TI 2016-063/I
Tinbergen Institute Discussion Paper

Gender and Redistribution:

Experimental Evidence

Thomas Buser[1,3]
Louis Putterman[2]
Joël van der Weele[1,3]

[1] *Amsterdam School of Economics, Faculty of Economics and Business, University of Amsterdam;*
[2] *Brown University, United States;*
[3] *Tinbergen Institute, the Netherlands.*

Gender and Redistribution: Experimental Evidence

Thomas Buser[*], Louis Putterman[†] and Joël van der Weele[‡]

August 19, 2016

Abstract

Gender differences in voting patterns and political attitudes towards redistribution are well-documented. The experimental gender literature suggests several plausible behavioral explanations behind these differences, relating to gender differences in confidence concerning future relative income position, risk aversion, and social preferences. We use data from lab experiments on preferences for redistribution conducted in the U.S. and several European countries to disentangle these potential mechanisms. We find that when choosing to redistribute income as a disinterested observer, women choose higher tax rates than men when initial income depends on performance in a task but not when it is randomly allocated. In a veil of ignorance condition with uncertainty about the income position of the decision maker, this effect is even stronger, leading to a 10ppt gender difference in average chosen tax rates in the performance conditions. We find that this gender difference is mainly due to men being more (over)confident about their task performance and the resulting income position, with gender differences in risk aversion and social preferences playing a minor role.

JEL: C91, J16, H24, D31

Keywords: gender, redistribution, overconfidence, risk attitudes, voting, taxation

[*]University of Amsterdam and Tinbergen Institute. t.buser@uva.nl. School of Economics, Roetersstraat 11, 1018WB Amsterdam, The Netherlands.

[†]Brown University. Louis_putterman@brown.edu

[‡]University of Amsterdam and Tinbergen Institute. vdweele@uva.nl

1 Introduction

In many Western countries, attitudes towards redistribution differ by gender. Based on data from large-scale surveys such as the Word Value Survey or the General Social Survey, multiple studies find that conditional on socio-economic controls, women are more in favor of governments redistributing incomes than are men (Alesina and Giuliano, 2010; Luttmer and Singhal, 2011). Consistent with such attitudes, women show higher support for left-wing parties than men, resulting in substantial voting gaps at the polls (Shapiro and Mahajan, 1986; Inglehart and Norris, 2000; Giger, 2009).

Findings from the experimental behavioral economics literature, surveyed in more detail below, suggest some tentative explanations for these gender gaps. First, the experimental literature finds that men are consistently more (over)confident concerning their own abilities and overestimate their relative income. Second, women are consistently found to be more risk averse, which could translate into higher support for redistribution as a form of insurance against income uncertainty. Third, some studies find that women make more egalitarian choices, although these gender differences seem to depend a lot on the exact experimental setting. However, none of these gender gaps have been directly linked to redistributive preferences.

We study the behavioral roots of gender differences in political attitudes, using data from two studies: Durante et al. (2014, henceforth DPW) and Farina et al. (2015, henceforth FGS). These studies share a similar design, in which participants redistribute an initial unequal income distribution in groups of 21 participants. Depending on the experimental condition, a participant's initial position in the income distribution is either random, based on the average income at their place of birth, or decided by their relative performance in a cognitive or effort task. Participants then choose tax (redistribution) rates under three different scenarios: as an impartial observer, behind a veil of ignorance about their own relative position, and with full knowledge about their position in the distribution. The experiments were conducted at eight different locations in the US, Italy, Germany and Norway.

When choosing as an impartial observer, we find that in the aggregate, women favor higher tax rates than men in the conditions where initial income depends on performance in a task, but not when it is allocated randomly or on the basis of a participant's origin. This gender difference is even stronger in a veil of ignorance condition, where participants' own incomes are affected by taxation but where they are ignorant of their exact place in the initial income distribution. We show that most of the gender difference in tax choices under the veil of ignorance is due to men being more confident than women in their own relative performance, conditional on actual rank. Risk preferences and social preferences explain a much smaller part of the gender difference. This pattern holds both in the US and in Europe and is robust across a range of model specifications. The data are less informative about the reason why female participants favor more equality when deciding to redistribute as an impartial observer, but we explain how the same factors could potentially be at work.

Our results add to the empirical literature in economics and political science about preferences for redistribution, which is almost exclusively based on survey evidence. While many survey studies can control for income, it is harder to control for income expectations and risk attitudes. By contrast, the laboratory allows us to use monetary tradeoffs to tease apart different potential sources of the gender gap, highlighting the role of uncertainty and confidence about future income. In doing so, we also contribute to the literature in experimental economics that has not explored gender differences in redistribution except in abstract two-person games. DPW and FGS elicit tax choices in larger groups (21 subjects) and with a pre- and post-tax incomes setting that is suggestive of a macro-economic or political economy framing. The combination of data from several studies and across several countries lends robustness and credibility to the results.

2 Literature

There is a substantial literature in experimental economics and psychology on gender differences in behavior. Here, we discuss briefly the main gender differences found in the experimental economic literature as well as gender differences in political attitudes and voting behavior.

Social preferences in experiments. The experimental economics literature mostly measures preferences for fairness or a concern with social outcomes through simple division games such as dictator and ultimatum games, or dilemma situations like the public goods game. Decisions in these games reveal whether people are willing to sacrifice in order to improve the income of others. Several surveys look at gender differences in such "social" preferences. Croson and Gneezy (2009) survey a large number of studies, and conclude that there is no strong evidence for gender differences, although women's social decisions are more context dependent than men's, i.e. more reactive to manipulations in framing and incentives. Engel (2011) shows in a meta study of the dictator game that women are marginally more generous than men. Eckel and Grossman (2008a) find that there are no large differences in social decisions in public good, ultimatum and dictator games when subjects are exposed to risk. However, when risk plays no role, women make less individually oriented and more socially oriented decisions. Cappelen et al. (2015) find that gender differences in trust games are more pronounced in a representative sample of the population than in the student samples that are usually used in experiments.

When it comes to equality-efficiency tradeoffs, several papers find that women are more focused on equality than men (e.g. Andreoni and Vesterlund, 2001; Schildberg-Hörisch, 2010; Cappelen et al., 2014; Fisman et al., 2014). In the study closest to ours, Ackert et al. (2007) study redistributive tax choices in groups of 9 people with randomly allocated endowments and find no evidence for gender differences. This is in line with our findings, which show no significant gender differences in redistribution when initial endowments are randomly distributed.

Risk aversion. There is evidence that women are more risk averse than men. Croson and Gneezy (2009) survey ten experimental studies that elicit risk preferences using both real and hypothetical gambles. All studies find that women are (weakly) more risk averse than men. This conclusion is echoed by field studies that look at portfolio selection. Eckel and Grossman (2008b) survey 16 studies in experimental economics and also find that women take less risk, unless the framing of the experiment is contextual. Charness and Gneezy (2012) survey a particular risk preference elicitation methods, the Investment Game of Gneezy and Potters (1997), and find that men take substantially more risk than women.

Filippin and Crosetto (2016) criticize these surveys for using a selective sample of elicitation methods, and offer a more nuanced conclusion from their own literature survey: gender effects depend a lot on the details of the task, like the presence of a safe option. Analyzing almost a 100 studies using the Holt and Laury (2002) risk elicitation task, they find evidence of a small gender difference of about 1/6th of a standard deviation. Byrnes et al. (1999) conduct a meta-analysis with 150 studies in psychology, using a a broad range of risky behaviors like smoking, driving and gambling. The elicitation methods include self-reports, incentivized experiments and observed choices. The study finds that in most categories, males take more risks than females. However, the effect size is usually small and not consistent across all studies.

Overconfidence. A large literature shows that people are generally overconfident about their own abilities (Moore and Healy, 2008). There is some evidence that men are generally more confident, and hence more overconfident than women. For example, Deaux and Farris (1977) find that men evaluate their task performance more favorably than women, and tend to attribute their performance more to skill rather than luck. Lundeberg et al. (1994) find that women are less likely to be overconfident about the accuracy of (wrong) answers to exam questions. Estes and Hosseini (1988) ask people to evaluate the financial report of a company and decide upon a (fictional) investment in the company. They find that men are more confident about the correctness of their investment decision. Niederle and Vesterlund (2007) administer a number-addition task in groups of four, and measure beliefs about relative performance, giving a $1 reward for a correct assessment of performance ranks. They find that 75 percent of the men think they are the best performer in their group of four, while only 43 percent of the women hold this belief. Buser et al. (2016) do a similar exercise for three different cognitive tasks, and find that controlling for ability, women are on average 3 percentage points less confident that they are above the median performance in their group. There is evidence that the gender gap in overconfidence is greatest for tasks that are perceived to be "masculine", like mathematical exercises (e.g. Beyer and Bowden, 1997).

Some survey studies focus explicitly on differences in economic expectations between women and men, finding that women are more pessimistic about the economy (Clarke et al., 2005). Smith and Powell (1990) show that among college graduates, men are also more likely to overestimate their future earnings.

Political preferences for redistribution. Surveys yield quite consistent evidence for gender differences in redistributive attitudes. Early studies simply compare self-reported attitudes between the sexes. Using the General Social Survey and several other surveys, Shapiro and Mahajan (1986) find about a 3 percentage points gender differences in support for economic policies to help the poor and targeted groups, with women being more in favor. Using the American National Election Study, Howell and Day (2000) find that more egalitarian attitudes and a higher valuation of helping others explain much of the political gender gap on redistributive attitudes. Eagly et al. (2004) find that women score higher on "socially compassionate" values, and these scores are mediated by self-reported ideological values (e.g. conservatism and support for social equality).

Several studies have used stated support for redistribution as an independent variable in regression analyses on background variables including gender. Alesina and La Ferrara (2005) show a highly significant gender dummy in the General Social Survey, controlling for income and education. Inglehart and Norris (2000) and Alesina and Giuliano (2010) echo these results across countries, using the World Values Survey, and Luttmer and Singhal (2011) find the same result in Europe using the European Social Survey. Funk and Gathmann (2015) use data from telephone surveys that follow individual voting behavior in Swiss referenda, and find that women are more prone to vote for generous welfare policies.

Voting patterns. In political science there has been much attention to the gender gap in voting behavior. In the U.S., women vote more Democratic starting in the 1970s (Shapiro and Mahajan, 1986; Manza and Brooks, 1998). The gender gap seems to have grown over time, with the 20 percentage points gender gap in the 2012 presidential elections being the largest in U.S. history.[1] The gender voting gap has also been documented in European countries, where women generally vote for more "left-wing" parties (Inglehart and Norris, 2000; Giger, 2009). Note however that this voting gap is a phenomenon that has emerged only over the last decades, with women more likely to back conservative parties before the 1970s. This suggests that preferences for redistribution cannot be the only reason for voting gender gaps, but other, structural factors are at work as well (Manza and Brooks, 1998; Iversen and Rosenbluth, 2006).

3 Design

In this study, we use existing data from DPW and FGS. As the design of FGS was inspired by DPW, we start by outlining the design of the latter. In the second part of this section, we discuss the main differences of FGS with respect to DPW.

[1] See http://www.gallup.com/poll/158588/gender-gap-2012-vote-largest-gallup-history.aspx

3.1 Design of the DPW study

Sixteen experimental sessions with twenty-one subjects each were conducted in a computer lab at Brown University. A total of 336 undergraduate students from a wide range of disciplines participated. Each session, which took about 90 minutes, began with instructions on participants' computer screens being read aloud by the experimenter. Subjects were promised a $5 show-up fee and were told there would be an additional payoff the size of which would depend on the outcome of the experiment.

The core decisions, made in two sequential "parts," centered on the choice of tax rates that could partly or wholly equalize an unequal set of twenty pre-tax payoffs, ranging from $0.11 to $100. These income levels were chosen in proportion to the pre-tax incomes of given twentieths (vigintiles) of the United States population, based on U.S. Census data for the year 2000. In each part, participants chose a tax rate between 0% and 100%, in increments of 10%. One individual's choice was randomly implemented for the session ex post ("Decisive Individual"), with each participants' choice being equally likely to be decisive. The chosen tax rate was applied to the pre-tax income distribution and the tax proceeds were distributed equally amongst all participants (except one, as we explain below). Thus, a tax rate of 0% left the original pre-tax distribution in place, while a tax rate of 100% induced full income equality. The consequences of each tax choice for the post-tax income distribution were introduced verbally, graphically and by means of a formula.

In each part, subjects participated in four conditions with differing bases (methods) for assigning pre-tax earnings, one of which was randomly selected near the end of the session. A different tax rate could be chosen by each subject in each of the four conditions. Specifically, there were two methods the experimenters expected to be viewed as arbitrary by subjects—strictly random assignment (henceforth dubbed Random) and assignment based on parental socioeconomic status (where initial income increased with the average income in the subject's place of origin, henceforth dubbed Origin). The two remaining methods were expected to be viewed by some as bases of just claims or entitlement—performance on a general knowledge Quiz, and success playing a computer game, Tetris. We henceforth combine decisions for the Quiz and Tetris conditions (and for two similar conditions in FGS) under the heading Performance.[2]

In Part 1 of the experiment, the decision-maker was in the role of a "disinterested observer", and her tax choice redistributed earnings among the other twenty participants. To place the decision-maker in as disinterested a position as possible, subjects were truthfully informed that if randomly selected as the decisive individual, their own earnings would be randomly drawn from a uniform distribution over a small interval (of span $2.00) with the minimum value set at the average of the pre-tax earnings of the remaining twenty subjects, $19.80.[3] Subjects completed the comprehension

[2]Place of origin information was obtained during the initial log in procedure before any information about the experiment, and the Quiz and Tetris tasks were completed by all participants following their tax decisions in the second part of the experiment.

[3]The random element was included so that own identity as decider could not be inferred from payout. This

questions and decisions of Part 1, knowing only that another experimental part sharing some (but not all) features would follow, and that each part was equally likely to determine their earnings.

In Part 2, the decision-maker was among the twenty participants whose income was affected by the tax choice, but subjects chose their favored tax rate from behind a veil of ignorance about what their own pre-tax earnings would be.[4] After reading the Part 2 instructions and answering the comprehension questions for that part, subjects were asked to predict their relative ranking (with the exception of subjects in the Random condition).[5]

After the completion of Part 2 and the Quiz and Tetris tasks, a coin toss decided whether Part 1 or Part 2 would determine earnings. If Part 2 was selected, each participant was then shown her pre-tax earnings rank and (without prior notice) given the option to change her tax choice(s). This decision, dubbed Part 3, generated a third take on redistribution: decisions by interested parties after lifting the veil of ignorance. Subjects received no feedback regarding others' tax decisions and learned their relative performance in the Quiz and Tetris portions only if Part 2 was chosen by the subsequent coin toss, leading to the invitation to reconsider taxes.

Before concluding the session and learning their payoffs, subjects participated in an incentivized task consisting of five choices between a certain payment and a lottery. This task was designed to elicit risk attitudes using the "multiple price list" method introduced by Harrison and Rutström (2008). Subjects also answered a background survey including a series of questions on personal characteristics and attitudes.

In order to gauge strength of preference for equalizing income, subjects were randomly assigned a direct cost of either zero, \$0.25, \$0.50 or \$1.00 per 10% increment of selected tax rate (so, for instance, selecting a tax rate of 80% cost \$4.00 in the \$0.50 tax cost treatment). This tax cost is identical for all subjects in the same session and constant across the three parts of the experiment.[6] Apart from this tax cost, the other dimension which varied across sessions was a shrinkage of the redistributed amounts, representing an "efficiency loss" or a "leaky bucket effect". This loss was 0% in some sessions, where subjects heard no mention of it, or equal to 12.5% or 25% of taxed-and-redistributed earnings in others.[7] The crossing of the four tax costs with three efficiency loss rates yields 12 treatments. Two sessions of each treatment with efficiency loss 0, and one session

was done to minimize concern over social discomfort following revelation of the final experimental outcome to all participants.

[4]Because the decider stood as a bystander to a group of twenty others in Part 1, but was one of the twenty directly affected by the tax in Part 2 and Part 3, symmetry called for random selection of one of the twenty non-decisive subjects to receive an amount drawn from the \$19.80 to \$21.80 interval if Part 1 were not selected. This randomly selected twenty-first individual was thus unaffected by the chosen tax rate, as the decider would be under a Part 1 outcome.

[5]In estimating their rank, subjects chose one of seven ranges grouping together ranks 1-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20. In addition subjects indicated whether they were "very confident", "somewhat confident" or "not confident at all" about their predictions.

[6]Subjects in the \$0 tax cost treatment heard nothing about such costs existing in other sessions, just as subjects in each of the other treatments knew of the parameters of their own treatment only.

[7]Note that the decider's income was not affected by efficiency loss, if present, in case a Part 1 decision was implemented, but she was affected if Part 2 (3) was applied.

of each treatment with efficiency loss 12.5% and 25%, were conducted.

A more detailed description of the design including the instruction materials can be found in DPW. All the instructions are available at: www.brown.edu/Research/IDE/walkthrough.

3.2 Design of the FGS study

The design of FGS is based on DPW and shares its essential features, including the division in Parts 1, 2 and 3. Sessions were performed on university campuses in seven cities in four different countries: Munich and Bremen in Germany, Milan and Salerno in Italy, Oslo in Norway, and Oxford (MS) and Pullman (WA) in the United States. In each of these locations, 168 university students participated in the experimental sessions.

There are some differences with respect to the DPW design that we enumerate here. First, the source of the initial income inequality (Random, Origin, Performance) is varied between subjects rather than within. Put differently, whereas in DPW the different sources of inequality are conditions present in each treatment, they constitute separate treatments in FGS. Second, in FGS there were no treatments featuring efficiency losses or direct costs of implementing higher taxes. Third, the knowledge quiz of DPW is replaced by an intelligence task, and the Tetris game of DPW is replaced by a boring real-effort task that rewards effort rather than skill. Parallel to our combining of Tetris and Quiz decisions in DPW, we treat decisions in both the intelligence and real effort tasks of FGS as falling in the condition Performance.

Fourth, the income distribution is linear and in tokens, instead of being calibrated on the U.S income distribution as in DPW. This means that it is also more compressed: income varies from \$1.30 for the poorest to \$27.30 for the richest participant at FGS's two U.S. sites (and similar amounts in the European samples), whereas income varies between \$0.11 and \$100 in DPW. Fifth, in FGS subjects perform a new task in each round before they choose their favored tax rate, while in DPW all tasks are performed only once after the second tax choice. Sixth, FGS features an additional round between Parts 2 and 3, where people learn their earnings in Part 2, but not yet the actual income in that round, which is determined on the basis of a new performance on the task. This means participants can form more informed opinions on their rank than in Part 2, without being entirely sure about it (as they are in Part 3). We excluded this condition from the data, as it is not comparable to any condition in DPW. Finally, there are some subtle differences in the risk aversion test and the phrasing of the questionnaire items on determinants of income mobility and political orientation. In the appendix, we detail how we standardized these items.

While we do not want to downplay these differences, we believe that the choices in the two studies can be meaningfully compared. In both studies, participants choose tax rates between 0% and 100% for a group of 21 individuals, facing very similar tradeoffs between equality, risk and personal payoffs. Moreover, in the appendix we perform an extensive series of robustness checks, including specifications where we control for the actual individual cost of taxation (taking into account the influence of the tax cost and efficiency loss and the different income distributions used in the two studies), interact confidence and risk measurements with dummies for data origin (DPW vs. FGS), and use non-parametric versions of our confidence and risk controls. None of these checks alters our results, reinforcing our confidence in the robustness of our findings.

4 Results

In this section we analyze choices both in the disinterested observer condition as well as behind the veil of ignorance. We have substantially more observations from the US than from Europe (1,726 vs 839 for Parts 1 and 2), which means that the sample averages are dominated by the US results. In the graphs and regressions, we therefore show weighted results (whereby each continent is given equal weight). As the US and Europe have quite different traditions when it comes to the level of redistribution, we also discuss the result separately by continent. In the appendix, we present results from unweighted regressions and show that all main results are robust.

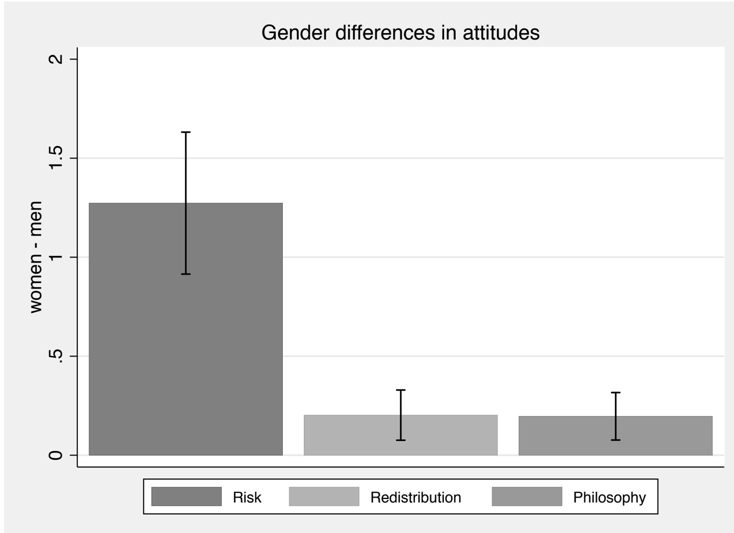
4.1 Gender differences in attitudes and disinterested redistribution choices

Figure 1 shows gender differences in risk aversion, self-rated survey attitudes towards redistribution[8] and political philosophy (where higher values mean more left-wing)[9]. The data replicate patterns in the literature, as women are significantly more in favor of redistribution and more left-leaning. Women also make significantly more risk-averse choices in the risk task.

[8]In DPW, attitudes towards redistribution are measured by the question "In general, do you think there is too little income redistribution". Answers are on a scale from 1 (disagree strongly) to 7 (agree strongly). In FGS, it is based on the question "Do you think gov't should reduce income differences?". Answers are on a scale from 1 (disagree strongly) to 5 (agree strongly). For the FGS data, the variable has been rescaled to have a range of 1 to 7 (see appendix).

[9]In DPW, political philosophy is measured by the question "Which of the following best describes your political philosophy (ideology)?" Answers are on a scale from 1 (very conservative) to 7 (very liberal). In FGS, it is based on the question "In political issues people often refer to positions of 'left' and 'right'. Where would you locate your opinions in the following scale, where 1 means 'Left' and 10 means 'Right'?". For the FGS data, the variable has been rescaled to have a range of 1 to 7 (see appendix).

Figure 1: Gender differences in attitudes

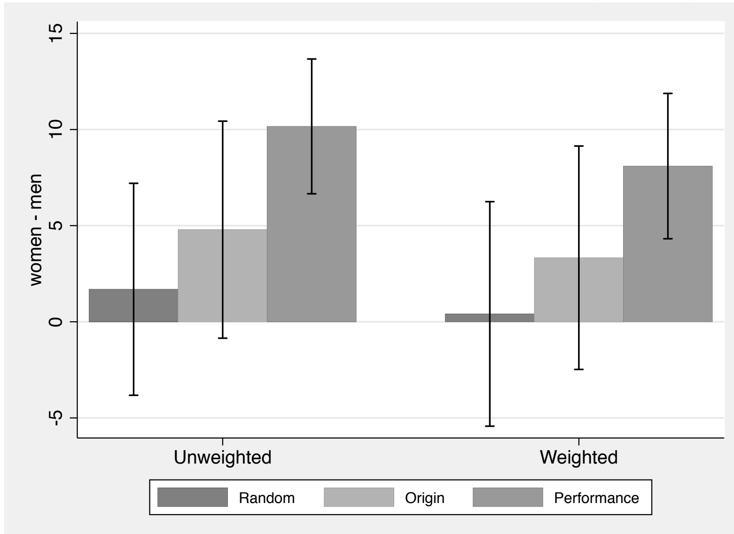


Note: The bars show gender difference whereby a positive value means a higher average value for women. "Risk" means risk aversion and is measured on a scale from 0 to 15. "Redistribution" means attitudes towards redistribution, where higher means more favorable, and is measured on a scale from 1 to 7. "Philosophy" means political philosophy on a left-right scale, where higher means more left-leaning, and is measured on a scale from 1 to 7. Observations from Europe are overweighted such that Europe and US are given equal weight.

In Figure 2 and Table 1, we analyze gender differences in favored tax rates in Part 1, where subjects are "impartial observers" choosing for others. Women do not act more inequality averse per se: women and men choose very similar tax rates when income is randomly allocated (49 percent for men vs. 51 percent for women; p=0.547, t-test). Men but not women choose somewhat lower tax rates when the distribution is based on people's origin (46 percent vs. 51 percent; p=0.096) but this difference is smaller and not significant when continents are given equal weight. Table 1 reports coefficients from regressions of a female dummy on chosen tax rates in the different conditions and also reports p-values for the difference in the female coefficient between conditions.[10] The difference in the gender effect between the Random and Origins conditions is not significant.

[10]Although determinant of income varies by treatment (i.e. between subjects) in FGS, it is experienced as one of multiple (within subject) conditions in DPW, as noted above. For convenience, we use the term "condition" in the remainder of the paper to cover both cases.

Figure 2: Tax by gender and condition (Part 1)



Note: The bars show gender differences in average chosen tax rates whereby a positive value means a higher average tax rate for women. "Weighted" means that observations from Europe are overweighted such that Europe and US are given equal weight.

Table 1: Gender differences in tax rates (Part 1: disinterested decision maker)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	4.742***	0.633	3.598	8.049***
	(1.694)	(2.889)	(2.920)	(2.066)
p-val vs. (2)			0.411	0.029
Observations	2565	644	628	1293
R^2	0.030	0.060	0.030	0.031

Note: All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Interestingly, women do redistribute significantly more than men when income depends on performance in the cognitive/effort tasks, which is mainly due to men favoring substantially lower tax rates in this case (33 percent vs. 43 percent; p=0.000). The data thus show that men are less in favor of redistribution when income differences are based on performance than when they are based on luck or origin, indicating that they see performance as an appropriate justification for inequality. This tendency is significantly weaker for women. Table 1 shows that the difference in the gender effect between the Random and Performance conditions is statistically significant. In the appendix, we present unweighted regressions (Table A1) and separate regressions for the US and European samples[11] (Table A5). While the difference in the gender effect between the

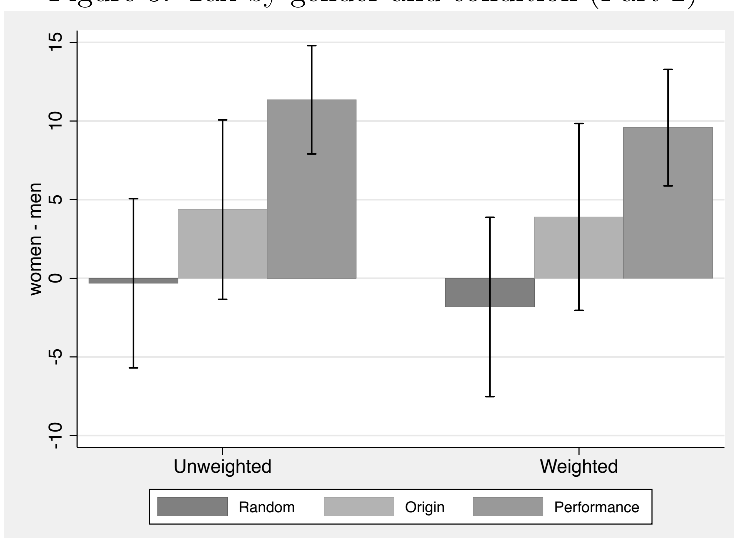
[11] In this and all following analyses, the US sample groups the data of DPW with the FGS data from the Pullman, WA and Oxford, MS sessions. The Europe sample consists of the FGS data from the Milan, Salerno, Oslo, Bremen and Munich sessions.

Random and Performance conditions goes in the same direction on both continents, it is roughly half as large and not statistically significant in Europe.

4.2 Redistribution choices behind the veil of ignorance

We will now look at the tax rates favored by men and women in Part 2. Here, the chosen tax rate affects the decider's own income, although his/her relative position in the income distribution is uncertain. Figure 3 shows that the pattern is similar to the one observed in Part 1. There is no gender difference in tax rates when income is randomly determined (men and women on average both favor a tax rate of 52 percent; p=0.91, t-test) but there is a large and significant gap when income is based on relative performance in the Performance conditions (average tax rates are 31 percent for men and 43 percent for women; p=0.000). This gender difference is slightly larger and more precisely estimated than in Part 1.

Figure 3: Tax by gender and condition (Part 2)



Note: The bars show gender differences in average chosen tax rates whereby a positive value means a higher average tax rate for women. Observations from Europe are overweighted such that Europe and US are given equal weight.

Table 2 shows results from weighted regressions that confirm that the difference in the gender gap between the Random and Performance conditions is large (11.4 percentage points) and statistically significant. Table A6 in the appendix shows that this result holds for the US and Europe separately. The difference in the gender gaps between the Random and Origin conditions is also larger than in Part 1 but still not statistically significant.

Table 2: Gender differences in tax rates (Part 2: behind the veil of ignorance)

	(1)	(2)	(3)	(4)
	All	Random	Origin	Performance
Female	5.206***	-1.740	4.081	9.680***
	(1.633)	(2.877)	(3.000)	(2.013)
p-val vs. (2)			0.137	0.001
Observations	2565	644	628	1293
R^2	0.015	0.033	0.011	0.034

Note: All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

We will now turn to the question of why men redistribute less than women in the Performance condition compared to the other conditions. Given that no significant gender difference in chosen tax rates is observed when income is randomly allocated, gender differences in pure preferences for equality are unlikely to be the main explanation. We will therefore concentrate on gender differences in (over)confidence and risk aversion as alternative potential explanations.[12]

Figure 4 shows the rank expectations of experimental subjects by true rank and gender. There is little variation in average estimated ranks across the performance distribution. That is, subjects are on average quite bad at estimating their relative performance (the correlation between true and guessed rank is highly statistically significant but quite weak at 0.23). This also means that subjects in the lower half of the performance distribution tend to be overconfident and subjects in the upper half of the performance distribution tend to be underconfident. Figure 4 also clearly demonstrates that men are more confident than women about their rank across the whole performance distribution. In Table 3, we regress the expected rank on true rank and a gender dummy. Men expect to rank 2.5 places higher than women with the same actual performance level. Men are also marginally significantly more confident about their rank in the "Origin" condition where they expect to rank 0.7 places higher than women conditional on true rank.

[12]To be sure, the similarity with the difference between men and women in Part 1 could suggest a preference effect specific to the source of earnings. For now, we focus on the more robust difference in Part 2. We return to the Part 1 difference in the discussion section.

Figure 4: Confidence by gender and true rank (weighted)

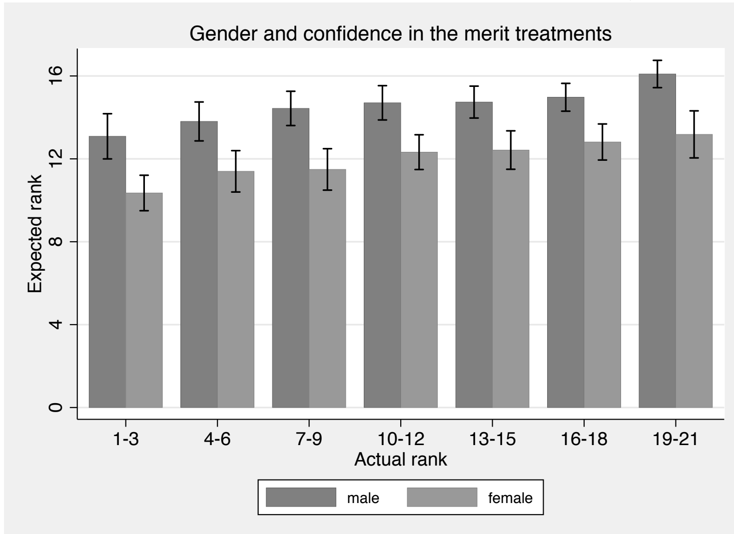


Table 3: Gender difference in confidence

	(1) Origin	(2) Performance
Female	-0.688*	-2.502***
	(0.400)	(0.247)
Rank dummies	√	√
Observations	628	1293
R^2	0.123	0.179

Note: All regressions control for rank dummies and site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

We will now investigate whether these differences in (over)confidence can provide an explanation for the gender differences in chosen tax rates. In Table 4, we regress chosen tax rates on gender and actual rank. In Column 7, we restrict the sample to observations from the Performance conditions, confirming that there is a significant gender difference in tax rates of 8.7 percentage points when controlling for actual rank (and therefore income). The explanatory power of risk attitudes for this gender difference in tax rates is quite small: the gender coefficient is reduced by 13 percent when the risk aversion control is added in Column 8. When we control for overconfidence (that is, guessed rank minus actual rank) in Column 9, the coefficient on the gender dummy is reduced by 62 percent. Together, confidence and risk attitudes can explain 70 percent of the gender difference in tax rates in the Performance condition and 53 percent of the difference in the gender coefficients between the Performance and Random conditions. The table also reports bootstrapped p-values for the significance of the change in the female coefficient caused by controlling for risk attitudes and confidence. Both changes are statistically significant at the 1-percent level.

In Table A3 in the appendix, we show that these results also hold when observations are

Table 4: The effect of confidence and risk aversion on the gender difference in Part 2 (weighted regressions)

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	-1.748	-2.773	3.645	2.284	1.413	0.311	8.709***	7.603***	3.316	2.654
	(2.882)	(2.902)	(2.997)	(3.032)	(2.813)	(2.819)	(2.004)	(2.022)	(2.060)	(2.072)
Actual Rank	-0.0465	-0.0313	-0.859***	-0.763***	-3.321***	-3.206***	-0.847***	-0.812***	-2.680***	-2.587***
	(0.237)	(0.236)	(0.250)	(0.248)	(0.331)	(0.333)	(0.154)	(0.154)	(0.284)	(0.286)
Risk aversion		0.768**		1.158***		0.964***		0.869***		0.665***
		(0.355)		(0.369)		(0.356)		(0.244)		(0.243)
Overconfidence					-3.045***	-3.002***			-2.150***	-2.073***
					(0.317)	(0.315)			(0.273)	(0.274)
p-value (Δ Female)		0.050		0.021	0.080	0.013		0.002	0.000	0.000
Observations	644	644	628	628	628	628	1293	1292	1293	1292
R^2	0.033	0.041	0.032	0.048	0.197	0.208	0.059	0.073	0.127	0.135

Note: All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

not weighted to give Europe and the US equal weight, and in Tables A7 and A8 we show that controlling for confidence significantly reduces the gender coefficient both in the US and in Europe separately. Keeping in mind that our controls are surely measured with error, we conclude that most of the difference in gender gaps between the Random and Performance conditions in Part 2 (choices behind a veil of ignorance) can be explained by these two factors, with gender differences in confidence playing by far the largest role.[13]

In Table 4, we also investigate whether gender differences in confidence and risk attitudes can explain the smaller gender difference in favored tax rates in the Origins condition (Columns 3 to 6). Controlling for risk attitudes reduces the gender coefficient by 37 percent and controlling for confidence leads to a reduction of 61 percent. Together the two factors explain almost the entire gap. While the estimates of the initial gender gap in the Origins condition are fairly imprecise and not statistically significant, the effects of confidence and risk attitudes on the gender gap are statistically significant. The same applies in the case of the Random condition (Columns 1 and 2) where controlling for risk attitudes changes the gender coefficient slightly but significantly.[14]

In Table 5, we analyze gender differences in Part 3 where subjects are fully informed about their own endowment. Conditional on actual rank, women choose slightly *lower* tax rates in all conditions except the Origins condition, but none of the coefficients are significant. There are also no significant differences in the gender coefficient across conditions, suggesting again that social preferences are unlikely to be an explanation for the higher tax rates chosen by women when income is uncertain.

Table 5: Gender differences in tax rates (Part 3: known rank)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	-1.879	-4.323	1.880	-2.339
	(1.743)	(3.074)	(3.305)	(2.311)
p-val vs. (2)			0.144	0.719
Observations	1813	456	440	917
R^2	0.368	0.412	0.325	0.385

Note: All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

4.3 Robustness checks

The appendix contains a range of further robustness checks. Because tax rates, our outcome variable, are bounded at 0 and 100, we show that the results hold when using Tobit instead of OLS (Tables A9 to A11). In Table A12, we show that the results hold when on top of actual and

[13]See Gillen et al. (2015) on a discussion of measurement error in experimentally elicited controls.

[14]The DPW experiment did not measure subjective rank (and therefore overconfidence) in the Random condition.

expected rank, we also control for the actual and expected individual marginal cost of taxation. That is, we calculate the net cost to each individual of raising the tax by 10 percentage points given their rank (actual or expected), as well as their tax cost and efficiency loss (for the DPW observations). This also takes into account the quite different income distributions in the two studies.

In Table A13, we show results from regressions that control for actual rank, risk aversion and overconfidence non-parametrically by using a separate dummy variable for each possible level of the variables. While this technique controls more completely for these factors (see Gillen et al., 2015) it makes it impossible to present all the coefficients, which is why we chose to present linear controls in the main tables.

Finally, in Table A14 we show results from regressions where we interact the controls for actual rank, risk aversion and overconfidence with a dummy for the data source to take into account that the different designs and the scaling of variables may lead to different effects on tax rates in the two datasets (note that all regressions control for the data source through site dummies). The results do not change but actual and expected rank tend to have a stronger negative effect on chosen tax rates in the DPW experiment, possibly reflecting the higher income inequality between top and bottom ranked subjects in the DPW design.

5 Discussion and conclusion

We investigate gender patterns in redistributive choices, using experimental data from multiple locations in the U.S. and Europe. We find that women tend to be more sensitive to the origin of inequality: Women specify higher tax rates than men when initial inequality is "earned" by performance on a task, but not when inequality is determined randomly or on the basis of socioeconomic background. When making decisions under a "veil of ignorance", where choices affect participants' own income and they face uncertainty about their position in the income distribution, the average gender gap is almost 10 percentage points in chosen tax rates when income is due to performance. We can explain this in large part by the fact that women are less (over)confident and more risk averse than men. This difference disappears when uncertainty is lifted in Part 3, demonstrating that social preferences are not a likely explanation for the observed gender differences in chosen tax rates. We find the same patterns both in the US and in Europe and our results are robust to a range of different specifications.

A similar pattern occurs in the condition where participants act as a disinterested observer (although it is less strong and not significant in the European sample). One interpretation of this result is that men see performance as a stronger justification to reduce redistribution, even if the lack of gender difference in Part 3 of the experiment indicates that this difference is not strong enough to influence choices where own income is at stake.[15] An alternative interpretation is that participants' ethical judgements in the disinterested observer condition are guided by their personal valuation of social insurance, which explains the similarities between choices in Parts 1 and 2 (the disinterested observer condition and under the veil of ignorance). For example, a participant in the role of a disinterested observer might ask himself what he would want others to choose for him, were he the affected party. A more confident subject might answer that he would wish for his achievement to stand, and not be rescinded by redistribution. In this way, confidence and risk aversion could affect even a disinterested choice.[16]

These findings contribute to the empirical debate about the sources of gender differences in voting patterns and redistributive attitudes in survey responses. Our results suggest that women and men think differently about performance as a justification for inequalities, and that aversion to uncertainty and lower (over)confidence about future earnings are important reasons for women to choose higher tax rates. These results complement structural explanations of the shifts in voting patterns, with women voting increasingly left wing in Europe and the U.S. compared to 40 years ago. Specifically, women have faced increased income uncertainty due to rising divorce rates, decreasing rates of (early) marriage and increased labor market participation (Edlund and Pande, 2002). Our results suggest that this increased uncertainty may drive women to vote for parties that advocate higher redistribution.

Our study links the experimental literature about gender differences in risk aversion and overconfidence (see Section 2), to the literature on redistribution. Specifically, we show that in redistribution decisions women are more likely to choose the "safe" option of redistribution, like others have found in more abstract gambles. This occurs both because they dislike risk per se, and because they are less likely to overestimate their actual income levels. While we view any direct extrapolations of our data to a political context as conjectures that require testing by other methods, our paper illustrates how experimental measures can generate insights into the motivations underlying redistributive choices.

Acknowledgments

We are deeply indebted to Francesco Farina, Gianluca Grimalda and Ulrich Schmidt for generously permitting us to extend our study of gender differences to the data from their study sites. While the patterns we report are present in the DPW data, which shares two co-authors with the present paper, the fact that we can show international robustness of our results is due to the painstaking

[15]In DPW, income was at stake in all parts, as decision-makers in many treatments did have to sacrifice \$0.25, \$0.50 or \$1.00 of own income to redistribute each 10% of income among the others in all parts of the experiment. However, even in DPW, earners in the highest ranks faced much larger costs to redistribute in Part 3 than in Part 1.

[16]Note that we can rule out confusion among participants about the payoff schemes as an explanation. Participants were not aware of the payoffs in the second part of the experiment when the first part was played.

work and kind collegiality of FGS. Thomas Buser gratefully acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a personal Veni grant. We thank seminar audiences at the University of Amsterdam and Marburg University for helpful comments.

References

Ackert, Lucy F., J Martinez-Vazquez, and Mark Rider, "Social preferences and tax policy design: some experimental evidence," *Economic Inquiry*, 2007, *45*, 487–501.

Alesina, Alberto and Eliana La Ferrara, "Preferences for redistribution in the land of opportunities," *Journal of public Economics*, 2005, *89* (5), 897–931.

_ **and Paola Giuliano**, "Preferences for Redistribution," in Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds., *Handbook of Social Economics, Vol. 1A*, 2010, pp. 93–131.

Andreoni, James and Lise Vesterlund, "Which is the fair sex? Gender differences in altruism," *Quarterly Journal of Economics*, 2001, pp. 293–312.

Beyer, Sylvia and Edward M. Bowden, "Gender differences in seff-perceptions: Convergent evidence from three measures of accuracy and bias," *Personality and Social Psychology Bulletin*, 1997, *23* (2), 157–172.

Buser, Thomas, Leonie Gerhards, and Joël van der Weele, "Measuring Responsiveness to Feedback as a Personal Trait," *Tinbergen Institute Discussion Paper 16-043/I*, 2016.

Byrnes, James P., David C. Miller, and William D. Schafer, "Gender Differences in Risk Taking: a Meta-Analysis," *Psychological Bulletin*, 1999, *125*, 367–383.

Cappelen, Alexander W., Knut Nygaard, Erik Ø. Sørensen, and Bertil Tungodden, "Social preferences in the lab : A comparison of students and a representative population," *Scandinavian Journal of Economics*, 2015, *117* (4), 1306–1326.

_ **, Sebastian Fest, Erik Ø Sørensen, and Bertil Tungodden**, "The freedom to choose undermines the willingness to redistribute," *NHH Discussion Paper 27/2014.*, 2014, pp. 1–17.

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

Clarke, Harold D., Marianne C. Stewart, Mike Ault, and Euel Elliott, "Men, Women and the Dynamics of Presidential Approval," *British Journal of Political Science*, jan 2005, *35* (1), 31–51.

Croson, Rachel and Uri Gneezy, "Gender differences in preferences," *Journal of Economic Literature*, may 2009, *47* (2), 448–474.

Deaux, Kay and Elizabeth Farris, "Attributing causes for one's own performance: The effects of sex, norms, and outcome," *Journal of research in Personality*, 1977, *11* (1), 59–72.

Durante, Ruben, Louis Putterman, and Joël van der Weele, "Preferences for Redistribution and Perception of Fairness: an Experimental Study," *Journal of the European Economic Association*, aug 2014, *12* (4), 1059–1086.

Eagly, Alice H., Amanda B. Diekman, Mary C. Johannesen-Schmidt, and Anne M. Koenig, "Gender gaps in sociopolitical attitudes: a social psychological analysis.," *Journal of personality and social psychology*, dec 2004, *87* (6), 796–816.

Eckel, Catherine C. and Philip J. Grossman, "Differences in the Economic Decisions of Men and Women: Experimental Evidence," in "Handbook of Experimental Economics Results," Vol. 1, Elsevier, 2008, pp. 509–519.

_ and _ , "Men, women and risk aversion: Experimental evidence," in "Handbook of Experimental Economics Results," Vol. 1, Elsevier, 2008, pp. 1016–1031.

Edlund, Lena and Rohini Pande, "Why have women become left-wing? The political gender gap and the decline in marriage," *The Quarterly Journal of Economics*, 2002, (August), 917–961.

Engel, Christoph, "Dictator Games: A Meta Study," *Experimental Economics*, 2011, *14*, 583–610.

Estes, Ralph and Jinoos Hosseini, "The gender gap on Wall Street: an empirical analysis of confidence in investment decision making," *The journal of psychology*, 1988, *122* (6), 577–590.

Farina, Francesco, Gianluca Grimalda, and Ulrich Schmidt, "A cross-country experimental comparison of preferences for redistribution," *unpublished manuscript, Institute for the World Economy, Kiel University*, 2015.

Filippin, Antonio and Paolo Crosetto, "A reconsideration of gender differences in risk attitudes," *Management Science*, 2016, *In press.*

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv, "The distributional preferences of Americans," Technical Report, National Bureau of Economic Research 2014.

Funk, Patricia and Christina Gathmann, "Gender Gaps in Policy Making: Evidence from Direct Democracy in Switzerland," *Economic Policy*, 2015, *30* (81), 141–181.

Giger, Nathalie, "Towards a modern gender gap in Europe?," *The Social Science Journal*, sep 2009, *46* (3), 474–492.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," *NBER Working Paper No. 21517*, 2015.

Gneezy, Uri and Jan Potters, "An experiment on risk taking and evaluation periods," *The Quarterly Journal of Economics*, 1997, pp. 631–645.

Harrison, Glenn W. and Elisabet Rutström, "Risk Aversion in the Laboratory," in J.C. Cox and G.W. Harrison, eds., *Risk Aversion in Experiments*, Vol. 12, Bingley, UK: Emerald, Research in Experimental Economics, 2008.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, 2002, *92* (5), 1644–1655.

Howell, Susan E. and Christine L. Day, "Complexities of the gender gap," *Journal of Politics*, 2000, *62* (3), 858–874.

Inglehart, R. and P. Norris, "The Developmental Theory of the Gender Gap: Women's and Men's Voting Behavior in Global Perspective," *International Political Science Review*, oct 2000, *21* (4), 441–463.

Iversen, Torben and Frances Rosenbluth, "The Political Economy of Gender: Explaining Cross-National Variation in the Gender Division of Labor and the Gender Voting Gap," *American Journal of Political Science*, jan 2006, *50* (1), 1–19.

Lundeberg, Mary A, Paul W Fox, and Judith Punćcohaŕ, "Highly confident but wrong: gender differences and similarities in confidence judgments.," *Journal of educational psychology*, 1994, *86* (1), 114.

Luttmer, Erzo F. and Monica Singhal, "Culture, Context, and the Taste for Redistribution," *American Economic Journal: Economic Policy*, 2011, *3* (February), 157–179.

Manza, Jeff and Clem Brooks, "The Gender Gap in U.S. Presidential Elections: When? Why? Implications?," *American Journal of Sociology*, mar 1998, *103* (5), 1235–1266.

Moore, Don A. and Paul J. Healy, "The trouble with overconfidence.," *Psychological review*, 2008, *115* (2), 502.

Niederle, Muriel and Lise Vesterlund, "Do women shy away from competition? Do men compete too much?," *The Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

Schildberg-Hörisch, Hannah, "Is the veil of ignorance only a concept about risk? An experiment," *Journal of Public Economics*, 2010, *94*, 1062–1066.

Shapiro, Robert Y. and Harpreet Mahajan, "Gender Differences in Policy Preferences : A Summary of Trends From the 1960s to the 1980s," *Public Opinion Quarterly*, 1986, *50* (1), 42–61.

Smith, Herbert L. and Brian Powell, "Great Expectations: Variations in Income Expectations Among College Seniors," *Sociology of Education*, 1990, *63* (3), 194–207.

Appendix 1: Robustness checks

Unweighted regressions

Table A1: Gender differences in tax rates (Part 1: disinterested dictator)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	6.433***	1.868	4.923*	9.891***
	(1.810)	(2.744)	(2.841)	(2.038)
p-val vs. (2)			0.310	0.005
Observations	2565	644	628	1293
R^2	0.030	0.054	0.025	0.037

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table A2: Gender differences in tax rates (Part 2: behind the veil of ignorance)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	6.749***	-0.0658	4.521	11.490***
	(1.700)	(2.701)	(2.890)	(1.978)
p-val vs. (2)			0.175	0.000
Observations	2565	644	628	1293
R^2	0.019	0.034	0.009	0.042

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table A4: Gender differences in tax rates (Part 3: known rank)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	-1.681	-4.166	2.077	-2.108
	(1.763)	(3.069)	(3.297)	(2.329)
p-val vs. (2)			0.139	0.703
Observations	1813	456	440	917
R^2	0.372	0.413	0.330	0.388

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table A3: The effect of confidence and risk aversion on the gender difference in Part 2 (unweighted regressions)

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	-0.104	-1.177	3.957	2.559	1.617	0.402	10.260***	8.986***	5.041**	4.232**
	(2.707)	(2.724)	(2.826)	(2.855)	(2.562)	(2.575)	(1.979)	(1.985)	(2.073)	(2.069)
Actual Rank	-0.123	-0.122	-1.251***	-1.179***	-3.720***	-3.639***	-0.839***	-0.809***	-2.736***	-2.643***
	(0.224)	(0.224)	(0.235)	(0.235)	(0.280)	(0.283)	(0.148)	(0.148)	(0.272)	(0.275)
Risk aversion		0.707**		1.014***		0.893***		0.909***		0.704***
		(0.338)		(0.339)		(0.317)		(0.246)		(0.248)
Overconfidence					-3.316***	-3.293***			-2.211***	-2.131***
					(0.285)	(0.284)			(0.263)	(0.264)
p-value (Δ Female)		0.054		0.019	0.069	0.009		0.001	0.000	0.000
Observations	644	644	628	628	628	628	1293	1292	1293	1292
R^2	0.034	0.042	0.053	0.065	0.243	0.252	0.067	0.082	0.138	0.147

Note: All regressions control for site fixed effects. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

US vs. Europe

Table A5: Gender differences in tax rates (Part 1: disinterested dictator)

US:

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	9.638***	4.205	7.465**	13.360***
	(2.426)	(3.373)	(3.584)	(2.574)
p-val vs. (2)			0.297	0.002
Observations	1726	434	419	873
R^2	0.027	0.030	0.011	0.050

Europe:

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	-0.135	-2.938	-0.158	2.685
	(2.343)	(4.710)	(4.600)	(3.221)
p-val vs. (2)			0.697	0.364
Observations	839	210	209	420
R^2	0.022	0.056	0.036	0.014

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A6: Gender differences in tax rates (Part 2: behind the veil of ignorance)

US:

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	9.675***	3.102	5.365	14.900***
	(2.242)	(3.281)	(3.618)	(2.490)
p-val vs. (2)			0.490	0.000
Observations	1726	434	419	873
R^2	0.030	0.039	0.006	0.063

Europe:

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	0.755	-6.581	2.833	4.401
	(2.353)	(4.739)	(4.786)	(3.151)
p-val vs. (2)			0.182	0.066
Observations	839	210	209	420
R^2	0.009	0.033	0.017	0.018

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A7: The effect of confidence and risk aversion on the gender difference in Part 2 (US)

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	2.954	1.857	4.681	3.258	2.148	0.695	13.24***	11.62***	8.306***	7.204***
	(3.295)	(3.340)	(3.418)	(3.453)	(3.005)	(3.052)	(2.526)	(2.539)	(2.654)	(2.646)
Actual Rank	-0.259	-0.279	-2.001***	-1.961***	-4.341***	-4.301***	-0.805***	-0.789***	-2.848***	-2.761***
	(0.273)	(0.274)	(0.281)	(0.281)	(0.283)	(0.284)	(0.185)	(0.185)	(0.342)	(0.345)
Risk aversion		0.584		0.806**		0.823**		0.972***		0.760**
		(0.421)		(0.390)		(0.358)		(0.319)		(0.323)
Overconfidence					-3.758***	-3.761***			-2.356***	-2.272***
					(0.348)	(0.345)			(0.329)	(0.330)
p-value (Δ Female)		0.196		0.080	0.141	0.033		0.003	0.000	0.000
Observations	434	434	419	419	419	419	873	872	873	872
R^2	0.041	0.046	0.113	0.121	0.335	0.343	0.085	0.102	0.164	0.174

Note: All regressions control for site fixed effects. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Table A8: The effect of confidence and risk aversion on the gender difference in Part 2 (Europe)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Random	Random	Origin	Origin	Origin	Origin	Perform.	Perform.	Perform.	Perform.
Female	-6.625	-7.402	3.005	1.977	1.298	0.600	4.225	3.552	-1.797	-2.083
	(4.744)	(4.730)	(4.792)	(4.760)	(4.657)	(4.602)	(3.113)	(3.137)	(3.178)	(3.202)
Actual Rank	0.191	0.257	0.258	0.459	-1.732**	-1.405**	-0.833***	-0.789***	-2.529***	-2.443***
	(0.385)	(0.388)	(0.398)	(0.390)	(0.685)	(0.689)	(0.247)	(0.248)	(0.438)	(0.442)
Risk aversion		0.933*		1.734***		1.414**		0.731**		0.523
		(0.563)		(0.635)		(0.640)		(0.363)		(0.353)
Overconfidence					-2.008***	-1.843***			-2.037***	-1.972***
					(0.549)	(0.546)			(0.422)	(0.423)
p-value (Δ Female)		0.285		0.314	0.218	0.127		0.130	0.000	0.000
Observations	210	210	209	209	209	209	420	420	420	420
R^2	0.034	0.048	0.019	0.055	0.092	0.116	0.043	0.052	0.106	0.111

Note: All regressions control for site fixed effects. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Tobit regressions

Table A9: Gender differences in tax rates (Part 1: disinterested dictator)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	8.018***	2.304	5.172	12.820***
	(2.488)	(4.408)	(4.294)	(2.922)
p-val vs. (2)			0.466	0.005
Observations	2565	644	628	1293

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table A10: Gender differences in tax rates (Part 2: behind the veil of ignorance)

	(1) All	(2) Random	(3) Origin	(4) Performance
Female	8.725***	-1.098	5.856	15.19***
	(2.373)	(4.243)	(4.523)	(2.846)
p-val vs. (2)			0.277	0.000
Observations	2565	644	628	1293

Note: All regressions control for site fixed effects. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Table A11: The effect of confidence and risk aversion on the gender difference in Part 2 (Tobit)

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	-1.101	-2.621	5.273	3.251	1.837	0.223	13.87***	12.29***	6.291**	5.352*
	(4.251)	(4.256)	(4.508)	(4.534)	(4.196)	(4.184)	(2.813)	(2.822)	(2.863)	(2.869)
Actual Rank	-0.0105	0.0156	-1.334***	-1.193***	-5.110***	-4.937***	-1.192***	-1.138***	-3.775***	-3.639***
	(0.346)	(0.345)	(0.375)	(0.369)	(0.531)	(0.530)	(0.222)	(0.222)	(0.419)	(0.422)
Risk aversion		1.183**		1.736***		1.402***		1.237***		0.950***
		(0.555)		(0.568)		(0.539)		(0.338)		(0.334)
Overconfidence					-4.620***	-4.548***			-3.020***	-2.909***
					(0.496)	(0.492)			(0.396)	(0.396)
p-value (Δ Female)		0.056		0.024	0.081	0.015		0.002	0.000	0.000
Observations	644	644	628	628	628	628	1293	1292	1293	1292

Note: All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Controlling for individual cost of redistribution

Table A12: The effect of confidence and risk aversion on the gender difference in Part 2

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	-1.864	-2.890	3.487	2.078	1.150	0.00837	8.705***	7.590***	3.053	2.440
	(2.885)	(2.903)	(2.975)	(3.006)	(2.788)	(2.794)	(2.008)	(2.026)	(2.060)	(2.071)
Actual cost	-1.374	-1.380	-4.641***	-4.790***	-0.963	-1.116	-0.0732	-0.214	0.719	0.586
	(1.237)	(1.236)	(1.181)	(1.194)	(0.984)	(1.010)	(0.886)	(0.863)	(0.821)	(0.811)
Actual Rank	0.216	0.232	0.0327	0.160	-1.844***	-1.687***	-0.834***	-0.771***	-1.706***	-1.637***
	(0.358)	(0.357)	(0.365)	(0.362)	(0.531)	(0.530)	(0.234)	(0.233)	(0.400)	(0.401)
Risk aversion		0.769**		1.194***		0.996***		0.871***		0.622**
		(0.355)		(0.366)		(0.350)		(0.244)		(0.242)
Expected cost					-5.692***	-5.719***			-6.320***	-6.071***
					(1.053)	(1.071)			(0.957)	(0.947)
Overconfidence					-1.854***	-1.799***			-1.031***	-1.003***
					(0.457)	(0.453)			(0.372)	(0.370)
p-value (Δ Female)		0.049		0.018	0.053	0.007		0.002	0.000	0.000
Observations	644	644	628	628	628	628	1293	1292	1293	1292
R^2	0.034	0.043	0.048	0.066	0.217	0.229	0.059	0.073	0.140	0.147

Note: Actual cost means the net cost to an individual of raising the tax rate by 10 percent given rank, tax cost and efficiency loss. Expected cost means the the net cost to an individual of raising the tax rate by 10 percent given their expected rank. All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Non-parametric controls

Table A13: The effect of confidence and risk aversion on the gender difference in Part 2

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	-2.318	-2.390	3.290	2.355	1.027	-0.160	8.941***	7.748***	3.382	2.722
	(2.927)	(3.004)	(3.008)	(3.089)	(2.757)	(2.808)	(2.004)	(2.048)	(2.088)	(2.123)
Rank dummies	√	√	√	√	√	√	√	√	√	√
Risk dummies		√		√		√		√		√
Expected rank dummies					√	√			√	√
p-value (Δ Female)		0.902		0.142	0.200	0.054		0.007	0.000	0.000
Observations	644	644	628	628	628	628	1293	1292	1293	1292
R^2	0.074	0.099	0.060	0.092	0.275	0.295	0.072	0.094	0.181	0.196

Note: All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

Controlling for data source (DPW vs FGS)

Table A14: The effect of confidence and risk aversion on the gender difference in Part 2

	(1) Random	(2) Random	(3) Origin	(4) Origin	(5) Origin	(6) Origin	(7) Perform.	(8) Perform.	(9) Perform.	(10) Perform.
Female	-1.737	-2.680	3.043	1.828	1.529	0.549	8.672***	7.557***	3.073	2.425
	(2.917)	(2.943)	(2.961)	(2.975)	(2.796)	(2.804)	(2.011)	(2.034)	(2.085)	(2.091)
Source	8.987	9.693	29.62***	33.12***	32.47***	36.78***	5.839	3.628	20.88***	19.23**
	(5.874)	(6.601)	(5.797)	(6.670)	(8.565)	(9.442)	(3.963)	(4.704)	(6.915)	(7.872)
Actual rank	0.149	0.194	0.243	0.428	-1.779***	-1.492**	-0.758***	-0.720***	-2.205***	-2.136***
	(0.331)	(0.332)	(0.342)	(0.334)	(0.602)	(0.602)	(0.213)	(0.214)	(0.363)	(0.366)
Actual rank x source	-0.538	-0.602	-2.785***	-2.950***	-2.768***	-3.030***	-0.238	-0.250	-1.471***	-1.437***
	(0.460)	(0.462)	(0.464)	(0.458)	(0.668)	(0.670)	(0.298)	(0.300)	(0.496)	(0.509)
Risk aversion		0.876*		1.597***		1.350***		0.749**		0.613**
		(0.495)		(0.514)		(0.516)		(0.312)		(0.304)
Risk aversion x source		-0.304		-0.900		-0.729		0.229		0.0460
		(0.698)		(0.686)		(0.661)		(0.492)		(0.493)
Overconfidence					-1.996***	-1.867***			-1.746***	-1.702***
					(0.474)	(0.470)			(0.347)	(0.347)
Overconfidence x source					-2.055***	-2.171***			-1.288***	-1.234**
					(0.628)	(0.625)			(0.479)	(0.487)
p-value (Δ Female)		0.099		0.086	0.158	0.042		0.002	0.000	0.000
Observations	644	644	628	628	628	628	1293	1292	1293	1292
R^2	0.005	0.014	0.077	0.098	0.216	0.231	0.051	0.063	0.128	0.135

Note: "Source" is binary indicator of data origin (DPW vs GFS). All regressions control for site fixed effects. Observations from Europe are overweighted such that Europe and US are given equal weight. The p-values for the change in the female coefficient are bootstrapped (10,000 repetitions, stratified by gender and site). The p-value in column 2 refers to the change in the female coefficient relative to column 1, the p-values in columns 4-6 are relative to column 3, and the p-values in columns 8-10 are relative to column 7. Clustered standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01

Appendix 2: Scaling of variables

Risk aversion

In DPW, risk attitudes were measured in an incentivized way using the "multiple price list" method introduced by Harrison and Rutström (2008). This consists of five choices between a certain payment of $1 and a lottery that pays zero or a positive amount x (x=1.8, 2, 2.33, 2.67, 3 for choices 1-5) with equal probability. Our risk aversion control is equal to the sum of the numbers of the choices where a subject picked the safe option, yielding a variable that ranges from 0 to 15 (e.g. if a subject picks the safe option for choices 1-3 and the risky option for choices 4-5, risk aversion is equal to 1+2+3=6).

In FGS, risk aversion was measured through three choices between a certain amount and a lottery. The lottery always pays 0 or 5 with equal probability while the certain amount is equal to 2.5 for choice 1, 2.1 for choice 2 and 1.7 for choice 3. Our risk aversion control is again equal to the sum of the numbers of the choices where a subject picked the safe option, yielding a variable that ranges from 0 to 6. To make the two measures comparable, in the FGS data we multiply the measure with 15/6. This means that in both cases, subjects with the highest risk aversion score of 15 rejected a gamble that in expectation pays roughly 1.5 times the certain amount.

Attitudes

In DPW, attitudes towards redistribution are measured by the question "In general, do you think there is too little income redistribution". Answers are on a scale from 1 (disagree strongly) to 7 (agree strongly). In FGS, it is based on the question "Do you think gov't should reduce income differences?". Answers are on a scale from 1 (disagree strongly) to 5 (agree strongly). For the FGS data, we rescale the variable by multiplying it with 7/5. In DPW, political philosophy is measured by the question "Which of the following best describes your political philosophy (ideology)?" Answers are on a scale from 1 (very conservative) to 7 (very liberal). In FGS, it is based on the question "In political issues people often refer to positions of 'left' and 'right'. Where would you locate your opinions in the following scale, where 1 means 'Left' and 10 means 'Right'?". For the FGS data, we rescale the variable by multiplying it with 7/10.