Gender and economic preferences in a large random sample*

Anne Boschini[a], Anna Dreber[b], Emma von Essen[c], Astri Muren[d] and Eva Ranehill[e]

September 4, 2014

Abstract:

We explore gender differences in preferences related to altruism, fairness, cooperation, trust, coordination, risk and competitiveness in an experiment with a large random sample of the Swedish population. In addition to a baseline treatment, we have treatments where participants are primed with their gender or know the counterpart's gender. We find no behavioral differences between treatments, but some gender differences within specific treatments: men are in some instances less generous, more trusting, and more competitive than women. Aside from a lack of gender differences in risk taking, our results are roughly in line with previous literature.

JEL: C91; C93; J16
Keywords: Gender differences; Random sample; Social preferences; Risk-taking; Competitiveness; Dictator games; Priming; Experiment

---

* We thank Thomas Buser, Karin Hederos Eriksson, Magnus Johannesson, Matthias Heinz, Johanna Möllerström, Mats Persson, Anna Sandberg, Heiner Schumacher and Roberto Weber, as well as seminar participants at Lund University, the Norwegian School of Economics, Stockholm School of Economics, Stockholm University, University of Gothenburg, and University of Neuchatel, for helpful comments and suggestions. We thank FORTE, the Jan Wallander and Tom Hedelius Foundation (Handelsbankens forskningsstiftelser) and the Knut and Alice Wallenberg Foundation for generous financial support.
[a] Department of Economics, Linköping University, anne.boschini@liu.se
[b] Department of Economics, Stockholm School of Economics, anna.dreber@hhs.se
[c] Department of Economics and Business, Aarhus University, vonessen@econ.au.dk
[d] Department of Economics, Stockholm University, astri.muren@ne.su.se
[e] Department of Economics, University of Zurich, eva.ranehill@econ.uzh.ch

1. Introduction

Gender differences in economic outcomes have in the academic literature often been explained by discrimination or gender differences in acquired skills. As women's educational attainment and labor market opportunities have improved, persisting gaps have caused researchers to propose additional explanations. One such proposition is that men and women differ with regard to their preferences and psychological attributes, and that this, in turn, causes men and women, on average, to select into different incentive structures and professions. In this paper we study whether there are gender differences in economic preferences in an experiment on a large random sample of the Swedish adult population. To measure preferences we use a battery of standard economic games, as well as measures of risk preferences and competitiveness in two tasks with different gender stereotypes. We also systematically vary the salience of gender through several different treatments.

Results from previous experimental studies suggest that a gender gap in preferences exists in some relevant areas: women are typically found to be more generous in the dictator game as well as less risk-taking, less competitive and less trusting than men (Eckel and Grossman 2008a; Eckel and Grossman 2008b; Croson and Gneezy 2009; Bertrand 2011; Engel 2011). These preferences have been related to important economic outcomes. For example, individuals who are risk averse tend to work in sectors with lower but less variable wages, and participate less in financial markets (e.g. Bonin et al. 2007; Fisher 2010; Dohmen et al. 2011). Research also indicates that individuals who self-report being more competitive and less altruistic earn more than less competitive and more altruistic individuals (Fortin 2008). Similarly, an increasing number of studies find that competitiveness correlates with educational choices, where individuals who are willing to compete in a mathematical task in the laboratory pursue more competitive, and more prestiguous, educational tracks (Buser, Niederle, and Oosterbeek,

forthcoming; Zhang 2013). The gender differences in preferences found in laboratory studies thus resonate with the gender gaps observed in aggregate economic outcomes, highlighting the need for more research on this topic.

It is, however, not clear to what extent experimental results on gender differences in preferences are generalizable to the general population. The vast majority of studies investigating gender differences in preferences do so using convienience samples, often relying on subject pools consisting of students (an approach that has been discussed in the literature, see, for example, Levitt and List (2007), and Falk and Heckman (2009)). Whether results from these samples can be generalized to other populations such as the inhabitants of a country, and thereby to general aggregate gender gaps at the population level, is still an open question (see, for example, Cappelen et al. 2011 who find a significant gender gap in altruism and trust among students, but not in a probabilistic sample from a broader population). Students are potentially a problematic subject pool when looking at gender differences in preferences since selection into universities may create a confound with gender and the dependent variables. Students typically cover a narrow age span whereas it could be the case that gender differences in preferences vary during the life cycle (as shown in e.g., Flory et al. 2013). Also, individuals from the higher end of the socioeconomic strata are over-represented among university students, and economic background may be correlated with gender stereotypes (Firestone, Harris, and Lambert 1999, e.g.).

In order to make generalizations about gender differences in preferences possible, it is necessary to study a broader sample than students. We study a simple random sample (of the Swedish adult population), representing the gold standard of sampling methods. Our sample is similar by gender, age, income and education to the Swedish population. In addition, we have a high response rate, and can detect no differences between non-response and response groups by gender and age (the two comparison variables we have

access to). We therefore argue that we, in all relevant respects, have a representative sample of the Swedish population.

In one of the larger samples in the experimental economics literature, we measure preferences in a broad range of standard incentivized decisions related to altruism, fairness, cooperation, trust, coordination, risk and competitiveness. This allows us to study a variety of economic preferences within the same sample, and gives us an idea of the general importance of gender gaps in preferences in this population. In different treatments we further vary the salience of the participant's own gender as well as the gender of the counterpart. Comparing behavior *between* different treatments thereby allows us to test for context dependence. In addition, our data enables us to examine gender differences in preferences *within* treatments. Finally, while gender differences in previous experimental studies typically are studied without controlling for sociodemographic characteristics, we include age, gender income and education; the most important sociodemographic controls.

Our study is conducted via telephone interviews with about 1000 participants, and the experiment includes four treatments. First, in the *Baseline* treatment, we investigate gender differences in a setting where individuals have no information about their counterpart. This treatment comes closest to what traditionally has been the norm in experimental economics when measuring preferences in the laboratory. Second, in the *Priming* treatment, participants are in a subtle way reminded of their own gender by being asked to state their gender before they make their decisions. Priming is a tool from social psychology, and suggests that behavior can be affected by the individual's sense of identity (e.g. Steele and Aronson 1995).[1] In the third and fourth treatments, the *Female counterpart* and *Male counterpart* treatments, we vary the degree of information about the counterpart by making participants aware

[1] For previous papers in economics using a similar technique to prime participants with their gender, see Benjamin, Choi and Strickland (2010) and Boschini, Muren and Persson (2012). For research in economics on the importance of identity, see, e.g., Akerlof and Kranton (2000).

of the gender of their counterpart. Here, we investigate whether gender differences in preferences arise in the interplay between members of the same and the opposite sex. *Gender priming* as well as knowing the *gender of the counterpart* may induce both men and women to act in accordance with gender stereotypes that for example define women as non-competitive, docile and generous, and men as more aggressive and self-assertive (e.g. Bakan 1966; Williams and Best 1982; Guimond et al 2006; Myers 2009). The priming treatment and the treatments where the gender of the counterpart is known are also an attempt to mirror decision-making outside of the laboratory, where we are often reminded of our gender or know the gender of our counterpart.

We find no behavioral differences between treatments but some gender differences within specific treatments. Men are more trusting than women in the baseline treatment. When participants are primed with their own gender, there is evidence of men being more competitive in the math task and less generous in the dictator game than women. In addition, men are more willing to compete in the math task when the counterpart is female. We find no evidence of gender differences in risk taking. Finally, including sociodemographic control variables does not change any of these results qualitatively.[2]

We thus find evidence of gender differences in those preferences where it was expected, with the exception of risk preferences. However, many of the gender differences arise in specific contexts, and are not general or present in all treatments. There are several possible explanations to our findings. There may be fewer gender differences in Sweden compared to many other countries,

[2] Pooling the data gives us a very similar picture. In the pooled data set men are significantly more trusting and competitive, and marginally significantly less generous. Controlling for sociodemographic characteristics in the pooled dataset decreases these gender differences somewhat; the gender difference in the dictator game is no longer significant, and the gender difference in the trust game becomes marginally significant.

given Sweden's relatively high degree of gender equality.[3] Differences in comparison to previous research may also be due to differences in the measurement method (individual phone interviews). It could also be the case that previous published research, and research using mainly student samples, has overestimated the occurrence of gender gaps in preferences. This remains to be explored further, in particular by studying random samples in cross-country comparisons.

The remainder of the paper is organized as follows. In section 2 we discuss previous experimental results on gender differences and present our hypotheses. In section 3 we describe our experimental design and the procedure of data collection. In section 4 we present the data, while in section 5 we present our results for the four treatments, looking at each treatment separately as well as comparing the baseline treatment with the other three treatments respectively. Finally, we discuss our results in comparison with previous literature and conclude in section 6.

2. Previous literature and hypotheses

For each of the games and measures studied here, we start by reviewing the existing studies that, like us, experimentally investigate gender differences using probabilistic sampling from a more general population. For some preferences there is either little or no previous research using probabilistic samples; here we also briefly review the experimental literature using other sampling methods.[4] We first discuss studies where the interacting parties are anonymous to each other, as is the case in our *Baseline* treatment. Thereafter we consider studies using either priming, framing or varying degrees of information about the counterpart to study gender differences in preferences. These are relevant for our three remaining treatments; *Priming*, *Female*

[3] Sweden scores in the top world-wide on gender equality indices (see, e.g., "The Global Gender Gap Report 2012" 2014).

[4] For reviews on gender differences in preferences, mainly based on non-probabilistic student samples, see Eckel and Grossman (2008a), Croson and Gneezy (2009) and Bertrand (2010).

counterpart and *Male counterpart.* These latter studies have so far only used non-probabilistic samples. Overall, the results suggest that both the existence and magnitude of gender gaps in different preferences vary depending on the context. We use the findings in the literature to form our hypotheses, which we present separately following the literature review.

2.1 Literature review

2.1.1 The dictator game

In behavioral economics and psychology, altruism or generosity is often studied through giving in the dictator game (DG) (e.g. Kahneman, Knetsch, and Thaler 1986; Forsythe et al. 1994). We have identified three studies investigating dictator game giving in probabilistic samples (Carpenter, Connolly, and Myers 2008; Cappelen et al. 2011; Fisman, Jakiela, and Kariv 2014). The first two studies compare a convenience sample of students (approximately 100 participants) to a representative sample (410 participants in Carpenter et al. 2008, and 136 in Cappelen et al. 2011).[5] In both studies, a gender gap in giving is present only among students, where men give less than women. Fisman et al. (2014) use a sample of 1002 individuals from the American Life Panel and find a small effect in the direction of women giving more. Other studies on non-probabilistic samples (mainly students) also find that, if anything, women tend to give more (meta-study: Engel 2011; two reviews on student samples: Eckel and Grossman, 2008a; Croson and Gneezy, 2009).

Expectations consistent with gender norms of women being more altruistic than men have also been identified in dictator games (see Eckel and Grossman 2008a; Aguiar et al. 2009; Grossman and Lugovskyy 2011). Further, Boschini et al. (2012) show that *priming* participants with their gender, by asking them to state their gender before the study begins, induces a significant gender gap

[5] In the study by Carpenter et al. (2008), the probabilistic sampling method was random in all respects except for gender; they oversampled men due to the nature of the population of interest.

in altruism that is in line with social expectations. This happens only when female and male participants and seated in the same room. More specifically, while there is no gender difference when participants are not primed with their gender, or when they are seated in single-sex groups, men primed with their gender in mixed-sex seating become more egoistic. Since a related study (Cadsby, Servátka, and Song 2010) finds no effect of using single-sex versus mixed-sex seating on gender behavior in dictator game giving, it appears to be the combination of seating and priming rather than the seating in itself that matters for gendered behavior.

The gender of the counterpart has also been shown to matter. Several studies have found that both men and women are more generous towards women (Eckel and Grossman 1998; Saad and Gill 2001; Dufwenberg and Muren 2006).

2.1.2 The ultimatum game

Fairness preferences are, in experimental economics, often measured with the ultimatum game (UG) (e.g. Güth, Schmittberger, and Schwarze 1982). To our knowledge, the only study looking at ultimatum game behavior using a large non-student sample is Bellemare, Kröger, and Van Soest. (2008), where the participants are about 2000 households randomly selected from the Dutch population.[6] This study identifies a weak tendency among men to accept lower offers than women. On the other hand, gender differences in non-probabilistic samples are inconclusive (e.g. Solnick 2001; Eckel and Grossman 2001; Eckel and Grossman 2008a; Croson and Gneezy 2009; Güth and Kocher 2013).

Several previous papers vary the available information about the *gender of the counterpart* in the ultimatum game; the results are, however, conflicting (Eckel and Grossman 2001; Saad and Gill 2001; Solnick 2001; McGee and Constantinides 2013).

[6] The recruitment for the CentERpanel is conducted by TNS-NIPO. Households complete an internet based survey every week. When a household leaves the panel it is replaced with another household with similar characteristics.

2.1.3 The prisoner's dilemma

The prisoner's dilemma (PD) is the most studied game of cooperation. We are aware of one study that investigates gender differences in cooperation in a representative sample. Thöni et al. (2012), study cooperation in a public goods game among 1448 participants from an internet panel recruited via invitations sent to a random draw from the Danish population, finding no significant effects of gender on cooperative behavior.[7] Additionally, studies on student samples tend to find mixed results when it comes to gender and cooperation (see Croson and Gneezy 2009; Ellingsen et al. 2013).

As far as we know there are no relevant studies on priming or information about the gender of the counterpart in the prisoner's dilemma.

2.1.4 The trust game

Trust and reciprocity have also been given substantial attention in economics, and are typically measured using the investment game - often referred to as the trust game (TG) (e.g. Berg, Dickhaut, and McCabe 1995). We have identified three studies of this game that use probabilistic sampling from a wider population. Fehr et al. (2003) study trust and trustworthiness in a probabilistic sample of 429 individuals in Germany. They find no correlations between trust or trustworthiness and gender. Falk and Zehnder (2013) compare the trust behavior in a random sample of students at the University of Zurich with that of a random sample of citizens of the city of Zurich. While there are slight differences in trust between the groups, there are no indications of any gender differences. Cappelen et al. (2011) find no gender differences in a probabilistic Norwegian sample, but find men to be more trusting than women among two samples of students. In contrast to these studies, Garbarino and Slonim (2009) find men to be more trusting than women in a study with 441 participants

[7] Highly educated individuals as well as high-income earners and the middle aged individuals are somewhat overrepresented in comparison with the Danish population.

from a national internet panel in the US.[8] In a similar vein, most studies on students find that men are more trusting than women, if anything (e.g. Cappelen et al. 2011, see also Croson and Gneezy 2009).

Concerning information about the gender of the counterpart, Slonim and Gullien (2010), using a design where the gender of the counterpart is known, find that participants trust counterparts of opposite gender more when they are able to select who they interact with based on gender, compared to when the counterpart is randomly assigned.

2.1.5 The battle of the sexes

The battle of the sexes (BoS) is a coordination game where participants have conflicting interests but can earn more money if they coordinate on either participant's preferred outcome (Luce and Raiffa 1989). We are not aware of any previous study that explores behavior in this game in a broader sample than university students.

Holm (2000) studies this game in student samples in Sweden and the US, and finds that participants behave significantly more hawkish against women than against men when they know the *gender of the counterpart*. This effect is present for women as well as men, and is found in both Sweden and the US.[9]

However, little is known about gender differences in the baseline version of the game. As pointed out by Holm (2000), the two parties in this game need a focal point to coordinate since they have conflicting interests and are unable to choose an equal split. This makes the battle of the sexes particularly suited for studying whether gender saliency induces gender stereotypic behavior.

[8] Garbarino and Slonim (2009) use the *StudyResponse's* national panel that consists of individuals that have agreed to participate in academic research including research online.
[9] Holm (2000) does not include any baseline version where the gender of the counterpart is unknown.

2.1.6 Risk preferences

Risk preferences have been measured in several ways in the experimental economics literature, using both incentivized and non-incentivized measures.

Harrison et al. (2007) investigate risk preferences using an incentivized measure in a random sample of the Danish population aged 19 to 75 (253 individuals). They find that individuals are, on average, risk averse, but they find no gender difference. This result holds also when controlling for a number of socio-economic variables. Dohmen et al. (2010) and Dohmen et al. (2011) study two different random samples of adults (1012 and 22019 individuals respectively, overall response rates are not reported). Using measures both with and without incentives, Dohmen et al. (2010) do not find gender differences in risk-taking.[10] Dohmen et al. (2011), on the other hand, find that women self-report to be less willing to take risks (and this measure is highly correlated with the incentivized measure). They also find that the willingness to take risks decreases with age. Using the same hypothetical measure, Almenberg and Dreber (2012) also find that men on average self-report to be more risk taking than women in a random sample of 1,300 Swedish adults. von Gaudecker et al. (2011), using both hypothetical and incentivized measures, find Dutch women to be more risk averse than men when investigating risk preferences using the CentER internet panel.[11] Finally, Beauchamp et al. (2012) use a large probabilistic sample (approximately 11,000 individuals) drawn from the Swedish twin population, which is on several dimensions representative of the general population. Using the same hypothetical risk measure as in Dohmen et al. (2010, 2011), they find that men are significantly more risk taking on average. In sum, the results from probabilistic samples are not conclusive but indicate that, if anything, men are more risk taking than women.

[10] Out of the larger sample of participants in Dohmen et al. (2010), 500 were invited to take part in a risk elicitation task and 94 % participated. In Dohmen et al. (2011) the response rate for the general risk question was 99.67 %.

[11] The panel is the same as in the study by Bellemare et al. (2008).

When it comes to studies on non-representative samples, many find women to be more risk averse than men (see Croson and Gneezy 2009 for an overview of this literature), though the effect sizes tend to be fairly small (Nelson 2014; Filippin and Crosetto 2014).

Exploring the impact of priming, Benjamin et al. (2010) find no effect of gender priming on risk preferences. However, Booth and Nolen (2012a) find that girls who are either randomly allocated to all-female groups or who attend single-sex schools are less risk averse than girls from mixed-sex groups or girls from co-ed schools.

2.1.7 Competitiveness

Competitiveness is typically measured as either the willingness to compete, by participants self-selecting into a tournament instead of a piece-rate payment scheme, or from the performance response as a reaction to a competitive setting compared to a non-competitive setting (e.g. Gneezy, Niederle, and Rustichini 2003; Niederle and Vesterlund 2007). To our knowledge there is no previous study on competitiveness using probabilistic sampling from a country's entire population. A number of other non-probabilistic studies show that women or girls typically are less willing to compete than men or boys (e.g. Gneezy and Rustichini 2004; Niederle and Vesterlund 2007; Sutter and Rützler 2010; Almås et al. 2014).[12] The gender gap in the willingness to compete sometimes varies and can even be reversed depending on the task performed (e.g. Gneezy 2005; Grosse and Riener 2010; Günther et al. 2010; Shurchkov 2012; Dreber et al. 2014). Typically, gender differences appear when the competitive task is math related, but there are also studies finding that the type of task does not matter (e.g. Wozniak et al. 2010).

[12] The results on gender differences in performance change are more mixed, with null results as well as conflicting ones (e.g. Gneezy et al. 2003; Gneezy and Rustichini 2004; Dreber, Essen, and Ranehill 2011; Datta Gupta, Poulsen, and Villeval 2013; Dreber, Essen, and Ranehill 2014; Khachatryan et al. 2014). There is also evidence suggesting that gender differences in competitiveness may be country-specific but not always easy to explain with e.g. women's power relative to men (see e.g. Gneezy, Leonard, and List 2009; Cárdenas et al. 2012; Andersen et al. 2013; Zhang 2013).

No study, to our knowledge, explicitly investigates the effect of gender *priming* on competitiveness. Booth and Nolen (2012b) find that girls from same-sex schools are more willing to compete than girls from co-ed schools, and that when girls are randomized to either single-sex groups or mixed-sex groups, girls in the former group are more willing to compete than girls in the latter. It is not clear whether this is due to (long-run) effects from the gender of the counterpart or to some type of priming, but these interesting results suggest that context may influence competitiveness.

Several studies on competitiveness look at the effect of the *gender of the counterpart*. The results are mixed (Gneezy, Niederle, and Rustichini 2003; Antonovics, Arcidiacono, and Walsh 2009; Cárdenas et al. 2012; Datta Gupta et al 2013).

2.2 Hypotheses

We summarize the literature review above in Table 1, which provides a basis for our hypotheses.

Table 1. Previous results - basis for hypotheses

	Random population samples	*Student and other non-random samples*	*Context effects*
DG giving	M ≤ W	M ≤ W	Gender priming can increase the gender difference. People give more to female counterparts
UG least acceptance	M ≤ W	M = W	Mixed results on effects of counterpart's gender*
PD cooperation	M = W	Mixed results	-
Trust investment	M ≥ W	M ≥ W	People send more to the opposite gender when the gender of the counterpart is a choice
BoS hawkish	-	-	People play more hawkish with female counterparts
Risk taking	M ≥ W	M ≥ W	Gender differences can be reduced in single-sex settings

Competitive-ness	-	M > W	The task, single-sex settings and cultural factors sometimes matter. Mixed results on effects of the counterpart's gender*

*Mixed results indicate that gender differences in opposite directions have been found in the literature.

2.2.1 Baseline treatment

The results from the literature review above suggest that in our *Baseline* treatment, we can expect women to give at least as much as men in the dictator game, and, if anything, men to be more trusting than women in the trust game, more risk-taking than women in the risk task, and more willing to compete than women in the math task but not necessarily in the word task. For the other games, we do not expect any gender differences in the baseline treatment.

2.2.2 Treatments: Priming and knowing the gender of your counterpart

In this section we present hypotheses for our three treatments.

Priming

Gender priming, by for example having men and women answer questions related to their gender, can induce men and women to act in accordance with gender stereotypes – see e.g. Spencer et al. (1999), and McGlone and Aronson (2006) – although the mechanisms are subtle and great care has to be taken in the implementation of the experiment as discussed in Bargh et al. (1996). Social psychology research (Eagly 1987; Guimond et al. 2006) suggests the presence of stereotypes depicting women as non-competitive and generous, and men as aggressive and self-assertive. If activated by priming, these stereotypes could be expected to take effect in many of the games and measures used in the present study. For example, in the the dictator game, priming may cause women to give more, or men less, or both. Priming could also affect responder behavior in the ultimatum game, leading to higher minimum acceptable proposals for men and the opposite for women; as well

as in the battle of the sexes, leading to more hawkish behavior by men and less by women. Gender stereotypes or norms could also be expected to influence risk-taking and competitive behavior, making men more and women less inclined to take risks, and to choose competition.[13] However, based on previous literature (see 2.1.6 above), we do not have strong expectations for a gender priming effect on risk preferences.

Compared to the baseline treatment we thus expect gender priming to strengthen, or introduce, gender differences in many of the games as well as in competitiveness especially in the math task.

Female and male counterpart

The aforementioned literature on gender matching or gender pairing effects in experimental games suggests that the gender of the counterpart in some instances affects behavior. The literature does not however suggest any clear direction or unifying intuition for these effects. We hope that our study by its combination of large sample size and several treatments contributes to clarifying previous contradictory results.

We also study whether men and women are more or less sensitive to the treatments*; priming* as well as *gender of the counterpart*. It has been argued that women react more to changes in the experimental setup (e.g., Croson and Gneezy 2009, Elingsen et al. 2013). However sometimes male behavior seems to be more malleable (e.g., Boschini et al. 2012), and we have no clear prior with respect to this.

3. Experimental design

This incentivized artefactual field experiment (following the definition of Harrison and List 2004) is conducted on a simple random sample of the

[13] The results in Booth and Nolen (2012a; 2012b), who find that girls in same-sex groups are more risk taking and competitive than girls in mixed-sex groups, can be interpreted in this light if individuals in single-sex environments are less likely to be reminded of sex-specific stereotypes than individuals in mixed-sex environments.

Swedish population aged 17-83. The sampling and data collection was performed in close collaboration with a professional polling company based in Stockholm, Sweden.[14] No ethical review was needed.[15] In this section we describe the design of the study. Section 4 below describes the sampling and data.

3.1 Setup and treatments

All sampled individuals received a letter inviting them to participate in a phone interview study on economic decision-making conducted by researchers at Stockholm University. The letter stated that the study would last approximately 30 minutes and that it would entail monetary rewards including a participation fee of SEK 100.[16] Phone interviews started within a few days after the letter was sent out, and the interviewers made up to 14 attempts to reach each individual in the sample.[17] The phone interview consisted of decisions made in eight independent and different situations, plus a few demographic questions. An interviewer read the instructions to each participant.[18] Before each decision, participants were asked to answer a number of control questions pertaining to each situation as a way to both measure and improve participants' understanding of each decision (except for the part measuring risk preferences). Participants received no feedback on outcomes during the course of the experiment. The interview time of about 30 minutes is in line with the standards of Statistics Sweden.

[14] MIND Research, Djupdalsvägen 7, 192 51 Solna.

[15] According to Swedish law all research that comprises treatment of sensitive personal information such as handling of social security numbers, is subject to an ethical review by the Central or Local Ethical Review Board (Etikprövningsnämnden, EPN). An application to EPN for the present project was submitted in June, 2011. EPN stated that our project did not need to undergo ethical review since we only handle anonymized information. MIND provided us with anonymized data and stored all code keys.

[16] SEK 100 was just over USD 14 at the time of the study.

[17] All interviews and attempts to contact participants were conducted in the afternoon and evening during normal working days.

[18] In order to minimize the individual differences between the interviews we conducted a pilot were we listened in on a few interviews.

Participants in the phone interview were randomly assigned to one of four treatments; *Baseline*, *Priming*, *Female counterpart* or *Male counterpart*. In our *baseline* treatment the interaction was fully anonymous for the participants vis-à-vis each other and no reference was made to gender.[19] In the *priming* treatment, participants were asked to state their gender at the beginning of the interview. Finally in the *female counterpart* and *male counterpart* treatments the gender of the counterpart was revealed and this information was repeated for each decision involving a counterpart. In addition to the random assignment to treatment, we arranged for there to be an equal number of women and men among the interviewers. This means that we are also able to consider, in an ex post analysis, whether the interviewer's gender matters for economic decision-making in each of the four treatments.

We also had a number of individuals participating in a postal survey. The postal survey was conducted for the purpose of having real counterparts to the participants in the phone interview in three decisions where the games were asymmetric.[20] These individuals were then, as described below, matched with counterparts from the phone interview in the corresponding three decisions.

3.2 Experimental measures and demographic questions

In the dictator game, the ultimatum game and the trust game the participant had a counterpart from the postal survey as explained below.[21] In the prisoner's dilemma game, the battle of the sexes, and when we measured competitive preferences, a counterpart was selected among the other

[19] Unlike in a standard laboratory experiment, participants are however not anonymous vis-à-vis the interviewer.

[20] Individuals in the postal survey received the survey with a cover letter stating that the survey was part of a study on economic decision-making conducted by researchers at Stockholm University and that filling out the survey would entail monetary rewards. Reminders containing a copy of the questionnaire were sent out to all non-response participants three and five weeks later. These individuals also answered some demographic questions.

[21] The participants in the phone interview were informed of having a counterpart among the participants in the study, but not what kind of survey or interview this counterpart participated in. Participants in the postal survey were neither primed nor did they receive information about the counterpart's gender.

participants in the phone interview. The measure of risk preferences involved no counterpart. All decision situations were presented to the participants in neutral language and the different alternatives were labeled neutrally, for example option “A” and option “B” rather than “cooperate” and “defect”. Participants were informed that one of the decisions would be randomly chosen for payment in accordance with the decision(s) taken by the participants involved.

In the dictator game the phone interview participants were the dictators. They were allocated SEK 400 and asked to decide how much to keep and how much to allocate to the other person (who was in the postal interview). Any division of the SEK 400 between the dictator and the recipient could be chosen. We study how much the dictator chose to give.

In the ultimatum game participants in the phone interview were responders while participants in the postal survey were proposers. We are mainly interested in responder behavior since this decision is without strategic concerns, and can directly be interpreted as due to fairness preferences and/or preferences for reciprocity. Responders were informed that proposers had SEK 400 to split. They were asked to state the minimum amount that they would accept (in integers), below which they would reject all other offers. Accepting the proposal meant that both the participants are allocated the money accordingly, whereas rejecting the proposal implied that the money was forfeited for both.

In the prisoner’s dilemma both players simultaneously chose whether to cooperate or defect. We described the game as one where cooperation entails paying a cost for someone else to receive a benefit, and where the cost is lower than the benefit. Thus, the decision to cooperate is presented as giving

zero to self and SEK 300 to the counterpart, while defection implies giving SEK 100 to both players.[22]

In the trust game one participant, the trustor, was endowed with an amount of money and could send some or all to the counterpart, the trustee. In this version, any money sent was tripled before the trustee receives it. The trustee could then decide to return some money to the trustor (this transfer was not multiplied). Amounts sent and returned are measures of trust and trustworthiness, respectively. We are mainly interested in potential gender differences in trustor behavior, since trust is crucial to the decision of whether to enter into many economic interactions. Trustee behavior involves both reciprocal and altruistic preferences which we also measure in the dictator and ultimatum games. Participants in the phone interview were thus trustors, while the trustees were in the postal survey. Our trustors were endowed with SEK 100 and were asked to choose between keeping the whole sum, sending half, or sending the whole to the trustee.[23]

In the battle of the sexes, each of our participants chose between an option that gave SEK 150 to the counterpart and SEK 300 to themselves and an option that gave SEK 300 to the counterpart and SEK 150 to themselves, as long as both participants choose the same option. If they chose different options each got zero. We measure whether participants chose the hawkish option, i.e., the option that could lead to their own preferred outcome, rather than the option that could lead to the counterpart's preferred outcome.

We elicit risk preferences by letting participants make seven choices between a risky and a safe option. The risky options all have the same expected outcome and variance; they gave SEK 200 or 0 with equal probability, while the safe options were amounts varying from SEK 40 to SEK 160 (with

[22] Thus if both cooperate they get 300 each, if one defects while the other cooperate the defector gets 400 and the other 100, and if both defect they get 200 each.
[23] Trustees in the postal survey decided how much to return for the two alternatives where the trustor sent some money (either half or all of the money).

increments of 20 SEK). We measure risk preferences as the number of times the participant chose the risky option. If this decision was selected for payment, one out of the seven decisions was randomly chosen for payment. An advantage with this risk measure is that it is easy to understand for the participants.

We measure competitiveness in two different tasks: a word task and a math task. The choice of these two tasks allows us to compare the gender gap in competitiveness in a math task with an implicit male stereotype, and a language task with a neutral, or implicit female stereotype.[24] In the word task participants were asked to form as many words as possible from eight given letters (words of at least three letters) during a period of two minutes. In the math task they were asked to find as many number combinations as possible that add up to 25 from nine given numbers, also during two minutes. After having had the task described to them, participants chose payment form – individual payment or tournament payment. In our individual payment scheme, participants were paid SEK 10 per correctly solved exercise. The tournament payment scheme involved comparison with the counterpart, and only the best performer was paid (implicitly, the counterpart would then also have chosen the tournament option). Participants who chose the tournament payment scheme got SEK 20 for each correctly solved exercise if they solved at least as many exercises as the counterpart did, and zero otherwise. Competitiveness in each task is measured by the extent to which participants choose the tournament payment and are thus willing to compete.

Lastly, we asked participants a set of socio-demographic questions. In particular, we asked for age, gender, income, and education (see Table A1 in Appendix A for a detailed description of the variables). These variables are included as controls since some previous work has indicated that they may correlate with gender differences in preferences. For example, Garbarino and

[24] See, for example, Nosek et al. (2002) and Steffens et al. (2010) who investigate tasks and implicit gender stereotypes.

Slonim (2009), find complex interactions of age and gender in reciprocating trust behavior, and Carpenter et al. (2008) find (young) male students to be particularly selfish.[25]

3.3 Method of data collection

Using phone interviews to collect data may have advantages and disadvantages. The phone interview method has the advantage of being associated with high response rates while being less costly than face-to-face interviews (de Leeuw 2008). The use of computer assisted phone interviewing techniques also gives control over the interviewing process. On the other hand, phone interviews cannot be as long as face-to-face interviews, and information cannot be presented visually. For questions that have many or complicated alternatives, a possible response effect when using phone interviews is that participants choose the last alternative in a long list (so-called recency). In internet surveys respondents may instead be impatient and tend towards the first alternative. Fricker et al. (2005) compare phone and web interviews experimentally and find higher response rates and somewhat shorter times of completion in phone interviews, but no clear differences in data quality. Another aspect with phone interviews is that they may influence anonymity. Although anonymity is arguably greater in phone interviews than in face to face interviews, the phone interview setting does clearly not allow for the same degree of anonymity as the laboratory.

4. Data

The sampling and data collection was performed in September through November 2011, with a follow-up collection of income and education data in October 2012. In order to draw inference to the population of Swedish citizens aged 18-73, minimize problems with external validity and simplify

[25] Other measures we elicited were: civil status, number of children below age 18, household income, occupation, occupational sector, and the position within the workplace. Including these variables in the current analysis does not change our result in a qualitative way.

calculations, we drew a simple random sample. The drawn sample comprises 2349 individuals for the phone interview and 800 individuals for the postal survey, using the official register of the Swedish population (2011-08-19). We had over-coverage[26] of 465 individuals drawn to the phone interview and 48 individuals drawn to the postal survey, implying a sample size of 1884 individuals for the phone interview and 752 for the postal survey (see Table A2).

In total, 997 individuals completed the phone interview and 374 responded to the postal survey. This leaves us with a response rate of 52.9% for the phone interview and 49% for the postal survey.[27] To test whether the non-response was systematic we compare characteristics of the responders and non-responders based on the variables available, gender and age. In this paper we are primarily interested in analyzing the response from the phone interviews; we therefore focus on the phone sample in this comparison. Among the 881 non-responders in the phone interview sample, 610 individuals declined to participate and 271 were not reachable after 14 attempts. Statistics Sweden's regulations regarding statistical disclosure control provided us with two dimensions of the non-responders: gender and age. Average proportions of gender and age do not display statistically significant differences between the two samples (see Table A3 in Appendix A).[28] We thus do not have any indication of systematic non-response along these two dimensions.

[26] In survey sampling, overcoverage is a type of selection bias. It occurs when some non-members of the population are inadequately represented in the sample. These individuals should not have been a part of the sample due to the definition of the population. As an example, 6 individuals in the sample happened to be older than the population we sampled from; these individuals are therefore excluded from the sample. Further, to be a part of the population an active phone number or postal adress was required, which was not the case for all drawn individuals in the sample.

[27] These response rates are comparable to standard surveys conducted by Statistics Sweden (www.scb.se). Comparing with previous literature, for example 29.5 percent accepted a recruitment interview with the Dutch CentERpanel (55.3 percent of these then joined the panel, see Teppa and Vis 2012).

[28] We also compared the interaction between gender and age, and we have similar distribution among responders and non-responders.

Among the 997 phone interview participants, 269 were randomly assigned to the anonymous treatment, 256 to the priming treatment, and 472 to the treatment where the gender of the counterpart is revealed (254 interacted with men and 218 interacted with women).[29] To test that the random assignment provided us with similar groups in each treatment we compared average socio-demographic characteristics between the treatment groups. The characteristics we measured (gender, age, education and income) were all collected at the end of the survey. As expected in a randomized study we find no differences among the participants in different groups (see Table A4 in Appendix A).

To test the external validity (representativeness) of our population we compare point estimates of averages and standard deviations of the total Swedish population from Statistics Sweden aged 18-73 in 2011 with those of our sample. Since the main focus of the study is gender we compare our sample to the population in this dimension. To further understand how our sample relates to the population we also look at age, income and education. The sample respondents seem to compare well to the population with respect to gender income and education, but are somewhat older than the population on average (see Table 2 below).[30]

Table 2. Comparing population and sample

	Population		**Our sample**	
	Mean	**Sd**	**Mean**	**Sd**
Gender (0=man, 1=woman)	0.502	0.50	0.488	0.50
Age	40.616	23.678	45.516	15.758
	Proportion	**Share of women**	**Proportion**	**Share of women**
Income				

[29] For the matching of respondents within the phone interview (decisions 3, 5, 7, 8) and between the phone interview and postal survey (decisions 1, 2 and 4), we use an ex-post matching process similar to that proposed in Fehr et al (2003). As it turned out, it was possible to achieve this matching without using any of the participating individuals' decisions more than once.

[30] Since our random sample seems to be fairly representative of the population we do not consider population weights necessary.

<= 100.000	0.196	0.470	0.133	0.520
100 001-250 000	0.315	0.392	0.277	0.576
250 001-375 000	0.299	0.478	0.360	0.493
375 001-500 000	0.114	0.326	0.113	0.306
500 001-750 000	0.056	0.261	0.084	0.325
750 001-1 000 000	0.012	0.226	0.020	0.263
>1 000 000	0.008	0.165	0.013	0.333
Education				
< 7 years	0.064	0.462	0.042	0.390
8-9 years	0.119	0.433	0.110	0.422
10-12 years	0.468	0.475	0.409	0.493
> 12 years	0.350	0.550	0.440	0.508

Statistics Sweden provided estimates based on individual data, from 2011 of the Swedish population aged 18-73. The distribution of age is somewhat flatter and more skewed to the right of the mean compared to the population distribution (comparing kurtosis and skewness).

5. Results

Table 3 below gives an overview of the average outcome for each decision, in the baseline treatment.[31] In the main analysis and here, we only include participants who correctly answered all the control questions for each specific game; the sample size thus varies across decisions. Including all participants does not change our results qualitatively (see section 5.3).

Table 3. Overview, baseline treatment.

Decision	Mean	SD	%	N
Dictator game (giving SEK 0-400)	128.112	96.451	32.028	241
Ultimatum Game (least acceptance SEK 0-400)	103.367	73.444	25.842	173
Prisoner's dilemma (share cooperating)	0.513	0.501		128
Trust game (investment SEK 0-100)	61.4538	37.001	61.538	169
Battle of the sexes (share hawkish)	0.676	0.469		179
Risk taking (6 lotteries, number of risky decisions)	3.613	2.142	60.223	269
Competition Word (share competing)	0.376	0.486		197
Compeittion Math (share competing)	0.254	0.436		205

[31] We only report results on preferences stated in the phone interview since these by design were our main focus; the results for the postal survey are available upon request.

Considering average levels (and not gender) in the baseline treatment, our results are similar to those found by previous studies on probabilistic samples. For example, Cappelen et al. (2011) find that dictator game giving is between 40-42 percent, while we have 32 percent in the baseline treatment. Bellemare et al. (2008) have dictator game giving of approximately 45 percent of the endowment, and minimum share accepted in the ultimatum game of 34 percent, where we have 26 percent in the baseline ultimatum game. Fehr et al. (2003) have trust levels of approximately 45 percent, Garbarino and Slonim (2009) have 50 percent, Falk and Zehner (2013) have 66 percent; we have 62 percent in the baseline. Dohmen et al. (2010) have about 55 percent risky choices, we have 52 percent in the baseline.[32]

5.1 Gender differences in preferences within and between treatments

To synthesize the presentation of our results, each decision is analyzed using robust OLS regressions. Throughout the analysis we employ the standard significance level of 5%, while considering a significance level of 10% as "marginally significant". In order to complement statistically significant findings, we also calculate the effect size of the gender difference, by using standardized beta for each significant result.

Table 4. Gender differences within each treatment. OLS regressions with female as independent variable. provides an overview of the results on gender differences in preferences *within* each of the four treatments. The within treatment results are presented without control variables. Including the control variables does not change our results qualitatively.[33] In the text below we

[32] We here refer to studies that report comparable results, or where these can be calculated from reported figures and/or tables (in Fehr et al. 2003; and Bellemare et al. 2008 we used the data for threshold players).

[33] See table A5 in Appendix A for an extended table including the results with control variables. The control variables are age, age squared, education, income and gender of the interviewer, where income and education are categorized into three categories respectively (low, middle and high). The original variable consisted of more categories but there were too few observations in some of them for meaningful interpretation. Detailed definitions of these variables are presented in Table A1 in appendix A.

summarize, for each decision, (1) whether gender differences exist within each treatment, with and without control variables, and (2) whether treatment differences exist between baseline and all other treatments respectively. The analysis of differences *between* treatments, for men and women separately as well as together, is reported in more detail in Table A5 in appendix A. Here we only present the results from the between treatment analysis without control variables.

Table 4. Gender differences within each treatment. OLS regressions with female as independent variable.

Treatment / Decision		Baseline	Priming	Female counterpart	Male counterpart
DG giving	*Female*	6.995	**21.679***	-7.727	21.18
		(12.52)	(12.45)	(14.50)	(12.88)
	# obs.	241	228	191	229
UG least acceptance	*Female*	5.135	6.93	-20.853	-0.269
		(11.19)	(10.26)	(13.98)	(12.11)
	# obs.	173	163	122	162
PD cooperation	*Female*	-0.048	0.008	0.025	-0.013
		(0.09)	(0.09)	(0.09)	(0.09)
	# obs.	128	135	111	138
Trust investment	*Female*	**-14.536****	-2.909	-4.055	-4.377
		(5.63)	(5.25)	(5.45)	(5.48)
	# obs.	169	167	134	162
BoS hawkish	*Female*	0.048	0.013	0.117	0.017
		(0.07)	(0.07)	(0.08)	(0.07)
	# obs.	179	180	150	166
Risk taking	*Female*	-0.055	0.107	-0.174	0.13
		(0.26)	(0.29)	(0.30)	(0.29)
	# obs.	269	255	218	252
Competition word	*Female*	-0.029	0.054	0.08	0.032
		(0.07)	(0.07)	(0.08)	(0.07)
	# obs.	197	201	165	186
Competition math	*Female*	-0.095	**-0.112***	**-0.184****	-0.07
		(0.06)	(0.06)	(0.07)	(0.06)
	# obs.	205	209	170	201

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.

In the *dictator game,* we find a marginally significant gender difference in giving only in the treatment where participants are primed with their own gender. Here, men give on average 16% less (SEK 21.68 less) compared to women (p=0.083, standardized beta=0.115). This result is strengthened and becomes significant when we add sociodemographic control variables (p=0.043, standardized beta=0.137). To investigate how the outcome in the priming treatment differ from those in the other treatments, we explore between treatment differences within genders and for the full sample. We find no significant between treatment effects in giving. However, comparing the priming treatment with the baseline for each gender specifically, men in our sample give on average SEK 8 less when being primed with their gender compared to baseline, whereas women give SEK 7 more. This change in behavior is not significant although the point estimates vary in accordance with the stereotypical gendered norm which prescribes men to behave egoistically and women altruistically.

In the *trust game*, men send on average SEK 14.54 (out of 100), or 27%, more than women in the baseline treatment. This difference is significant (p=0.011, standardized beta=-0.197). The gender difference is in the same direction but smaller and not significant in the other treatments. As with the dictator game we look at treatment effects in the pooled sample of men and women, as well as for each gender separately. Again, we find no significant treatment effects. Although the gender gap in trust is smaller in the other treatments compared to our baseline, with men on average sending less money and women more compared to baseline, these differences are not significant (see table A5 in Appendix A).

Studying *risk preferences*, we find no gender differences in any of the four treatments. Looking at between treatment effects we find a marginally significant effect, indicating that participants are somewhat less risk taking when primed compared to the baseline (p=0.039, standardized beta=0.090).

Analyzing men and women separately, this pattern is only present among males (p=0.069, standardized beta=0.110) – a direction that is not in line with the hypothesized stereotypical norm. All other comparisons between treatments are insignificant.

Regarding *willingness to compete in the math task* we find a marginally significant gender difference in the priming treatment (p=0.071, standardized beta=-0.124), which does not change when control variables are included (p=0.067, standardized beta=-0.130). We also find a significant gender difference in the treatment with a female counterpart (p=0.010, standardized beta=-0.197), also including control variables (p=0.049, standardized beta=-0.158). In the priming treatment, men are approximately 10% more likely to compete compared to women. In the treatment where the opponent is a woman, this number increases to 18%. There are no between treatment effects. The gender gap in competitiveness in the math task found here is substantially smaller than what is typically found. For example, Niederle and Vesterlund (2007) find boys are 38 percentage points more likely to compete compared to girls.

In the *ultimatum game*, the *prisoner's dilemma*, the *battle of the sexes* and *willingness to compete in the word task* we find no within or between treatment gender differences. The lack of gender gap in competitiveness in the word task and the evidence of it in the math task is as hypothesized.

Finally, we use the non-parametric Kolmogorov-Smirnov test of equality of distributions to see whether gender differences arise at other parts of the distributions of all our outcome variables. This analysis does not identify any effects other than our main results.[34]

In sum, our results indicate that there are some significant gender differences among the Swedish population aged 18-73. The gender gaps arising are largely in line with previous reported results. However, differences arise in

[34] These results are available upon request.

specific treatments. Contrary to previous literature, we do not find a gender difference in risk preferences. Including sociodemographic control variables does not change any of these results qualitatively. We find almost no differences between treatments.

5.2 Further analysis

In the following section we study how the gender of the interviewer correlates with behavior, and whether gender gaps in behavior vary systematically between different age groups.

In our design the gender of the interviewer is *not* randomly assigned and we do therefore not specify a precise hypothesis on the effect of the interviewer's sex. We also have no hypotheses regarding the existence of systematic interactions of age and gender in preferences, given the mixed previous findings (if any at all). These variables are therefore investigated ex-post.

5.2.1 Gender of the interviewer

Although the effect of the gender of the experimenter, or of the interviewer, on experimental outcomes has not been studied extensively, some previous studies suggest that the presence of the opposite sex affects behavior. The gender of the interviewer may have a priming effect by reminding the participant of their own gender. Further, the presence of an interviewer may give rise to an audience effect, causing concerns for social image or adherence to gender norms and stereotypes to become more salient. Such audience effects may interact with the gender pairing of the audience and the decision-maker (Brandts and Garofalo 2012). Previous studies of gender-of-interviewer effects do find such effects, and mainly but not exclusively for sensitive/private questions (Flores-Macias and Lawson 2008). The effects on sensitive questions seem to be stronger for women than for men. Both the size and direction of effects tend to depend both on the interviewer's gender and on the question asked (e.g., Fuchs 2009), thus not giving grounds for specific hypotheses.

Although no random assignment of interviewer to participants was made when the polling company assigned the interviews to the interviewers, about 40% of the interviews were made by male interviewers and about 60% were made by female interviewers.[35]

We conduct regressions including dummies for each gender pair; female participant and female interviewer, female participant and male interviewer, male participant and male interviewer and male participant and female interviewer. Male participant and male interviewer is the baseline category. These regressions yield one effect which is significant both without and with controls included. Men are significantly more competitive in the math task if the interviewer is a woman rather than a man ($p=0.010$ without controls and $p=0.012$ with socio-demographic controls; see table A7 in Appendix A for all results, both without and with control variables.). There are also some marginally significant effects reported in table A7 in Appendix A. In sum, the gender of the interviewer does not appear to be of major importance.

5.2.2 The gender gap among younger and older individuals

With respect to age, previous research indicates that potential gender gaps in preferences can vary with age. For example, Carpenter et al. (2008) and Cappelen et al. (2011) find that student samples and representative samples exhibit different gender gaps in preferences.[36] More specifically, Flory et al. (2012) find that gender differences in competitiveness are the most pronounced for young adults and the least pronounced for older individuals. A relevant question is therefore whether the focus on university students in previous literature has over-estimated the average gender gap in the population as a whole.

[35] The share of female interviewers is similar across treatments as well as in the full sample.
[36] We only have 92 students in our sample, which reduces the power to detect a difference between students and non-students considerably. Comparing students as a group to the rest of the sample, we do not find any differences in behavior which are significant at the 5% level in specifications with, and without, controls. (These results are available on request.)

Most previous research on gender differences in economic decision-making is made on younger populations than ours, and we use the diversity of our sample to study whether gender differences vary between younger and older individuals. In this analysis we divide our sample into young and old. According to Statistics Sweden the average age when having the first child was 29 for women and 31 for men in 2011. Several gender differences seem to become more pronounced after having children – see e.g. (Angelov et al. 2013). We therefore label individuals in our sample as young if they are 30 years old or less.[37] In total, our sample comprises 216 individuals under the age of 31. Due to the small sample size, and low power, we only report the results from regressions based on the entire sample (of individuals who correctly answered the relevant control questions in each game), with dummies for each treatment. We conduct OLS regressions with and without control variables, including gender and age category, together with an interaction of these two variables.[38] The full results are available in table A8 in Appendix A.

Our results suggest that the gender gap in preferences sometimes varies with age group. However, it is difficult to generalize our results from the different games. Older men give the least of all groups in the dictator game, and significantly less than women of the corresponding age ($p=0.047$ and $p=0.095$ without and with control variables) while there is no gender gap in the younger sample. Young women accept lower offers than older women in the ultimatum game ($p=0.004$ and $p=0.001$ without and with controls), but young individuals accept lower offers in general, and there is no gender difference within any age category. In the trust game, a gender gap in trust arises between young women and young men ($p=0.040$ and $p=0.043$ without and with controls). In the battle of the sexes, younger women choose the most

[37] In addition, we also tested the cut off points 40 and 50 years. Apart from the battle of the sexes, where differences are only significant between young women and young men using the cut off at 30 years, the results follow a similar pattern independent of age cut off point.
[38] In the main analysis we control for age as a continuous variable.

hawkish of all categories, and significantly more so than younger men (p=0.007 and p=0.006 without and with controls). Similarly, in the math competition, the gender gap in competitiveness only arises in the older age group (p=0.001 and p=0.014 without and with controls). In sum, this analysis indicates that gender differences in preferences can vary with age, and that this relationship also varies depending on the preference explored.[39] It is interesting to note that in most of the cases where differences in the gender gap arise in different age groups, the differences are smaller among younger individuals, or vary across ages due to young women behaving more like the average man.

5.3 Robustness

In order to assess the stability of our results we conduct an ex-post sample size analysis, a Bonferroni correction, and perform the main analysis also including the participants who did not answer the control questions correctly. We also check whether our results may be affected by ability (or perhaps motivation), crudely measured as the number of correct answers to the control questions, and test for gender differences in the distributions of the outcome variables.

We conducted a sample size analysis for each test with null-results, within as well as between treatments. The sample size analyses show that we would need a substantial additional sample to reach significant results, ranging from above 400 up to more than 12 000 000 depending on game and treatment. (see Table A9 in Appendix A). In the cases where we have significant results, we conducted a power analysis – the power of these tests are all below 0.8 which is a commonly used standard. Hence, our tests have low sensitivity (power to correctly detect a false null-hypothesis). The outcome of the sample size and

[39] There are some additional effects of the socio-demographic control variables: in the baseline treatment, individuals with higher incomes accept lower offers in the ultimatum game and compete more in both tasks, while individuals with high education levels cooperate less in prisoner's dilemma. These effects also suggest that the telephone interview method is no more noisy than other methods, thereby supporting our design.

power analysis suggests that there are few strong gender differences in preferences in the Swedish population, and that the null-results are most likely not an issue of sample size.

We conduct a series of tests in this paper, which may increase the probability of making Type I errors. When we designed the study, we did not calculate an optimal sample size assuming a pre-specified power and size, and did, hence, not take multiple testing corrections into account. We therefore do not include the corrections in the text partly in order to avoid an increase of Type II errors (i.e. false negatives). To provide a more cautious interpretion of the results we here also look at the results using Bonferroni correction for multiple comparisons. We investigate gender differences in preferences in 8 games and 4 treatments. We further conduct comparisons between the baseline and the three other treatments within the 8 respective games for the full sample and for the two genders respectively. In total 104 tests were thus conducted, correlated as well as uncorrelated.[40] Taking this into account when adjusting alpha leaves us with a cut-off point of $p=0.05/104=4.808e^{-4}$. An adjustment of the significance level would then imply that none of our results remain significant.

Control questions were included in the phone study both to make participants think through each decision and to indicate if they had understood the decision. There were two, three or four control questions for each decision except the risk decision which had no control questions, altogether 23 control questions. The share of participants answering the control questions correctly varies over decisions, with the largest share for the dictator game (89 percent) and the smallest share for the prisoner's dilemma game (51 percent). There is no trend of fewer correct answers through the experiment. We compared, for each measure, the group that answered at least one of the control question for that measure incorrectly, with the main sample (including only individuals that answered all control questions correctly), with respect to gender, age, income

[40] This does not include the additional ex-post tests on gender of the interviewer and age.

and education (see Appendix B). We find no gender differences between the two samples for any of the measures. However, for most of the measures, participants who failed to answer at least one control question are more frequent in the lower income and educational brackets compared to the main sample. They are also on average older than the participants in the main sample.[41]

To support our results we performed the same analysis as for our main results, with the sample including those participants who did not answer the control questions correctly. This does not change our results qualitatively. The finding that men are less generous compared to women in the priming treatment of the dictator game is weakened; it is marginally significant but only when we include the control variables ($p=0.123$ and $p=0.059$, without and with control variables). The gender difference in trust within the baseline remains stable ($p=0.016$ and $p=0.011$ without and with control variables), and the result that men are more competitive than women in the math task when they know their counterpart is a woman remains, although somewhat weakened ($p=0.040$ and $p=0.075$ without and with control variables).

Answering the control questions correctly could also be thought of as a rough measure of ability or motivation. We create a variable that is the total number of correctly answered control questions (maximum 23). There is no difference between men and women in the number of correctly answered control questions ($p=0.981$), and including this variable in our regressions does not change our results. However, the number of correct answers is correlated with preferences in some treatments. For example, it reduces the minimum acceptable amount in the baseline and priming treatments of the ultimatum game, increases PD cooperation in the baseline treatment and when the counterpart is a woman, and increases trust overall. The suggested correlation

[41] In the dictator game there is no difference between the groups on any of the characteristics.

between this kind of variable and pro-social behavior should be explored further in future research.

6. Conclusion

We have tested to what extent there are gender differences in preferences in a large random sample of the Swedish population. Previous results in the experimental literature suggest that men, compared to women, are on average less generous in the dictator game, more trusting, more risk taking and more willing to compete in math tasks, though there are conflicting results when it comes to gender differences in preferences in general (see, e.g., Eckel and Grossman 2008a; 2008b, Croson and Gneezy 2009, Bertrand 2010).[42] Our results, based on one of the larger experimental economics samples on gender differences in preferences, partially support these findings among the Swedish adult population: in our study, men are to some extent less generous, more trusting, and more willing to compete in math than women. One important exception is that we find no gender gap in risk preferences.

Previous studies on probabilistic samples, exploring settings that resemble our Baseline treatment, have either found no gender gap (as in our study), or that men are less generous in the overall population in the dictator game (Carpenter et al. 2008, Cappelen et al. 2011, Fisman et al. 2014). In line with previous studies we find no significant gender difference in the ultimatum game. In the trust game, studies have either found no gender gap (Fehr et al. 2003, Bellemare et al. 2008) or men have been found to be more trusting (Garbarino and Slonim 2009). Our result is thus in line with the latter study. Looking at risk preferences, some studies find that men are more risk taking than women (Dohmen et al. 2011, von Gaudecker et al. 2011), whereas some find no gender gap (Harrison et al. 2007, Dohmen et al. 2010). In particular, in

[42] This may have several explanations, such as small sample sizes (which are rarely more than 100 participants per treatment), differences in sampling methods, differences in populations, differences in meaures, or subtle differences in the experimental procedures used in the different studies.

contrast with two previous studies on probabilistic samples of the Swedish population, we find no gender difference in risk taking. Beauchamp et al. (2013), using a large sample of the Swedish twin population, find that men are significantly more risk taking than women on average. Almenberg and Dreber (2012), also using a sample of the Swedish general population, find that men are on average more risk taking than women. However, these two studies use hypotetical risk questions (as in e.g. Dohmen et al. 2011) whereas we use an incentivized measure. To what extent this can explain the difference in results should be tested in future work.

Most of our gender differences in preferences are observed in treatments where gender is made salient. This has previously only been explored in non-probabilistic samples. We find that men are less generous than women in the dictator game when participants are primed with their own gender. We also find some evidence suggesting that men are more competitive than women in the priming treatment. Boschini et al. (2012) have previously found that among students in Sweden, this form of priming reduces male generosity. However, it has not been previously explored if priming can affect competitiveness. Although there is no treatment effect, we find evidence of men being somewhat more competitive than women in the math task in the priming treatment. We also find that men are more willing than women to compete in the math task when the counterpart is a woman. Previous results on the role of the gender of the counterpart for gender differences in competitiveness are so far inconclusive, thus to what extent our result can be generalized needs to be explored. We observe no gender differences in competitiveness in a word task, in the ultimatum game or in the prisoner's dilemma. These results are largely in line with the overall results from previous studies on students. We also find that the gender of the counterpart has no importance in the battle of the sexes, which contrasts with a previous result on students where both men and women play more hawkishly against women (Holm 2000).

Our results suggest that gendered behavioral norms for some preferences can be activated in specific gendered situations. However, our study is largely silent as to the mechanisms of our specific results. For example, there are no significant treatment effects that could suggest why there are gender differences in some treatments and not in others. Thus, our results do not specify that men or women are more or less sensitive to the experimental context. Previous results on this have been mixed (see, e.g. Croson and Gneezy 2009 and Boschini et al. 2012). This thus remains to be explored further. Moreover, our significant results should also be interpreted with some caution. For example, we have a power below 0.8 and all our significant results have an effect size of less than half a standard deviation. Nelson (2012) advocates that researchers should be more cautious when drawing inference to aggregate popualtions from laboratory samples. When studying gender differences in preferences, Nelson suggests that the standard statistical analysis should be complemented by reporting effect sizes. Interestingly, in a review of gender differences in risk preferences, Nelson finds that in 75% of the articles reviewed the size of the gender difference is small (less than half a standard deviation). Calculating effect sizes in probabilistic samples, as we have done here, facilitates comparison between populations and enables clearer inference to aggregate populations.

Our sample spans a large age range: participants are 18-73 years old. We are thus able to study to what extent there are gender differences in preferences in some age groups but not others. If anything, we find that older individuals behave more according to conventional gender stereotypes. For example, the gender gap in the dictator game and in competitiveness in the math task is only significant among older individuals, and young women behave the most hawkish of all groups in the battle of the sexes. When it comes to the interaction between age and gender for the willingness to compete, our results differ from what has been previously reported (e.g. Mayr et al. 2012, Flory et al. 2013): this thus remains to be explored further. While the gender of the

interviewer turns out to be unimportant for most decisions and treatments, we do find men to be more competitive in the math task when the interviewer is a woman rather than a man.

Here we study a random sample from the population of Sweden, an egalitarian and fairly gender-equal country. This might mean that gender differences would be more pronounced in random samples in less gender-equal countries. It is interesting to note that even in Sweden, however, gender differences in labor market outcomes are still pervasive. While the gender wage gap, corrected for sector, experience and education, is small in an international comparison, the Swedish labor market is as horizontally segregated as other European countries (see Halldén, forthcoming).

In sum, this is the first experiment on a random sample focusing on gender differences in preferences. We find that gender differences mainly appear in specific contexts. Gendered contexts are universal in both social and working life, and to the extent that they influence long-term educational and professional choices, as well as daily decisions, they may have considerable influence over the economic outcomes of men and women. It would seem a promising line of research to focus more on these contexts, including possible interventions to close the gender gaps. Using probabilistic sampling from different types of populations will be a useful method for this type of research.

References

Aguiar, Fernando, Pablo Brañas-Garza, Ramón Cobo-Reyes, Natalia Jimenez, and Luis M. Miller. 2009. "Are Women Expected to Be More Generous?" *Experimental Economics* 12 (1): 93–98. doi:10.1007/s10683-008-9199-z.

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–53. doi:10.1162/003355300554881.

Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Sorensen, and Bertil Tungodden. 2014. *Willingness to Compete: Family Matters*. SSRN Scholarly Paper ID 2382592. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2382592.

Almenberg, Johan, and Anna Dreber. 2012. *Gender, Stock Market Participation and Financial Literacy*. SSRN Scholarly Paper ID 1880909. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=1880909.

Andersen, Steffen, Seda Ertac, Uri Gneezy, John A. List, and Sandra Maximiano. 2013. "Gender, Competitiveness, and Socialization at a Young Age: Evidence From a Matrilineal and a Patriarchal Society." *Review of Economics and Statistics* 95 (4): 1438–43. doi:10.1162/REST_a_00312.

Antonovics, Kate, Peter Arcidiacono, and Randall Walsh. 2009. "The Effects of Gender Interactions in the Lab and in the Field." *Review of Economics and Statistics* 91 (1): 152–62. doi:10.1162/rest.91.1.152.

Bakan, David. 1966. *The Duality of Human Existence: Isolation and Communion in Western Man*. Beacon Press.

Bargh, John A., Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71 (2): 230–44. doi:10.1037/0022-3514.71.2.230.

Bellemare, Charles, Sabine Kröger, and Arthur Van Soest. 2008. "Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities." *Econometrica* 76 (4): 815–39. doi:10.1111/j.1468-0262.2008.00860.x.

Benjamin, Daniel J, James J Choi, and A. Joshua Strickland. 2010. "Social Identity and Preferences." *American Economic Review* 100 (4): 1913–28. doi:10.1257/aer.100.4.1913.

Berg, Joyce, John Dickhaut, and Kevin McCabe. 1995. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10 (1): 122–42. doi:10.1006/game.1995.1027.

Bertrand, Marianne. 2011. "New Perspectives on Gender" 4B: 1543–90.

Bonin, Holger, Thomas Dohmen, Armin Falk, David Huffman, and Uwe Sunde. 2007. "Cross-Sectional Earnings Risk and Occupational Sorting: The Role of Risk Attitudes." *Labour Economics*, Education

and Risk Education and Risk S.I., 14 (6): 926–37. doi:10.1016/j.labeco.2007.06.007.
Booth, Alison L., and Patrick Nolen. 2012. “Gender Differences in Risk Behaviour: Does Nurture Matter?” *Economic Journal* 122 (558). http://ideas.repec.org/a/ecj/econjl/v122y2012i558pf56-f78.html.
Booth, Alison, and Patrick Nolen. 2012. “Choosing to Compete: How Different Are Girls and Boys?” *Journal of Economic Behavior & Organization* 81 (2): 542–55. doi:10.1016/j.jebo.2011.07.018.
Boschini, Anne, Astri Muren, and Mats Persson. 2012a. “Constructing Gender Differences in the Economics Lab.” *Journal of Economic Behavior & Organization* 84 (3): 741–52.
Brandts, Jordi, and Orsola Garofalo. 2012. “Gender Pairings and Accountability Effects.” *Journal of Economic Behavior & Organization* 83 (1): 31–41.
Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek. 2012. *Gender, Competitiveness and Career Choices*. SSRN Scholarly Paper ID 2183036. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2183036.
Cadsby, C. Bram, Maroš Servátka, and Fei Song. 2010. “Gender and Generosity: Does Degree of Anonymity or Group Gender Composition Matter?” *Experimental Economics* 13 (3): 299–308. doi:10.1007/s10683-010-9242-8.
Cappelen, Alexander W., Knut Nygaard, Erik Ø Sørensen, and Bertil Tungodden. 2011. *Social Preferences in the Lab : A Comparison of Students and a Representative Population*. CESifo Working Paper / Behavioural Economics ; 3511. München : CESifo. http://www.cesifo-group.de/portal/page/portal/ifoHome/b-publ/b3publwp/_wp_abstract?p_file_id=16157950&category=; http://www.cesifo-group.de/portal/pls/portal/docs/1/1207563.PDF; http://hdl.handle.net/10419/49506.
Cárdenas, Juan-Camilo, Anna Dreber, Emma von Essen, and Eva Ranehill. 2012. “Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden.” *Journal of Economic Behavior & Organization* 83 (1): 11–23.
Carpenter, Jeffrey, Cristina Connolly, and Caitlin Knowles Myers. 2008. “Altruistic Behavior in a Representative Dictator Experiment.” *Experimental Economics* 11 (3): 282–98. doi:10.1007/s10683-007-9193-x.
Croson, Rachel, and Uri Gneezy. 2009. “Gender Differences in Preferences.” *Journal of Economic Literature* 47 (2): 448–74.
Datta Gupta, Nabanita, Anders Poulsen, and Marie Claire Villeval. 2013. “Gender Matching and Competitiveness: Experimental Evidence.” *Economic Inquiry* 51 (1): 816–35. doi:10.1111/j.1465-7295.2011.00378.x.

De Leeuw, Edith D. 2008. “Choosing the Method of Data Collection.” In *International Handbook of Survey Methodology*, edited by E. D. de, J. J. Hox, and D. A. Dillman, 113–35. New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2010. “Are Risk Aversion and Impatience Related to Cognitive Ability?” *American Economic Review* 100 (3): 1238–60. doi:10.1257/aer.100.3.1238.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences.” *Journal of the European Economic Association* 9 (3): 522–50. doi:10.1111/j.1542-4774.2011.01015.x.

Dreber, Anna, Emma Essen, and Eva Ranehill. 2014. “Gender and Competition in Adolescence: Task Matters.” *Experimental Economics* 17 (1): 154–72.

Dreber, Anna, Emma von Essen, and Eva Ranehill. 2011. “Outrunning the Gender Gap—boys and Girls Compete Equally.” *Experimental Economics* 14 (4): 567–82. doi:10.1007/s10683-011-9282-8.

Dufwenberg, Martin, and Astri Muren. 2006. “Generosity, Anonymity, Gender.” *Journal of Economic Behavior & Organization*, Journal of Economic Behavior & Organization, 61 (1): 42–49.

Eagly, Alice Hendrickson. 1987. *Sex Differences in Social Behavior: A Social-Role Interpretation*. L. Erlbaum Associates.

Eckel, Catherine C., and Philip J. Grossman. 1998. “Are Women Less Selfish Than Men?: Evidence From Dictator Experiments.” *The Economic Journal* 108 (448): 726–35. doi:10.1111/1468-0297.00311.

———. 2001. “Chivalry and Solidarity in Ultimatum Games.” *Economic Inquiry* 39 (2): 171–88.

———. 2008a. *Differences in the Economic Decisions of Men and Women: Experimental Evidence*. SSRN Scholarly Paper ID 1883696. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=1883696.

———. 2008b. *Men, Women and Risk Aversion: Experimental Evidence*. Handbook of Experimental Economics Results. Elsevier. http://ideas.repec.org/h/eee/expchp/7-113.html.

Ellingsen, Tore, Magnus Johannesson, Johanna Mollerstrom, and Sara Munkhammar. 2013. “Gender Differences in Social Framing Effects.” *Economics Letters* 118 (3): 470–72. doi:10.1016/j.econlet.2012.12.010.

Engel, Christoph. 2011. “Dictator Games: A Meta Study.” *Experimental Economics* 14 (4): 583–610.

Falk, Armin, and James J. Heckman. 2009. “Lab Experiments Are a Major Source of Knowledge in the Social Sciences.” *Science* 326 (5952): 535–38. doi:10.1126/science.1168244.

Falk, Armin, and Christian Zehnder. 2013. “A City-Wide Experiment on Trust Discrimination.” *Journal of Public Economics* 100 (April): 15–27. doi:10.1016/j.jpubeco.2013.01.005.

Fehr, Ernst, Urs Fischbacher, Jürgen Schupp, Bernhard von Rosenbladt, and Gert Georg Wagner. 2003. *A Nationwide Laboratory Examining Trust and Trustworthiness by Integrating Behavioural Experiments into Representative Surveys*. CEPR Discussion Paper 3858. C.E.P.R. Discussion Papers. http://ideas.repec.org/p/cpr/ceprdp/3858.html.

Filippin, Antonio, and Paolo Crosetto. 2014. *A Reconsideration of Gender Differences in Risk Attitudes*. Departmental Working Paper 2014-01. Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano. http://ideas.repec.org/p/mil/wpdepa/2014-01.html.

Firestone, Juanita M., Richard J. Harris, and Linda C. Lambert. 1999. “Gender Role Ideology and the Gender Based Differences in Earnings.” *Journal of Family and Economic Issues* 20 (2): 191–215. doi:10.1023/A:1022158811154.

Fisher, Patti J. 2010. “Gender Differences in Personal Saving Behaviors.” *Journal of Financial Counseling and Planning*, January. http://www.highbeam.com/doc/1P3-2080645731.html.

Fisman, Raymond, Pamela Jakiela, and Shachar Kariv. 2014. *The Distributional Preferences of Americans*. Working Paper 20145. National Bureau of Economic Research. http://www.nber.org/papers/w20145.

Flory, Jeffery, Uri Gneezy, Kenneth Leonard, and John List. 2012. *Sex, Competitiveness, and Investment in Offspring: On the Origin of Preferences*. Artefactual Field Experiments 00072. The Field Experiments Website. http://ideas.repec.org/p/feb/artefa/00072.html.

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton. 1994. “Fairness in Simple Bargaining Experiments.” *Games and Economic Behavior* 6 (3): 347–69. doi:10.1006/game.1994.1021.

Fortin, Nicole M. 2008. “The Gender Wage Gap among Young Adults in the United States.” *Journal of Human Resources* 43 (4): 884 –918. doi:10.3368/jhr.43.4.884.

Fricker, S. 2005. “An Experimental Comparison of Web and Telephone Surveys.” *Public Opinion Quarterly* 69 (3): 370–92. doi:10.1093/poq/nfi027.

Garbarino, Ellen, and Robert Slonim. 2009. “The Robustness of Trust and Reciprocity across a Heterogeneous U.S. Population.” *Journal of Economic Behavior & Organization*, Journal of Economic Behavior & Organization, 69 (3): 226–40.

Gneezy, Uri, Kenneth L Leonard, and John A List. 2009. “Gender Differences in Competition: Evidence From a Matrilineal and a Patriarchal Society.” *Econometrica* 77 (5): 1637–64. doi:10.3982/ECTA6690.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. “Performance in Competitive Environments: Gender Differences.” *The Quarterly Journal of Economics* 118 (3): 1049 –1074. doi:10.1162/00335530360698496.

Gneezy, Uri, and Aldo Rustichini. 2004. “Gender and Competition at a Young Age.” *American Economic Review* 94 (2): 377–81. doi:10.1257/0002828041301821.

Grossman, Philip J, and Oleksandr Lugovskyy. 2011. “An Experimental Test of the Persistence of Gender-based Stereotypes.” *Economic Inquiry* 49 (2): 598–611. doi:10.1111/j.1465-7295.2009.00265.x.

Guimond, Serge, Armand Chatard, Delphine Martinot, Richard J Crisp, and Sandrine Redersdorff. 2006. “Social Comparison, Self-Stereotyping, and Gender Differences in Self-Construals.” *Journal of Personality and Social Psychology* 90 (2): 221–42. doi:10.1037/0022-3514.90.2.221.

Günther, Christina, Neslihan Arslan Ekinci, Christiane Schwieren, and Martin Strobel. 2010. “Women Can’t jump?—An Experiment on Competitive Attitudes and Stereotype Threat.” *Journal of Economic Behavior & Organization* 75 (3): 395–401. doi:10.1016/j.jebo.2010.05.003.

Güth, Werner, and Martin G. Kocher. 2013. *More than Thirty Years of Ultimatum Bargaining Experiments: Motives, Variations, and a Survey of the Recent Literature*. CESifo Working Paper Series 4380. CESifo Group Munich. http://ideas.repec.org/p/ces/ceswps/_4380.html.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. “An Experimental Analysis of Ultimatum Bargaining.” *Journal of Economic Behavior & Organization* 3 (4): 367–88. doi:10.1016/0167-2681(82)90011-7.

Holm, Håkan J. 2000. “Gender-Based Focal Points.” *Games and Economic Behavior* 32 (2): 292–314. doi:10.1006/game.1998.0685.

Harrison, Glenn W., Morten I. Lau, and E. Elisabet Rutström. 2007. “Estimating Risk Attitudes in Denmark: A Field Experiment*.” *Scandinavian Journal of Economics* 109 (2): 341–68. doi:10.1111/j.1467-9442.2007.00496.x.

Harrison, Glenn W., and John A. List. 2004. *Field Experiments*. SSRN Scholarly Paper ID 698961. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=698961.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. “Fairness and the Assumptions of Economics.” *The Journal of Business* 59 (4): S285–300.

Khachatryan, Karen, Anna Dreber, Von Essen, Emma, and Eva Ranehill. 2014. *Gender and Preferences at a Young Age: Evidence from Armenia*. SSRN Scholarly Paper ID 2397117. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2397117.

Levitt, Steven D, and John A List. 2007. “What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?” *Journal of Economic Perspectives* 21 (2): 153–74. doi:10.1257/jep.21.2.153.
Luce, R. Duncan, and Howard Raiffa. 1989. *Games and Decisions: Introduction and Critical Survey*. Reprint edition. New York: Dover Publications.
McGee, Peter, and Stelios Constantinides. 2013. “Repeated Play and Gender in the Ultimatum Game.” *Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics)* 42 (C): 121–26.
McGlone, Matthew S., and Joshua Aronson. 2006. “Stereotype Threat, Identity Salience, and Spatial Reasoning.” *Journal of Applied Developmental Psychology* 27 (5): 486–93. doi:10.1016/j.appdev.2006.06.003.
Myers, Dee Dee. 2009. *Why Women Should Rule the World*. HarperCollins.
Nelson, Julie A. 2014. “Are Women Really More Risk-Averse Than Men? A Re-Analysis of the Literature Using Expanded Methods.” *Journal of Economic Surveys*, May, n/a–n/a. doi:10.1111/joes.12069.
Niederle, Muriel, and Lise Vesterlund. 2007. “Do Women Shy Away From Competition? Do Men Compete Too Much?” *The Quarterly Journal of Economics* 122 (3): 1067–1101. doi:10.1162/qjec.122.3.1067.
Niels D. Grosse, and Gerhard Riener. 2010. *Explaining Gender Differences in Competitiveness: Gender-Task Stereotypes*. Jena Economic Research Papers. Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics. http://ideas.repec.org/p/jrp/jrpwrp/2010-017.html.
Nosek, Brian A, Mahzarin R Banaji, and Anthony G Greenwald. 2002. “Math = Male, Me = Female, Therefore Math Not = Me.” *Journal of Personality and Social Psychology* 83 (1): 44–59.
Saad, Gad, and Tripat Gill. 2001. “The Effects of a Recipient’s Gender in a Modified Dictator Game.” *Applied Economics Letters* 8 (7): 463–66. doi:10.1080/13504850010005260.
Shurchkov, Olga. 2012. “Under Pressure: Gender Differences in Output Quality and Quantity Under Competition and Time Constraints.” *Journal of the European Economic Association* 10 (5): 1189–1213. doi:10.1111/j.1542-4774.2012.01084.x.
Slonim, Robert, and Pablo Guillen. 2010. “Gender Selection Discrimination: Evidence from a Trust Game.” *Journal of Economic Behavior & Organization* 76 (2): 385–405. doi:10.1016/j.jebo.2010.06.016.
Solnick, Sj. 2001. “Gender Differences in the Ultimatum Game.” *Economic Inquiry* 39 (2): 189–200. doi:10.1111/j.1465-7295.2001.tb00060.x.
Spencer, Steven J., Claude M. Steele, and Diane M. Quinn. 1999. “Stereotype Threat and Women’s Math Performance.” *Journal of Experimental Social Psychology* 35 (1): 4–28. doi:10.1006/jesp.1998.1373.
Steele, Claude M., and Joshua Aronson. 1995. “Stereotype Threat and the Intellectual Test Performance of African Americans.” *Journal of*

Personality and Social Psychology 69 (5): 797–811. doi:10.1037/0022-3514.69.5.797.

Steffens, Melanie C., Petra Jelenec, and Peter Noack. 2010. "On the Leaky Math Pipeline: Comparing Implicit Math-Gender Stereotypes and Math Withdrawal in Female and Male Children and Adolescents." *Journal of Educational Psychology* 102 (4): 947–63. doi:10.1037/a0019920.

Sutter, Matthias, and Daniela Rützler. 2010. "Gender Differences in Competition Gender Differences in Competition Emerge Early in Life." *IZA Discussion Paper* 5015 (5015).

Teppa, Federica, and Corrie Vis. 2012. *The CentERpanel and the DNB Household Survey: Methodological Aspects*. DNB Occasional Studies 1004. Netherlands Central Bank, Research Department. http://ideas.repec.org/p/dnb/dnbocs/1004.html.

"The Global Gender Gap Report 2012." 2014. *The Global Gender Gap Report 2012 | World Economic Forum*. Accessed May 25. http://www.weforum.org/reports/global-gender-gap-report-2012.

Thöni, Christian, Jean-Robert Tyran, and Erik Wengström. 2012. "Microfoundations of Social Capital." *Journal of Public Economics* 96 (7–8): 635–43. doi:10.1016/j.jpubeco.2012.04.003.

Uri Gneezy, Aldo Rustichini. 2005. "Executive versus Teachers: Gender, Competition and Self-Selection."

Von Gaudecker, Hans-Martin, Arthur van Soest, and Erik WengstrÖm. 2011. "Heterogeneity in Risky Choice Behavior in a Broad Population." *American Economic Review* 101 (2): 664–94. doi:10.1257/aer.101.2.664.

Williams, John E., and Deborah L. Best. 1982. *Measuring Sex Stereotypes: A Thirty-Nation Study*. SAGE Publications.

Zhang, Y. Jane. 2013. *Culture and the Gender Gap in Competitive Inclination: Evidence from the Communist Experiment in China*. MPRA Paper 47356. University Library of Munich, Germany. http://ideas.repec.org/p/pra/mprapa/47356.html.

Appendix A

Table A1. Definition of variables

Name of variable	Description	Values of variable
Female	Gender of participant	1 if female, 0 otherwise
Treatment	One of the four treatment group the participant was assigned to	1=baseline 2 = priming 3=female counterpart 4=male counterpart
Age*	Age of the participant in number of years	Number of years at time of interview
Income	The income in SEK of the participants, broken down in 3 categories.	Low income=0-250000 Middle income=250001-750000 High income=750001-
Education	The education in number of years the participant had received at the time of the interview, broken down in 3 categories.	Low education=0-9 Middle education=10-12 High education=<12
Gender of the interviewer	Gender of the person interviewing the participant	1 if female, 0 otherwise

* Since we only have information on birth year we have defined age as the year the study was conducted deducted by the birth year. This implies we assume all individuals are born the 1st of January and the sample will therefore include some indivduals that are 74 years old.

Table A2. Sample sizes, non-response and under-coverage

		Telephone	Postal
Sample size		2349	800
Response		996	374
Non-response			
	Declined	610	8
	No contact	271	370
Under-coverage			
	Not part of the population	89	
	No active phone number/address	320	37
	Wrong sampling	56	
	Late postal responses		11
	Older than the population	6	

Table A3. Phone interview: non-response analysis

	Non-responders		**Responders**			
	Mean	SD	Mean	SD	p-value ttest	p-value KS**
Gender	0.4915	0.50	0.4885	0.50	0.896	1
Age category*	6.5142	3.0298	6.5166	3.1839	0.987	0.901

* Age categories are; 1=18-20, 2=21-25, 3=26-30, 4=31-35, 5=36-40, 6=41-45, 7=46-50, 8=51-55, 9=56-60, 10=61-65, 11=66-70, and 12=71-73. Since the age for responders is collected as a continuous variable when comparing the figures we forced the responders actual age into the same categories as for non-responders.
** We also tested the equality of distribution between the non-responders and the responders using a Kolmogorov-Smirnov test with the combined p-value for large samples.

Table A4. Confirmed random assignment

	Baseline		**Priming**		**Female Counterpart**		**Male counterpart**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Gender	0.4647	0.500	0.484	0.501	0.523	0.501	0.488	0.501
Age	45.647	15.663	46.439	15.539	44.326	15.843	45.472	16.030
Income category*	2.758	1.115	2.923	1.331	2.809	1.294	2.898	1.343
Interaction: gender and income category	1.210	1.488	1.258	1.547	1.287	1.485	1.275	1.621
Interaction: gender and age	21.699	25.587	22.839	25.920	22.477	24.181	21.456	24.547

*In line with regulations of statistical disclosure control from Statistics Sweden we collected income in categories. Income is categorized in 7 brackets of SEK; 1=<100 000, 2=101-250k, 3=251-375k, 4=376-500k, 5=501-759k, 6=751-1000k, and 7= >1000k.
None of the variables are significantly different between the treatments using a ttest.

Table A5. Gender differences within each treatment. OLS regressions with female as independent variable.

Decision \ Treatment		Baseline		Priming		Female counterpart		Male counterpart	
		No controls	Controls	No controls	Controls	No controls	Controls	No controls	Controls
DG giving	*Female*	6.995	7.758	**21.679***	**25.775****	-7.727	-6.751	21.180	19.230
		(12.521)	(12.917)	(12.452)	(12.646)	(14.502)	(15.394)	(12.883)	(14.014)
	# obs.	241		228		191		229	
UG least acceptance	*Female*	5.135	-2.847	6.930	11.231	-20.853	-17.233	-0.269	4.968
		(11.185)	(10.612)	(10.259)	(11.146)	(13.979)	(15.750)	(12.110)	(12.946)
	# obs.	173		163		122		162	
PD cooperation	*Female*	-0.048	-0.023	0.008	0.021	0.025	0.054	-0.013	-0.009
		(0.089)	(0.098)	(0.085)	(0.084)	(0.094)	(0.103)	(0.085)	(0.087)
	# obs.	128		135		111		138	
Trust investment	*Female*	**-14.536****	**-13.328****	-2.909	-2.248	-4.055	0.344	-4.377	-3.371
		(5.628)	(5.865)	(5.252)	(5.468)	(5.451)	(5.964)	(5.477)	(6.144)
	# obs.	169		167		134		162	
BoS hawkish	*Female*	0.048	0.069	0.013	0.016	0.117	0.102	0.017	0.011
		(0.070)	(0.072)	(0.070)	(0.075)	(0.078)	(0.081)	(0.072)	(0.081)
	# obs.	179		180		150		166	
Risk taking	*Female*	-0.055	-0.029	0.107	0.457	-0.174	-0.291	0.130	0.259
		(0.263)	(0.270)	(0.288)	(0.301)	(0.303)	(0.310)	(0.285)	(0.292)
	# obs.	269		255		218		252	
Competition word	*Female*	-0.029	0.012	0.054	0.042	0.080	0.089	0.032	0.025
		(0.069)	(0.070)	(0.069)	(0.074)	(0.076)	(0.080)	(0.072)	(0.079)
	# obs.	197		201		165		186	
Competition math	*Female*	-0.095	-0.046	**-0.112***	**-0.118***	**-0.184****	**-0.148****	-0.070	-0.040
		(0.061)	(0.062)	(0.062)	(0.064)	(0.071)	(0.075)	(0.062)	(0.072)
	# obs.	205		209		170		201	

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 Number of observations include participants that answered the control questions correctly for each decision. Control variables include age, age squared, education, income and gender of interviewer.

Table A6. Between treatment effects: OLS estimates of differences in means (no control variables)

	Baseline vs Priming			Baseline vs Female counterpart			Baseline vs Male counterpart		
	Full sample	**Women**	**Men**	**Full sample**	**Women**	**Men**	**Full sample**	**Women**	**Men**
DG giving	0.177	7.616	-7.068	0.735	-2.909	4.452	-1.074	1.274	-3.454
	(8.815)	(12.976)	(11.978)	(4.768)	(6.906)	(6.637)	(2.989)	(4.397)	(4.065)
# obs.	469	225	244	432	216	216	470	227	243
UG least acceptance	-11.389	-10.200	-11995	0.023	-6.210	6.784	-0.864	-1.764	0.038
	(7.590)	(11.077)	(10.376)	(4.478)	(6.320)	(6.334)	(2.725)	(4.167)	(3.582)
# obs.	336	162	174	295	151	144	335	161	174
PD cooperation	0.061	0.089	0.033	0.032	0.050	0.014	-0.027	-0.022	-0.033
	(0.061)	(0.089)	(0.086)	(0.032)	(0.047)	(0.045)	(0.020)	(0.029)	(0.029)
# obs	263	127	136	239	116	123	266	135	131
Trust investment	-0.161	5.925	-5.702	-0.359	2.513	-2.727	0.166	1.981	-1.405
	(3.860)	(5.741)	(5.130)	(1.968)	(2.823)	(2.716)	(1.314)	(1.942)	(1.756)
# obs.	336	151	185	303	147	156	331	156	175
BoS hawkish	0.007	-0.010	0.025	-0.011	0.004	-0.030	0.004	-0.002	0.009
	(0.049)	(0.070)	(0.070)	(0.026)	(0.035)	(0.039)	(0.017)	(0.024)	(0.024)
# obs.	359	174	185	329	169	160	345	172	173
Risk taking	**-0.402****	-0.318	**-0.480***	0.020	0.015	0.075	0.012	0.020	-0.041
	(0.194)	(0.288)	(0.263)	(0.100)	(0.148)	(0.135)	(0.064)	(0.094)	(0.088)
# obs.	525	249	276	487	239	248	523	249	274
Competition word	-0.003	0.041	-0.041	0.003	0.030	-0.024	0.006	0.016	-0.005
	(0.049)	(0.071)	(0.067)	(0.026)	(0.037)	(0.036)	(0.017)	(0.024)	(0.023)
# obs.	398	189	209	362	180	182	383	190	193
Competition math	0.033	0.021	0.038	0.035	0.013	0.058	0.005	0.009	0.000
	(0.044)	(0.059)	(0.064)	(0.024)	(0.030)	(0.035)	(0.015)	(0.020)	(0.021)
# obs.	414	199	215	375	188	187	406	197	209

Standard errors in parentheses.

Table A7. OLS regressions with gender of the interviewer

Preference	Variables	Baseline		Priming		Female counterpart		Male counterpart	
Controls		**No**	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**	**Yes**
DG giving	FemaleFemale	24.156	23.777	1.893	0.755	17.172	16.532	40.513**	38.636*
		(19.647)	(20.147)	(16.362)	(17.072)	(20.189)	(21.104)	(18.744)	(19.832)
	FemaleMale	1.664	1.016	29.195*	33.409*	4.156	3.021	16.851	17.105
		(15.870)	(16.271)	(17.350)	(17.762)	(19.614)	(19.899)	(17.681)	(18.329)
	MaleFemale	8.335	9.343	-29.229*	-30.235*	38.442*	33.688	13.215	15.538
		(16.458)	(16.604)	(17.432)	(16.788)	(20.481)	(20.529)	(17.599)	(17.924)
	#obs	241	241	228	228	191	191	229	229
UG least acceptance	FemaleFemale	-6.766	-18.479	9.414	9.858	7.413	7.935	-6.331	-7.327
		(17.260)	(16.308)	(13.072)	(14.122)	(22.790)	(24.516)	(17.379)	(18.379)
	FemaleMale	3.812	-2.545	-2.021	0.504	-27.187	-26.181	9.504	15.422
		(14.506)	(13.791)	(14.301)	(14.860)	(18.516)	(18.613)	(18.407)	(17.923)
	MaleFemale	-13.194	-13.877	-13.779	-20.769	15.408	19.030	3.765	-1.018
		(14.647)	(15.647)	(15.089)	(14.619)	(20.159)	(19.184)	(15.628)	(15.252)
	#obs	173	173	163	163	122	122	162	162
PD cooperation									
	FemaleFemale	-0.052	0.017	0.050	0.088	-0.056	-0.034	0.014	0.021
		(0.135)	(0.144)	(0.112)	(0.113)	(0.134)	(0.140)	(0.123)	(0.128)
	FemaleMale	0.007	0.038	-0.030	-0.064	0.117	0.126	0.018	-0.012
		(0.124)	(0.126)	(0.129)	(0.124)	(0.127)	(0.133)	(0.129)	(0.135)
	MaleFemale	0.025	0.055	0.054	0.026	0.017	-0.019	0.028	0.002
		(0.127)	(0.127)	(0.116)	(0.122)	(0.130)	(0.142)	(0.122)	(0.129)
	#obs	128	128	135	135	111	111	138	138
Trust investment									
	FemaleFemale	-15.412**	-13.551*	-0.689	1.446	-2.485	0.287	-0.635	-0.856
		(7.608)	(8.022)	(6.848)	(7.087)	(8.261)	(8.885)	(8.124)	(8.335)
	FemaleMale	-15.888**	-14.915*	-2.475	-1.844	-2.440	1.944	4.064	3.500
		(7.892)	(8.209)	(7.852)	(7.961)	(8.142)	(8.484)	(7.595)	(7.875)
	MaleFemale	1.731	2.060	2.346	3.795	-0.495	-3.143	12.677*	11.637
		(7.486)	(7.756)	(7.133)	(7.485)	(7.763)	(7.495)	(7.240)	(7.301)
	#obs	169	169	167	167	134	134	162	162
BoS hawkish									
	FemaleFemale	0.049	0.069	-0.032	-0.035	0.231**	0.217*	-0.039	-0.040
		(0.101)	(0.105)	(0.091)	(0.099)	(0.109)	(0.112)	(0.101)	(0.102)

	FemaleMale	0.019	0.028	-0.053	-0.056	0.185*	0.180*	-0.070	-0.077
		(0.094)	(0.094)	(0.096)	(0.100)	(0.106)	(0.108)	(0.107)	(0.110)
	MaleFemale	-0.054	-0.079	-0.151	-0.146	0.165	0.171	-0.084	-0.085
		(0.102)	(0.103)	(0.098)	(0.100)	(0.116)	(0.115)	(0.100)	(0.102)
	#obs	179	179	180	180	150	150	166	166
Risk taking									
	FemaleFemale	-0.286	-0.247	-0.091	0.340	-0.778*	-0.944**	-0.065	0.053
		(0.392)	(0.395)	(0.372)	(0.377)	(0.451)	(0.442)	(0.381)	(0.391)
	FemaleMale	-0.019	0.026	-0.141	0.275	-0.179	-0.356	0.329	0.444
		(0.330)	(0.337)	(0.406)	(0.415)	(0.375)	(0.375)	(0.402)	(0.395)
	MaleFemale	-0.078	-0.068	-0.613	-0.423	-0.543	-0.759*	-0.254	-0.241
		(0.364)	(0.376)	(0.394)	(0.405)	(0.424)	(0.427)	(0.392)	(0.389)
	#obs	269	269	256	255	218	218	254	252
Competition word									
	FemaleFemale	0.002	0.060	-0.004	-0.014	0.160	0.178*	0.087	0.107
		(0.104)	(0.103)	(0.088)	(0.092)	(0.105)	(0.107)	(0.098)	(0.102)
	FemaleMale	-0.060	-0.011	0.116	0.096	0.073	0.074	0.195*	0.208*
		(0.091)	(0.091)	(0.101)	(0.108)	(0.099)	(0.101)	(0.101)	(0.110)
	MaleFemale	-0.002	0.016	0.043	0.033	0.061	0.061	0.157	0.192**
		(0.098)	(0.098)	(0.095)	(0.097)	(0.109)	(0.109)	(0.098)	(0.095)
	#obs	197	197	201	201	165	165	186	186
Competition math									
	FemaleFemale	-0.058	0.008	-0.135*	-0.147*	-0.164*	-0.143	-0.050	-0.026
		(0.084)	(0.083)	(0.078)	(0.081)	(0.094)	(0.094)	(0.073)	(0.080)
	FemaleMale	-0.008	0.040	-0.086	-0.090	-0.088	-0.075	0.154*	0.167*
		(0.078)	(0.081)	(0.087)	(0.089)	(0.093)	(0.096)	(0.087)	(0.093)
	MaleFemale	0.132	0.133	0.071	0.057	0.162	0.116	0.225***	0.220**
		(0.091)	(0.090)	(0.094)	(0.095)	(0.111)	(0.114)	(0.086)	(0.087)
	#obs	205	205	209	209	170	170	201	201

Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1.
FemaleFemale is 1 if the participant is a woman and the interviewer is a woman, 0 otherwise.
FemaleMale is 1 if the participant is a woman and the interviewer is a man, 0 otherwise.
MaleFemale is 1 if the participant is a man and the interviewer is a woman, 0 otherwise.

Table A8. P-values for OLS estimates of age and gender effects

	Within age: between genders				Within gender: between ages			
	> 30 women vs. men		≤ 30 women vs. men		Young vs. old women		Young vs. old men	
Controls	No	Yes	No	Yes	No	Yes	No	Yes
DG: 889 obs	0.047	0.095	0.934	0.978	0.633	0.571	0.337	0.572
UG: 620 obs	0.769	0.940	0.364	0.368	0.004	0.001	0.176	0.042
PD: 512 obs	0.784	0.751	0.897	0.843	0.526	0.285	0.766	0.291
TG: 632 obs	0.096	0.413	0.040	0.043	0.759	0.948	0.280	0.055
BoS: 675 obs	0.973	0.995	0.007	0.006	0.355	0.644	0.015	0.008
Risk: 997 obs	0.683	0.263	0.413	0.429	0.041	0.096	0.428	0.945
Word: 749 obs	0.359	0.235	0.825	0.926	0.688	0.373	0.456	0.119
Math: 785 obs	0.001	0.014	0.289	0.269	0.631	0.279	0.663	0.465

Table A9. Sample size and power analysis

	Within treatments				Between treatments: for full sample and within genders								
		Women vs. men			Baseline vs. priming			Baseline vs. female count'p			Baseline vs. male count'p		
	Baseline	Priming	Female count'p	Male count'p	Full sample	Women	Men	Full sample	Women	Men	Full sample	Women	Men
DG:													
required*	6065	593	5284	668	9084741	5130	5503	142912	9559	3773	28585	21237	2645
actual*	241	228	191	229	469	225	244	432	216	216	470	227	242
UG: req.	6446	2807	434	2571176	1176	1503	1026	84526025	1231	990	26152	6554	12388155
actual	173	163	122	162	336	162	174	295	151	144	335	227	174
PD: req.	3463	129473	12891	46470	2068	989	6995	1944	791	10141	1175	1888	799
actual	128	135	111	138	263	127	136	239	116	123	266	135	131
TG: req.	**0.7281****	4277	1905	1994	1511503	1117	1178	71578	1459	1219	162418	1179	2147
actual	169	167	134	162	336	151	185	303	147	156	331	156	175
BoS: req.	2975	38412	524	24948	127042	774300	10990	13855	84568	2098	59175	282308	9786
actual	179	180	150	166	359	174	185	329	169	160	345	172	173
Risk: req.	48473	14579	5155	9604	963	1608	656	21796	181981	6378	127436	41701	9937
actual	269	256	218	254	525	249	276	487	239	248	523	249	274
Word: req.	8734	2613	1160	7268	1182448	4381	4321	195861	2055	3144	26354	3289	40247
actual	197	201	165	186	398	189	209	362	180	182	383	190	193
Math: req.	659	502	**0.7327****	1248	5564	12298	4695	1343	7714	551	27068	8226	7958106
actual	205	209	170	201	414	199	215	375	188	187	406	197	209

*Required indicates the sample size required to receive a significant result in the respective tests, given the current sample and assuming a power of 0.80.
**The bolded numbers are tests where we find significant or marginally significant results. In these cases we have conducted a power analysis given the current sample – the figure in these cells are therefore power of the test rather than required sample sizes.

Appendix B

Control questions

Below we compare, for each measure, the full sample of participants with the group of participants who answered the control questions for that measure correctly. The sample comparison is made for gender, age, income and education. The results are shown in table B1. Table B2 presents the results of an OLS regression analysis of gender differences in preferences within each treatment, with and without controls. This analysis is similar to the analysis presented in Table 5.1 of the paper, but includes also those participants who did not answer one or more control questions correctly.

The dictator game had two control questions.[48] Of the 997 participants, 88% of the men and 90% of the women answered both questions correctly; this gender difference is not significant (p=0.2412). The two groups of participants (those that answered all questions correctly and those that did not) are similar in other characteristic as well: in average age (p=0.7926) and in proportions of high, medium and low income and education (income: p=0.0982, p=0.6442, p=0.8931; education: p=0.0953, p=0.2775 and p=0.2854 for high, medium and low respectively). There is no difference in dictator game giving between the two groups (p=0.2364).

In the ultimatum game, 63% of the men and 62% of the women answered the two control questions correctly, again this gender difference is not significant (p=0.7101).[49] The two groups of participants differ however on other characteristics. Participants who failed to answer the control questions correctly are on average four years older (p=0.0005), and the proportions of high and medium income and high education participants are also smaller (and the proportions of low income and low education participants is larger) in this group (income: p=0.0195, p=0.0149, p=0.0030; education: p<0.0001, p=0.0785 and p<0.0001 for high, medium and low respectively). This group also had a SEK 35 higher minimum accepted offer in the ultimatum game compared to the other group (p<0.0001).

[48] These were: 1. If you choose to keep 200 crowns, how much does the other person get? 2. If you choose to keep nothing, how much will the other person get?

[49] The control questions for the ultimatum game were: 1. What happens if you've said the minimum sum you would agree to is 200 crowns? 2. What happens if you've said the minimum sum you would agree to is 10 crowns?

In the prisoner's dilemma there were four control questions, asking about the payoffs in each of the four strategy combinations.[50] 52% of the men and 51% of the women answered these questions correctly; again there is no significant gender difference (p=0.7909). As with the ultimatum game, the two groups differ however in other characteristics. Participants who failed the control questions are on average five years older compared to the others (p<0.0001). The proportions of participants with high income and education are also smaller (and the proportion of participants with low education is larger) in this group (income: p=0.0384, p=0.3610, p=0.0404; education: p=0.0025, p=0.3354 and p=0.0004 for high, medium and low respectively). We find no differences in the propensity to cooperate between the two groups of participants (p=0.1193).

In the trust game, 66% of the men and 61% of the women answered the three control questions correctly, this difference is marginally significant (p=0.0948).[51] Participants who failed to answer the control questions correctly are on average three years older (p=0.0013). This group also have fewer high income participants, more low income participants, fewer high education participants and more low education participants than the group of participants who answered control questions correctly (income: p=0.0004, p=0.2952, p=0.0008; education: p=0.0032, p=0.6016 and p<0.0001 for high, medium and low respectively). The group that answered the control questions correctly invested on average SEK 14 more than the group that failed one or more control questions (p<0001).

The battle of the sexes decision had four control questions, and 67% of the men and 69% of the women answered these correctly.[52] The difference is not significant (p=0.4744). Participants who failed the control questions are on average four years older (p=0.0003) than the other group of participants. The proportion of participants with high income and education is also smaller and

[50] 1. If both of you choose A, how much do you get and how much does the other person get? 2. If you choose A and the other person chooses B, how much do you get and how much does the other person get? 3. If you choose B and the other person chooses A, how much do you get and how much does the other person get? 4. If both of you choose B, how much do you get and how much does the person/man/woman you have been paired with get?

[51] The questions for the trust game were: 1. How much money do you get and how much does the other party get if you decide to keep the 100 crowns? 2. How much money does the other party have for distribution if you send 50 crowns? 3. How much money do you get if you send 100 crowns to the person/man/woman you are paired with, and that person/he/she decides to send back half?

[52] 1. If both of you choose A, how much do you get and how much does the other person get? 2. If you choose A and the other person chooses B, how much do you get and how much does the other person get? 3. If you choose B and the other person chooses A, how much do you get and how much does the other person get? 4. If both of you choose B, how much do you get and how much does the person/man/woman you have been paired with get?

the proportion of participants with low income and education is larger in this group (income: p=0.0373, p=0.1246, p= 0.0094; education: p<0.0001, p=0.0843 and p=0.0002 for high, medium and low respectively). The propensity to play hawkish is marginally higher among those who failed to answer the control questions correctly compared to the other group (p=0.0825).

The risk decision had no control questions and we therefore use the full sample in our main analysis of this decision. The seven risk questions were ordered in increasing value of the safe alternative, with identical risky alternatives.[53] This means that we can identify behavior suggesting that the participant had understood the decision situation. Define answers as consistent if the participant does not swith from safe to risky choices as the value of the safe choice increases. With this definition, 88% of the men and 81% of the women acted consistently. This difference is significant (p=0.017). The consistent group is no different in age (p=0.1143), but the proportions of consistent participants are higher in the high income group (p<0.0001) and in the high education group (p=0.0163) and lower in the low income group (p<0.0001). There is no significant gender difference between the two groups in number of risky choices (p=0.5289).

The word competition had four control questions, and 75% of the women and 75% of the men answered them correctly.[54] There is no gender difference between the two groups (p=0.8998), but participants who failed the control questions are on average two years older (p=0.0478). These participants also seem to have slightly lower income and education compared to those that did answered the control questions correctly (income: p=0.1339, p=0.0810, p=0.0393; education: p=0.0512, p=0.9398 and p=0.0026 for high, medium and low respectively). The average competitiveness is similar between the two groups (p=0.9670).

The math competition decision had the same control questions as the word competition. 79% of the men and 78% of the women answered these questions correctly. There is also no gender difference in failing the control questions (p=0.5941), but participants who failed are on average three years older compared to others (p=0.0148). These participants also seem to have slightly

[53] The risk questions were formulated: Which alternative do you choose: X crowns or the toss of a coin to either win 200 crowns or get nothing at all, with X being SEK 40, 60, 80, 100, 120, 140 and 160.

[54] The control questions in the word task were: 1. How much will you get paid per word if you choose Payment Method 1? 2. How much will you get paid per word if you choose Payment Method 2 and you compose more words than the person you are paired with? 3. How much will you get paid per word if you choose Payment Method 2 and you compose fewer words than the person you are paired with? 4. How much will you get paid per word if you choose Payment Method 2 and you compose the same number of words as the person/man/woman you are paired with?

lower income and education compared to those that answered the control questions correctly (income: p=0.0010, p=0.6329, p=0.0185; education: p=0.0063, p=0.6261 and p=0.0022 for high, medium and low respectively). The average preference is similar between the two groups (p=0.9659).

Table B1. Comparing the full sample (all 997 participants) with the main sample for each measure

Characteristic	**Gender**	**Age**	**Proportion in income group** (SEK thousands)			**Proportion in education group** (years)		
Sample	0=male, 1=female		Low < 250	Medium 250-375	High > 375	Low ≤ 9	Medium 10-12	High > 12
All, N=997								
Mean	0.488	45.516	0.392	0.344	0.220	0.150	0.405	0.436
DG: Mean	0.495	45.133	0.392	0.346	0.227*	0.146	0.399	0.445*
p-value	0.241	0.793	0.893	0.644	0.098	0.285	0.278	0.095
UG: Mean	0.484	44.169***	0.356***	0.373**	0.244**	0.113***	0.384*	0.494***
p-value	0.710	< 0.000	0.003	0.015	0.020	<0.000	0.079	< 0.000
PD: Mean	0.484	42.752***	0.361**	0.357	0.246**	0.111***	0.391	0.482***
p-value								
TG: Mean	0.468*	44.299***	0.353***	0.356	0.255***	0.108***	0.411	0.472***
p-value								
BoS: Mean	0.496	44.278***	0.364***	0.36	0.239**	0.121***	0.387*	0.483***
p-value								
Word: Mean	0.487	44.951**	0.374**	0.359*	0.231	0.131***	0.405	0.454*
p-value								
Math: Mean	0.484	44.888**	0.373**	0.348	0.242***	0.132***	0.401	0.459
p-value								

P-values refer to comparisons with the sample including all 997 participants. *** p<0.01, ** p<0.05, * p<0.1.

Table B2. Gender differences within each treatment with the full sample (all 997 participants. OLS regressions with female as independent variable

Decision \ Treatment		Baseline		Priming		Female counterpart		Male counterpart	
		No controls	Controls	No controls	Controls	No controls	Controls	No controls	Controls
DG	*Female*	3.244	2.022	18.145	**23.122***	-5.314	-5.916	10.965	11.345
		(11.806)	(12.520)	(11.723)	(12.183)	(13.450)	(14.322)	(12.462)	(13.285)
UG	*Female*	-9.322	**-17.384***	8.073	12.021	-11.248	-14.755	-2.525	2.428
		(9.210)	(9.301)	(9.557)	(10.42)	(11.143)	(12.350)	(10.400)	(10.212)
PD	*Female*	-0.017	-0.000	-0.029	-0.054	0.045	0.028	-0.073	-0.070
		(0.061)	(0.064)	(0.063)	(0.065)	(0.068)	(0.073)	(0.063)	(0.065)
TG	*Female*	**-10.477****	**-11.497****	**-7.551***	-6.346	-6.630	-2.555	-5.986	**-8.718***
		(4.326)	(4.475)	(4.386)	(4.573)	(4.528)	(4.669)	(4.481)	(4.593)
BS	*Female*	0.002	0.027	-0.011	0.023	0.077	0.067	0.002	0.001
		(0.056)	(0.057)	(0.712)	(0.064)	(0.064)	(0.067)	(0.058)	(0.065)
Word	*Female*	-0.013	0.026	0.064	0.039	0.093	0.113	0.025	0.040
		(0.060)	(0.061)	(0.061)	(0.066)	(0.066)	(0.070)	(0.060)	(0.064)
Math	*Female*	-0.060	-0.021	-0.076	-0.088	**-0.130****	**-0.121***	-0.074	-0.039
		(0.054)	(0.055)	(0.057)	(0.060)	(0.063)	(0.067)	(0.055)	(0.061)
# obs.		269	265	256	247	218	214	254	246

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 Control variables are age, age squared, education, income and gender of interviewer.