yes
μ_{Male}	0.032	0.040
	(0.032)	(0.032)
μ	0.213***	0.210***
	(0.014)	(0.015)
Obs.	29716	28474
Clusters	646	619
AIC	38255.810	36605.728
BIC	38289.008	36688.296

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: PT1 model estimates

	(1)	(2)
r_{Male}	0.117***	0.121***
	(0.031)	(0.032)
r	0.413***	0.368***
	(0.013)	(0.042)
Controls	No	Yes
γ_{Male}	-0.032***	-0.029***
	(0.009)	(0.010)
γ	0.990***	1.022***
	(0.005)	(0.024)
Controls	No	Yes
μ_{Male}	0.044	0.056
	(0.039)	(0.040)
μ	0.251***	0.244***
	(0.015)	(0.015)
Obs.	39238	37352
Clusters	853	812
AIC	51243.926	48684.741
BIC	51295.391	48838.247

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: PT1 model estimates for those who did not violate stochastic dominance

	(1)	(2)
r_{Male}	0.109***	0.112***
	(0.033)	(0.034)
r	0.407***	0.346***
	(0.014)	(0.045)
Controls	No	Yes
γ_{Male}	-0.033***	-0.029***
	(0.010)	(0.010)
γ	0.989***	1.016***
	(0.005)	(0.025)
Controls	No	Yes
μ_{Male}	0.035	0.041
	(0.034)	(0.035)
μ	0.214***	0.210***
	(0.015)	(0.015)
Obs.	29716	28474
Clusters	646	619
AIC	38173.599	36507.379
BIC	38223.395	36656.000

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: PT2 model estimates

	(1)	(2)
r_{Male}	0.137***	0.138***
	(0.033)	(0.034)
r	0.441***	0.385***
	(0.015)	(0.047)
Controls	No	Yes
δ_{Male}	0.000	-0.001
	(0.008)	(0.009)
δ	1.043***	1.028***
	(0.004)	(0.026)
Controls	No	Yes
γ_{Male}	-0.033***	-0.029***
	(0.010)	(0.010)
γ	0.971***	1.010***
	(0.005)	(0.025)
Controls	No	Yes
μ_{Male}	0.067	0.072
	(0.047)	(0.049)
μ	0.292***	0.285***
	(0.018)	(0.018)
Obs.	39238	37352
Clusters	853	812
AIC	50856.580	48290.993
BIC	50925.199	48512.725

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: PT2 model estimates for those who did not violate stochastic dominance

	(1)	(2)
r_{Male}	0.125***	0.123***
	(0.034)	(0.035)
r	0.430***	0.354***
	(0.015)	(0.049)
Controls	No	Yes
δ_{Male}	0.002	0.001
	(0.008)	(0.008)
δ	1.034***	1.001***
	(0.004)	(0.027)
Controls	No	Yes
γ_{Male}	-0.035***	-0.029***
	(0.010)	(0.010)
γ	0.973***	1.014***
	(0.005)	(0.025)
Controls	No	Yes
μ_{Male}	0.050	0.047
	(0.038)	(0.039)
μ	0.243***	0.240***
	(0.016)	(0.017)
Obs.	29716	28474
Clusters	646	619
AIC	37915.212	36214.049
BIC	37981.608	36428.725

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: ESVT for those who did not violate stochastic dominance

	(1)	(2)
M_{Male}	16.715*	32.555*
	(9.806)	(16.870)
M	12.827***	13.334***
	(1.637)	(4.567)
Controls	No	Yes
α_{Male}	-0.132	-0.286***
	(0.094)	(0.081)
α	1.069***	0.957***
	(0.060)	(0.159)
Controls	No	Yes
μ_{Male}	-0.023***	-0.030***
	(0.006)	(0.005)
μ	0.046***	0.047***
	(0.004)	(0.004)
Obs.	29716	28474
Clusters	646	619
AIC	37895.297	36189.271
BIC	37945.094	36337.892

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: ESVT1 model estimates

	(1)	(2)
M_{Male}	9.381*	27.870**
	(5.307)	(11.946)
M	11.358***	10.649***
	(1.050)	(3.613)
Controls	No	Yes
α_{Male}	-0.088	-0.320***
	(0.100)	(0.078)
α	1.157***	1.108***
	(0.055)	(0.161)
Controls	No	Yes
γ_{Male}	-0.028***	-0.010
	(0.011)	(0.011)
γ	0.969***	1.005***
	(0.006)	(0.030)
Controls	No	Yes
μ_{Male}	-0.025***	-0.037***
	(0.007)	(0.005)
μ	0.059***	0.060***
	(0.004)	(0.004)
Obs.	39238	37352
Clusters	853	812
AIC	50774.148	48195.634
BIC	50842.767	48417.366

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: ESVT2 model estimates

	(1)	(2)
M_{Male}	22.976	45.055**
	(16.926)	(19.571)
M	14.986***	14.183***
	(2.102)	(5.364)
Controls	No	Yes
α_{Male}	-0.083	-0.197***
	(0.101)	(0.072)
α	0.991***	0.905***
	(0.055)	(0.158)
Controls	No	Yes
δ_{Male}	0.005	0.004
	(0.009)	(0.010)
δ	1.025***	1.002***
	(0.005)	(0.027)
Controls	No	Yes
γ_{Male}	-0.030***	-0.018
	(0.011)	(0.012)
γ	0.964***	1.010***
	(0.006)	(0.030)
Controls	No	Yes
μ_{Male}	-0.026***	-0.033***
	(0.008)	(0.005)
μ	0.051***	0.051***
	(0.004)	(0.005)
Obs.	39238	37352
Clusters	853	812
AIC	50699.228	48081.644
BIC	50785.002	48371.601

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Controls include age, income, wealth, whether they obtained a university level education, whether they have children, and whether they have a long-term health condition. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.12: ESVT model estimates with Baillon et al. (2020) data

	(1)	(2)
M_{Male}	37.515***	36.523***
	(12.231)	(11.023)
M	25.755**	-94.911*
	(11.805)	(48.889)
Age	No	Yes
α_{Male}	1.004**	0.839*
	(0.404)	(0.433)
α	1.087***	2.121***
	(0.264)	(0.519)
Age	No	Yes
μ_{Male}	-0.010	0.006
	(0.024)	(0.021)
μ	0.110***	0.092***
	(0.021)	(0.016)
Obs.	9587	9587
Clusters	137	137
AIC	12584.720	12567.378
BIC	12627.729	12624.724

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13: ESVT1 model estimates with Baillon et al. (2020) data

	(1)	(2)
M_{Male}	29.342**	30.752**
	(14.737)	(12.674)
M	31.417**	-105.329**
	(14.205)	(52.742)
Age	No	Yes
α_{Male}	0.978**	0.789*
	(0.464)	(0.470)
α	1.120***	2.154***
	(0.320)	(0.512)
Age	No	Yes
γ_{Male}	0.101**	0.087**
	(0.040)	(0.034)
γ	0.961***	1.056***
	(0.033)	(0.104)
Age	No	Yes
μ_{Male}	-0.008	0.008
	(0.025)	(0.021)
μ	0.111***	0.092***
	(0.022)	(0.015)
Obs.	9587	9587
Clusters	137	137
AIC	12556.317	12538.245
BIC	12613.663	12617.095

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14: ESVT2 model estimates with Baillon et al. (2020) data

	(1)	(2)
M_{Male}	32.852**	29.486*
	(13.996)	(15.509)
M	1.020***	0.989***
	(0.027)	(0.091)
Age	No	Yes
α_{Male}	1.155**	0.910
	(0.517)	(0.563)
α	1.099***	2.260***
	(0.280)	(0.547)
Age	No	Yes
δ_{Male}	0.019	0.026
	(0.032)	(0.029)
δ	1.025***	0.978***
	(0.005)	(0.013)
Age	No	Yes
γ_{Male}	0.100***	0.084***
	(0.038)	(0.032)
γ	0.953***	1.067***
	(0.029)	(0.101)
Age	No	Yes
μ_{Male}	0.001	0.015
	(0.024)	(0.022)
μ	0.112***	0.094***
	(0.020)	(0.015)
Obs.	9587	9587
Clusters	137	137
AIC	12544.800	12529.091
BIC	12616.481	12629.445

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.15: EU and PT model estimates with gender dependent reference points

	EU	PT
r^g_{Male}	0.124***	0.553***
	(0.031)	(0.109)
r^g	0.414***	0.840***
	(0.014)	(0.015)
r^l_{Male}		0.169
		(0.113)
r^l		1.316***
		(0.021)
μ_{Male}	0.040	6.542
	(0.036)	(4.217)
μ	0.250***	3.534***
	(0.015)	(0.200)
Obs.	39238	39238
Clusters	853	853
AIC	51310.934	51246.883
BIC	51345.244	51298.347

Notes: θ_{Male} is the mean difference between males and females in parameter θ. Robust standard errors clustered on individual in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B. Figures

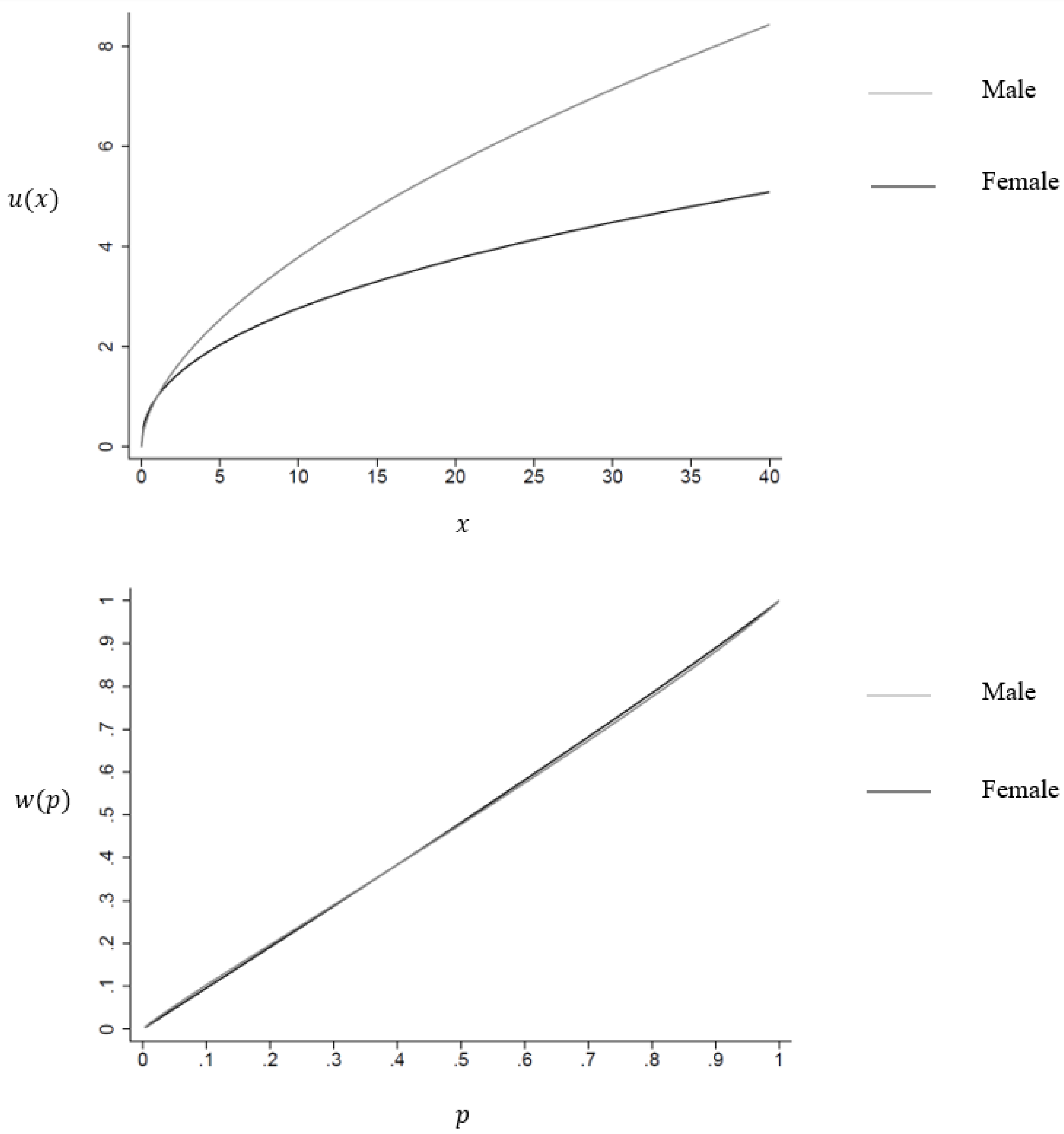

Figure B.1: Utility and probability weighting functions by participant gender

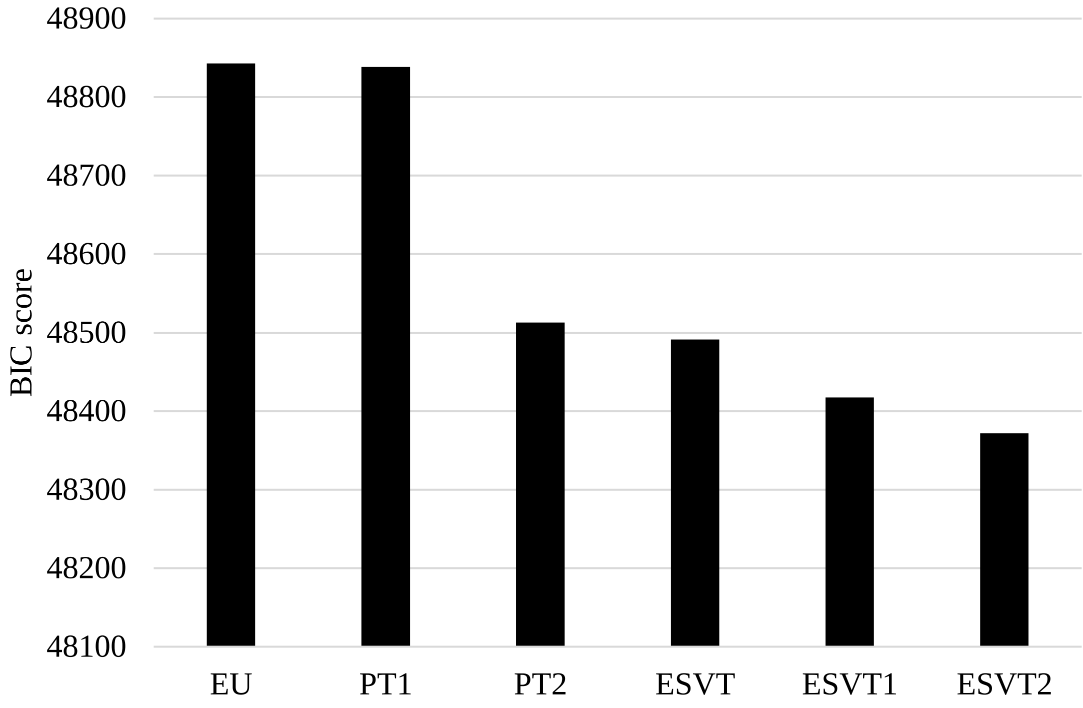

Figure B.2: BIC scores across models

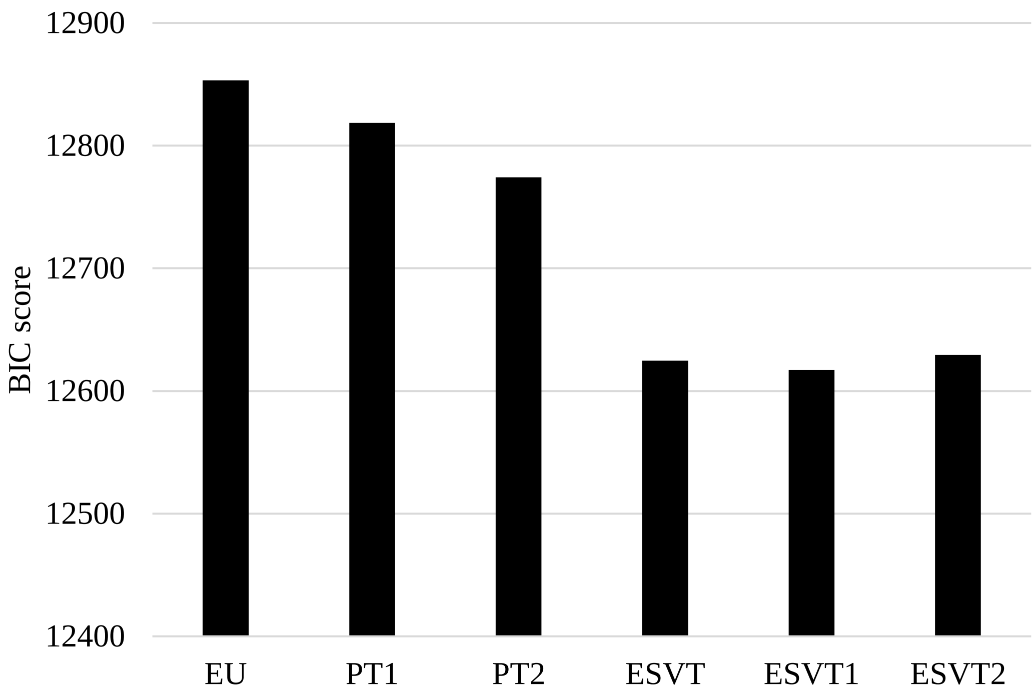

Figure B.3: BIC scores across models with Baillon et al. (2020) data

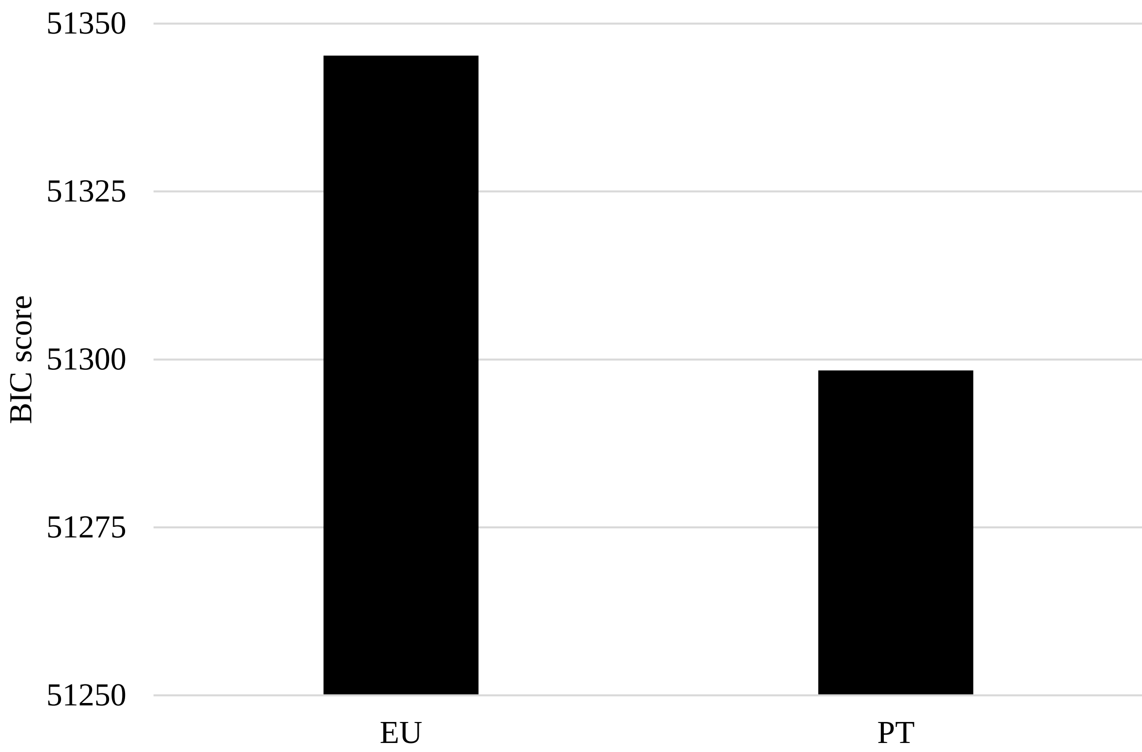

Figure B.4: BIC scores across models

C. Models used for estimation

Assume, the expected utility function v of receiving $\$x$ with probability p is given by:

$$v(p, x) = w(p)u(x)$$

where w is the probability weighting function and u is the utility function.

We employed the following six specifications for the above equation:

1. EU: expected utility

$$v(p, x) = px^r$$

2. PT1: prospect theory with one parameter probability weighting $w(p)$ as in Prelec (1998)

$$v(p, x) = e^{-(-\ln(p))^\gamma} x^r$$

3. PT2: prospect theory with two parameter probability weighting $w(p)$ as in Prelec (1998)

$$v(p, x) = e^{-\delta(-\ln(p))^\gamma} x^r$$

4. ESVT: expected subjective value theory as in Glimcher & Tymula (2023)

$$v(p, x) = p \frac{x^\alpha}{x^\alpha + M^\alpha}$$

5. ESVT1: expected subjective value theory as in Glimcher & Tymula (2023) with one parameter probability weighting $w(p)$ as in Prelec (1998)

$$v(p, x) = e^{-(-\ln(p))^\gamma} \frac{x^\alpha}{x^\alpha + M^\alpha}$$

6. ESVT2: expected subjective value theory as in Glimcher & Tymula (2023) with two parameter probability weighting $w(p)$ as in Prelec (1998)

$$v(p, x) = e^{-\delta(-\ln(p))^\gamma} \frac{x^\alpha}{x^\alpha + M^\alpha}$$