

Pre-analysis plan for “Culture and Gender Differences in Competition”*

Karen Hauge,^a Andreas Kotsadam,^a and Anine Riege^b

Date: 29.04.2019

Abstract

Women are less likely than men to enter into competitions and this difference has been argued to be partly rooted in cultural differences. Using a large sample of Norwegian second-generation immigrants we will test the effect of culture using the epidemiological approach. In this plan we describe the hypotheses to be tested and how they will be tested. The description includes how the variables are coded, how we will deal with missing values, and the specification of the estimation equations. We also conduct a power analysis which suggests that we are able to identify relatively small differences. All deviations from the plan will be highlighted in the final paper.

*The research has been funded by NFR number 270772 and 250506/F10. We thank Kjell Arne Brekke, Åshild Johnsen, Johanna Möllerström, Gaute Torsvik, Oddbjørn Raaum, Ole Røgeberg, and participants at the Bargaining workshop at BI in Oslo for useful comments.

a The Ragnar Frisch Centre for Economic Research, Gaustadalleen 21, N-0349 Oslo, Norway.

b Department of Management, Kingston University London, Kingston Hill, KT2 7LB Kingston upon Thames, United Kingdom

1. Introduction

The classic finding of Niederle and Vesterlund (2007) that women are less willing to compete than men has survived countless replications and is generally robust to various design decisions made by researchers (see e.g., Niederle, 2015; Niederle and Vesterlund, 2011 for reviews). The gender difference is also of a considerable size: combining findings from 53 published papers using competition experiments shows that 52 percent of the men and 34 percent of the women choose to compete (Dariel et al, 2017). We ask whether the gender difference is due partly to culture.

Previous research suggests that cultural factors are important as there are large variations in the gender differences across space. The best-known examples in the economics literature are from a series of papers comparing preferences for competition in a Masai society in Tanzania to the ones in a Khasi society in India (Andersen et al. 2008; 2013; Gneezy et al. 2009). The two societies are described as being on the opposite poles of the gender spectrum with the Masai as particularly patriarchal and the Khasi are, if not matriarchal, at least matrilineal with respect to inheritance and women usually being heads of the household. In these diverse settings, men choose to compete in the Masai community and women in the Khasi community. The result is interpreted as cultural differences affecting willingness to compete. However, it is likely that other differences exist between the societies investigated.

Because societies differ along both institutional and cultural dimensions, and these dimensions affect each other, we cannot infer causal relationships between the factors by simply comparing societal outcomes. To empirically identify the importance of one of the dimensions, we need to separate the influence of culture from that of institutions. One can do so by either examining people with different cultures facing the same institutions, or individuals with similar culture facing different institutions. Several recent papers have used exposure to different institutions to show that women growing up under a communist ideology (Booth et al. 2018), before and after reforms (Zhang 2018), and in single-sex schools have different competitive behavior than other women. The present work uses cultural variation under the same institution. We use the so-called epidemiological approach (see Fernández 2011 for a review of the use of the approach in the economics literature) to investigate cultural influences on gender differences in willingness to compete. In particular, we follow Fernández and Fogli (2009), and restrict ourselves to study second generation immigrants, and proxy for culture by gender roles in the parents' country of ancestry. Norwegian second-generation immigrants are all born and raised

in Norway, and thus face the same labor market and the same institutions, while differing in their cultural heritage from their parents.

The gender difference in competition does not always point in the direction of more gender equal settings producing smaller differences in competitiveness. Cardenas et al. (2012), find that boys in Sweden are more willing to compete than girls, but find no gender difference in Colombia. Tungodden finds that both Norwegian children and their parents are more likely to compete if they are male, and Boschini et al. (2018) find the same for adults in a math task in Sweden. Khachatryan et al. (2015) find no difference in the willingness to compete in Armenia, Dariel et al. (2017) find that women do not want to compete less in the United Arab Emirates, and Zhang (2013) finds no difference in China. More generally, there are opposing theories with respect to individual gender differences and macro level gender equality. On the one hand, the social role hypothesis predicts that gender differences diminish in more gender equal places as the views of women's and men's roles and positions in society become more equal. On the other hand, the resource hypothesis suggests that more resources reduces the barriers for gender-specific ambitions and goals. Falk and Hermle (2018) find larger individual level differences in preferences between men and women in more gender equal countries.¹ We also show that the correlation between the gender difference in previous studies and gender equality in the country where the study is conducted is positive (see Table 1). Hence, it is not clear that culture has an effect nor in what direction such an effect would point.

2. Experimental design

We will access phone numbers from the Population Registers and recruit participants from different ancestry backgrounds by SMS text messages. The text message describes that participating is voluntary and that participants can earn money by following a link and answer some questions.² The link takes the respondents to a Qualtrics page where we have them conduct the competition experiment and answer some survey questions.

¹ There is also a literature in psychology arguing that there is a gender paradox in personality whereby gender differences are larger in more egalitarian countries (e.g., Kaiser, 2018; Schwartz and Rubel-Lifschitz, 2009; Stoet and Geary, 2018). The sociological literature highlights that the gender egalitarianism that has evolved in many western countries is one compatible with individualism, and hence, differences may increase in areas of social life where gender essentialism is believed to be strong (England 2010; Charles 2011).

² Invitation to survey sent by SMS (split into two SMS's due to character limit in each SMS). SMS 1: "The University of Oslo and the Frisch Centre invites you to participate in a research study. Contribute to research and earn between 50-400 NOK. Voluntary, approx. 20 minutes." SMS 2: "Read more about the research study from the University of Oslo and the Frisch Centre here: [Frisch.uio.no/s](https://frisch.uio.no/s). Participate here (mobile/pc/tablet): "

We follow the classical design of the competition experiment by Niederle and Vesterlund (2007). As there may be differences across men and women in risk preferences, altruism, feedback aversion, and confidence, the classical design includes a set of features that are tailored to rule out these other mechanisms and isolate unwillingness to compete. We return to these possible mechanisms below when we present the design.

Our experiment consists of 3 parts where the participants are solving as many tasks as possible for 90 seconds, in addition to answering a survey. All subjects receive a show-up fee of 50 NOK, and in addition one part of the experiment is randomly chosen for payment. The tasks consist of calculating the number of 1's on a screen as in Figure 1.

Figure 1: Example of a task.

Part 1 – Piece rate

This is Part 1. Please count 1's in each table below and write in the answer in the line below the respective table.

0	0	1	1	0
0	0	1	0	1
0	1	1	0	1
1	0	1	1	0
1	0	0	0	1

How many 1's are there?

In Part 1 of the experiment, participants earn 5 NOK for every task solved. In Part 2, payment is based on performance so that the best player in a group of 4 receives payment (if several people have equally high scores all of them win). The payment to the winner in Part 2 is 20 NOK per task. The groups are composed by randomly drawing three additional

individuals from a pilot sample (see section 4 for details about the pilot sample). This information is known for the respondents.

After Part 2 is played, the respondents are asked to self-assess their relative position in the group. The answer they give is incentivized by giving them 5 NOK if they answer correctly. This is a classic design solution to assess the importance of gender difference in confidence.

In Part 3 the participants will perform the same counting task, but now they are asked to choose what type of payment they want: either piece rate as in part 1 or competitive payment as in Part 2. This choice constitutes our main dependent variable measuring willingness to compete. After they make their decisions, the respondents start the counting task in Part 3. Importantly, in the case of competitive pay the respondents are competing against their group's (the same group as in part 2) performance in Part 2. This assures that payment is based on a setting where everyone competes, and therefore that the competition is with everyone and not only the ones with a high willingness to compete. In the classic design, competing with individuals that have already played is also the solution to control for altruism as the respondent's choices cannot affect the payoffs of the other players. However, in the present experiment, the altruism component is always controlled for as the respondents play against people that have already finished their experiment.

After Part 3 participants answer questions concerning how many correct answers they believe they provided in Part 1, guess how many correct answers the other participants have provided on average, and are asked about causal attribution. They also receive information about how many correct answers they actually provided in each of Part 1, 2 and 3 before proceeding to Part 4.

Part 4 consists of answering some survey questions about risk aversion, altruism, trust, attitudes related to gender equality, integration in the Norwegian society, survey measures of competitiveness and a vignette experiment measuring discrimination. They are told that they will receive between 25 and 50 NOK if this part is drawn for payment.

After the survey questions, respondents are told that they will get 25 NOK if Part 4 is chosen for payment. They are then asked if they want to negotiate a higher payment. If they answer yes they are asked to write in the amount they require as their payment (between 25 and 50), and they are given this amount up to a maximum of 50 NOK if part 4 is chosen.

Respondents are then informed about which part is selected for payment, and how much they will earn for participating. Finally, the respondents are given the choice to be paid to their bank accounts or to donate the money they have earned to a charity (The Red Cross, Doctors without borders, or UNICEF).

3. Data and coding of main variables

3.1 Main dependent variable

Willingness to compete (Compete):

Our main dependent variable measures the willingness to compete by the choice made in Part 3. We code *Compete* to equal 1 if the respondent chose competitive remuneration and zero if piece rate is chosen.

3.2 Main independent variables

Logged female labor force participation (FLFP)

Our main measure of gender equality culture is the female labor force participation rates (FLFP) in the country of ancestry in 2000, as reported in the World Development Indicators online database (<https://databank.worldbank.org/data/home>, variable SL.TLF.ACTI.FE.ZS). We measure FLFP in 2000 to have as complete coverage for our countries of ancestry as possible. For Yugoslavia we take the average of Bosnia and Herzegovina and Croatia, which are the two former Yugoslavian countries in our sample. For Faroe Islands we use data from Statistics Faroe Islands (https://statbank.hagstova.fo/pxweb/en/H2/H2_AM/) for the year 2005. Finseraas and Kotsadam (2017) take the FLFP in 1970 to proxy for the culture that parents brought with them to Norway. They show, however, that the results are almost identical when using FLFP in 2000 and that the two measures are very highly correlated. We take the natural log of FLFP as this is the standard measure used in papers using the epidemiological approach in economics (for a review see Fernandez 2011). FLFP generally correlates with other aspects of gender equality (Fernandez and Fogli 2009; Fernandez 2011) and allows for comparisons with other studies of ancestry culture in Norway (Finseraas and Kotsadam 2017; Finseraas et al. 2018).

Gender equality index (GGI)

As a complement, we also include a second measure, namely the World Economic Forums Gender gap index (GGI) as a proxy of culture. The index is based on 14 different country level indicators and is taken from the Global Gender Gap reports. We use the values from the 2016 report. This measure is also popular in the literature and has previously been used to study

culture by e.g. Blau et al. (2018), Fryer and Levitt (2010), Guiso et al. (2008), Kaiser (2018), Nollenberger et al. (2016), and Rodriguez-Planas and Nollenberger (2018). We only have this variable for 46 of our countries and several countries have the exact same value.

3.3 Other variables

Performance under piece rate: Number of problems correctly solved in in Part 1.

Performance under competitive incentives: Number of problems correctly solved in Part 2

Performance under choice: Number of problems correctly solved in in Part 3.

Earnings: How much money is made in each part (not including the show-up fee) if that part would have been chosen.

Guess about own performance: Based on a direct question of how many tasks the respondent thinks she solved in Part 1.

Overestimation: the ratio of *Guess about own performance* to actual performance (following Demiral, Apicella and Mollerstrom (2017)). A value greater than one means that one has overestimated own performance, a value less than one means that one has underestimated own performance.

Overconfidence: the ratio of *Guess about own performance* to guess of the performance of an average other (following Demiral, Apicella and Mollerstrom (2017)). A value greater than one means that the person thinks she did better than the average other, a value less than one means that the person thinks she did worse than the average other.

Confidence: Based on a direct question of relative tournament performance as in Niederle and Vesterlund (2007). The variable takes 4 values (1st, 2nd, 3rd, 4th).

Causal attribution: Answer to the question (used by Demiral, Apicella and Mollerstrom (2017)) “To what extent do you think your result in Part 1 is due to controllable (i.e., effort) versus uncontrollable (i.e., chance and difficulty) factors?” The variable ranges from 1 to 10, where 1 is labeled “no control” and 10 is labeled “full control”.

3.4 Variables from the survey questions in Part 4

From the survey questions in Part 4 we get the following variables:

Risk aversion: Answer to the question “In general, how willing are you to take risks?” The answer categories are from 1 to 10 where 1 is labeled “not willing to take risk at all”, and 10 is labeled “very willing to take risk”.

Attitude index: We choose to include three questions that have been used extensively in the World Values Survey (42 of our 50 ancestry countries, see below, have values from the WVS on these questions). These questions/statements are: “Being a housewife is just as fulfilling as working for pay”; “On the whole, men make better political leaders than women do”; and “A university education is more important for a boy than for a girl”. The response categories in the WVS are 1: Strongly agree, 2: Agree, 3: Disagree, and 4: Strongly disagree. There is also an option for Do not know, which is coded as -1 in the WVS. We recode these variables so that 0: Strongly agree, 1: Agree, 2: Do not know 3: Disagree, and 4: Strongly disagree. We then create an attitudes index by simply adding the three variables and dividing by three.

Believed gender difference in performance: The difference to the answers to the two following questions (both questions used in Gupta et al 2013): “What do you think is the average numbers of correct answers provided in Part 1 of women participating in this survey (including yourself if you are a woman)” and “What do you think is the average numbers of correct answers provided in Part 1 of men participating in this survey (including yourself if you are a man)”.

Vignette score: Part 4 includes a vignette experiment based on Finseraas et al. (2016) and adapted by us to the current study, where each respondent are asked to evaluate either a female (Ida) or a male (Martin) candidate for a leadership position. The candidates are asked to give a score for how good they think the candidate fits the position ranging from 1=very poorly to 6=very good. The vignette also varies whether the position is stated to be good for someone who likes to compete and whether the candidate likes to compete. We will, however, pool this information and only compare the versions with Ida and the versions with Martin in the main specification.

Negotiation: A dummy variable equal to 1 if the respondent choses to negotiate for a higher payment after Part 4 (a similar measure of negotiation is used in Small et al. (2007).

Integration: We aim to measure integration by the answers to the following questions:

Live with parents: Answer to the question: “Do you live with your parents?” (question from the World Value Survey). Yes=1 and No=0.

Language at home: Answer to the question: “Which language do you usually speak at home?” The answer categories are: “Only Norwegian”, Norwegian and other language”, “Other language”.

Language with parents: Answer to the question: “Which language do you usually speak with your parents?” The answer categories are: Only Norwegian”, Norwegian and other language”, “Other language”.

Vote: Answer to the question “Did you vote at the last election?” Yes=1 and No=0.

Been a leader: Answer to the question, Have you ever been a leader in an organization (for instance at work, leisure activities, in sport, culture, or politics)? Yes=1 and No=0.

Control: “Some people feel they have completely free choice and control over their lives, while other people feel that what we do has no real effect on what happens to them. Please use this scale (from 1 to 10) where 1 means “none at all” and 10 means “a great deal” to indicate how much freedom of choice and control in life you have over the way your life turns out” (question from World Value Survey).

Make parents proud: Based on the statement in WVS: “One of my main goals in life has been to make my parents proud”. We code it in the same way as we code the gender equality variables. Also this variable in the WVS for 42 of our countries as comparison.

Upbringing: We ask respondents about what characteristics they think are especially important to encourage kids to develop. They can choose up to 5 characteristics out of 14. We also ask the respondents what characteristics their parents put focus on when they themselves grew up. Based on these answers we will create ten (five for whether the characteristics are important to the individual herself and five for whether they are important for the parents) binary variables for whether religiosity, gender equality, competitiveness, hard work, and obedience are chosen among the 5 things.

Personal development motives: Based on the four questions from Bönte et al. (2017): “I can improve my competence by competing”, “Competition allows me to judge my level of competence”, “I use competition as a way to prove something to myself” and “Competition allows me to measure my own success”. Answers are given on a scale from 1 (does not apply at all) to 7 (fully applies). We make an index by adding the answers to these four questions and dividing by four.

General self-reported competitiveness: We also follow Bönthe et al. (2017) to create a measure of general self-reported competitiveness using the following four other questions: “I enjoy competing against others”, “I find competitive situations unpleasant”, “I like situations where I compete against others” and “When I try to reach a goal I prefer to compete against others instead of trying to reach the goal on my own”. Answers are given on a scale from 1 (does not apply at all) to 7 (fully applies). We reverse code the answers to the question “I find competitive situations unpleasant” and we create the index by using the average score to the four questions.

Variables for exploratory purposes

Generalized trust: Answer to the question: “Generally speaking, would you say that most people can be trusted or that you need to be careful in dealing with people” This question is not planned to be used in this paper but is asked to the respondents.

Hypothetical altruism: Answer to the question: Imagine the following situation: You unexpectedly received NOK 10 000. How much of this amount would you donate to a good cause? (Values between 0 and 10 000 are allowed.) This question is not planned to be used in this paper but is asked to the respondents.

Charitable giving: Equal to 1 if the respondent donates the money instead of being paid herself.

3.5 Data from registers (provided by Statistics Norway)

The following information from the administrative registers will be linked to our data set: Gender, age, place of residence, earnings, welfare benefits, family relationships (marital status, number of kids, number of siblings), and highest completed education. In addition, we will add parents’ country background, earnings, and highest completed education.

Data from Statistics Norway (SSB) will serve as a quality check for the data on parental country background from the Population Register. The data will be updated to be in accordance with SSB if there are any differences between the two sources. We will proceed in the following way: If a person in our sample according to the SSB data does not have a parent from the country of origin stated in the data from the Population Register, we will recode the person in accordance with the SSB data. We will drop a person from our study if according to the SSB data none of the parents are born in any of the 59 countries in our sample and both parents are not born in Norway. We will recode a person to our sample of Norwegians if both parents are born in Norway. If a person in the sample with Norwegian parents have at least one parent from the 59

ancestry countries they will be recoded into that group and if they have a parent from another country but none from the 59 countries they will be dropped. The coding and use of the other variables is explained in section 5.4.

4. The samples

4.1 The population

The main population is people born in Norway between 1980 and 2000 with at least one parent from one of 59 chosen countries (see below) and people born in Norway between 1980 and 2000 with two parents born in Norway.

4.2 The pilot sample

The first sample we have is a pilot sample. We conducted a pilot of 40 students at the University of Oslo in the beginning of April 2019. Participants were recruited through the mailing list of OECONLAB of students who previously have agreed to receive invitations to experiments. The pilot serves several purposes. First of all, we wanted to test the experiment, the survey, and the time it took the respondents to answer. Secondly, we need the pilot in order to have a group of people that have already played so that we can match the later players by randomly drawing people in order to determine payouts. For this sample we will not add any register data and we do not have any information on e.g., country background.

4.3 The study sample

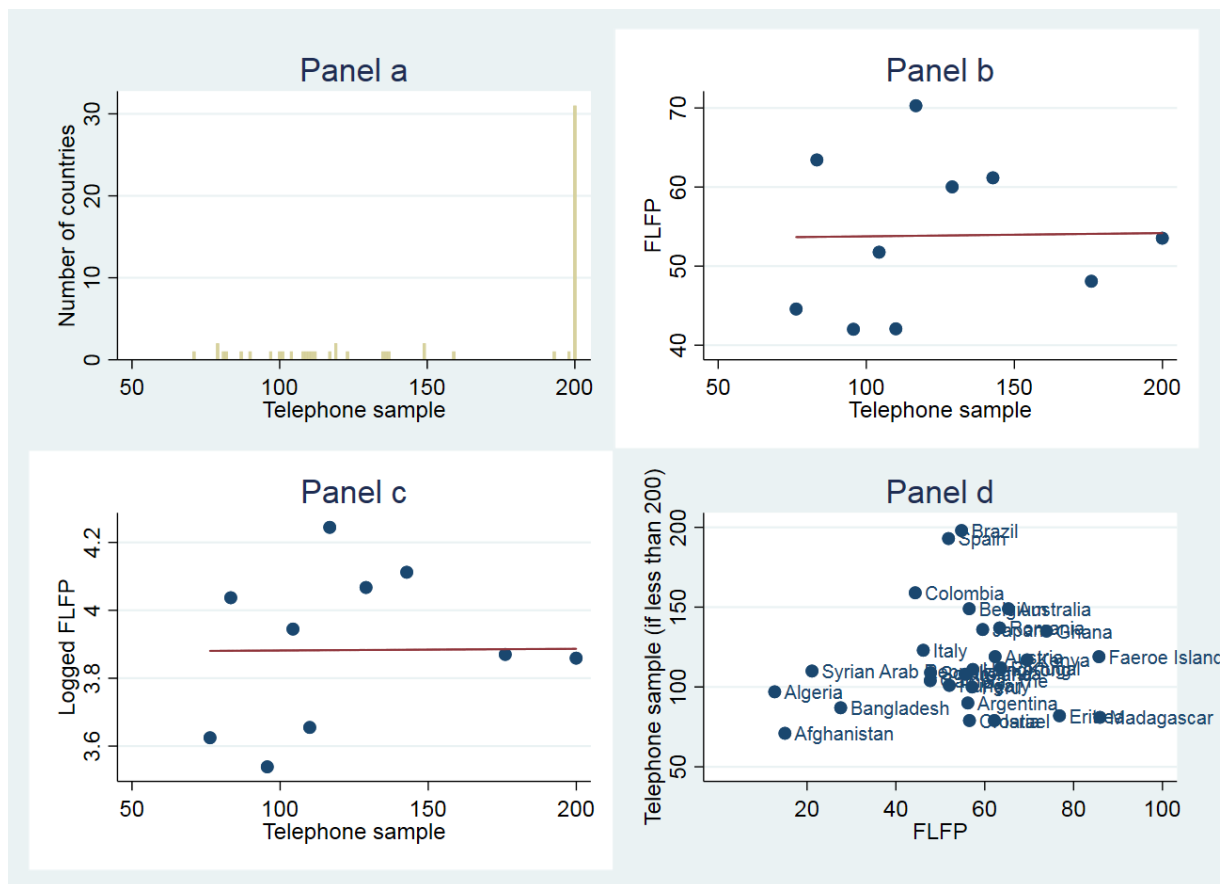
Our budget allows us to include around 2400 individuals with a second generation background. As we want variation in the ancestry country people are coming from and as we are uncertain about the response rate, we have chosen to try to contact people with at least one parent from the countries with the highest number of individuals recorded in the Population Register. We have calculated (see power analysis below) that we need around 2350 people and around 55 countries to have a reasonably well powered design. We want the samples to be similar, so we also restrict the sample in age. With these criteria in mind we find that there are 59 country backgrounds represented in the Population Registers with at least 71 Norwegian second generation immigrants born between 1980 and 2000, for which the Population register could provide a phone number. We asked for phone numbers for up to 200 people from these backgrounds. For the countries with more than 200 people we asked for a random draw of 200 people from the Population Register. We will stop data collection from a given country and sex when we have received 20 answers from such a group. Our aim is to collect data on 2360

individuals. As we expect that we will not be able to reach our quota of 20 people in every group (given that we will receive phone numbers for less than 200 people from some groups) we will oversample individuals at the tails of the FLFP distribution in order to increase power (see arguments related to increasing power below). More specifically, if we do not fill the quota of individuals from a country with a FLFP higher than the median FLFP (that is $FLFP > 57$), we will fill up with individuals of the same sex from the three countries in the upper tail of the FLFP distribution with most potential second generation respondents. These three country backgrounds are Sweden, Denmark, and Vietnam. If we do not fill the quota of individuals from a country with a below-median FLFP ($FLFP < 57$), we will fill up with individuals of the same sex from the three countries in the lower tail of the FLFP distribution with most potential second generation respondents. These three countries are Pakistan, Somalia, or Iran. Once we know the estimated response rates (see below) we will buy more phone numbers from these six tail countries.

In addition, we are including 300 individuals born in Norway between 1980 and 2000 with Norwegian born parents. This sample is based on random draw of 3000 individuals from the Population Register, for which phone numbers could be provided. This group will serve as a comparison group and a means to investigate the selection of our sample (more below). In particular, we are recruiting 100 of this sample using the same incentive scheme as for the individuals with parents from other countries, 100 using around half the incentives, and 100 with doubled monetary incentives. If there are no statistically significant differences between the groups in response rates and competitive behavior, we will use all 300 to compare with our other sample. If there are differences, we will only use the sample of individuals with Norwegian born parents that faced the same incentives and the eventual other sample that is not different.

The distribution of the raw sample (telephone sample) of individuals with parents from different countries is shown in Panel A of Figure 2. We see that we have a random draw of 200 individuals from 31 country backgrounds but for some country backgrounds, there are considerably fewer individuals. Fortunately, we see in Panel B and C that the country background raw sample sizes seem unrelated to our main measure of culture. Panel D shows the country names for the country backgrounds where there are less than 200 people in the raw samples.

Figure 2: Distribution of raw sample sizes (individuals with telephone numbers in the Population Register with different backgrounds) and relation to FLFP.



In Panels A and B of Figure 3 we show the correlation between our two measures of culture, FLFP and the gender equality index, and note that they are highly correlated also when we use our preferred operationalization and log FLFP as seen in Panel B. We also correlate our cultural proxies with the gender differences in competition found by previous studies. Dariel et al (2017) report estimates on the share of men and women willing to compete in studies following a design similar to Niederle and Vesterlund (2007). They included all the 53 studies found that were either published or forthcoming in November 2017. Based on these studies we have created a variable, Difference, which is equal to the percent of men willing to compete minus the percent of women in each study. In total, the gender difference is around 18 percentage points as 52 percent of men wanted to compete and 34 of the women. In Panels C and D of Figure 3 we correlate the gender differences with our two proxies. Interestingly, we find a positive correlation between gender equality and the gender difference in competition. We run regressions in Table 1 for the 43 studies from countries that overlap with our sample.

These regressions corroborate the positive correlation and show that our preferred measure (FLFP) is statistically significantly related to the difference and explains more of the variation in the difference in competitiveness.

Figure 3: Measures of culture in our sample and their relationship with gender differences in competition based on data from previous studies.

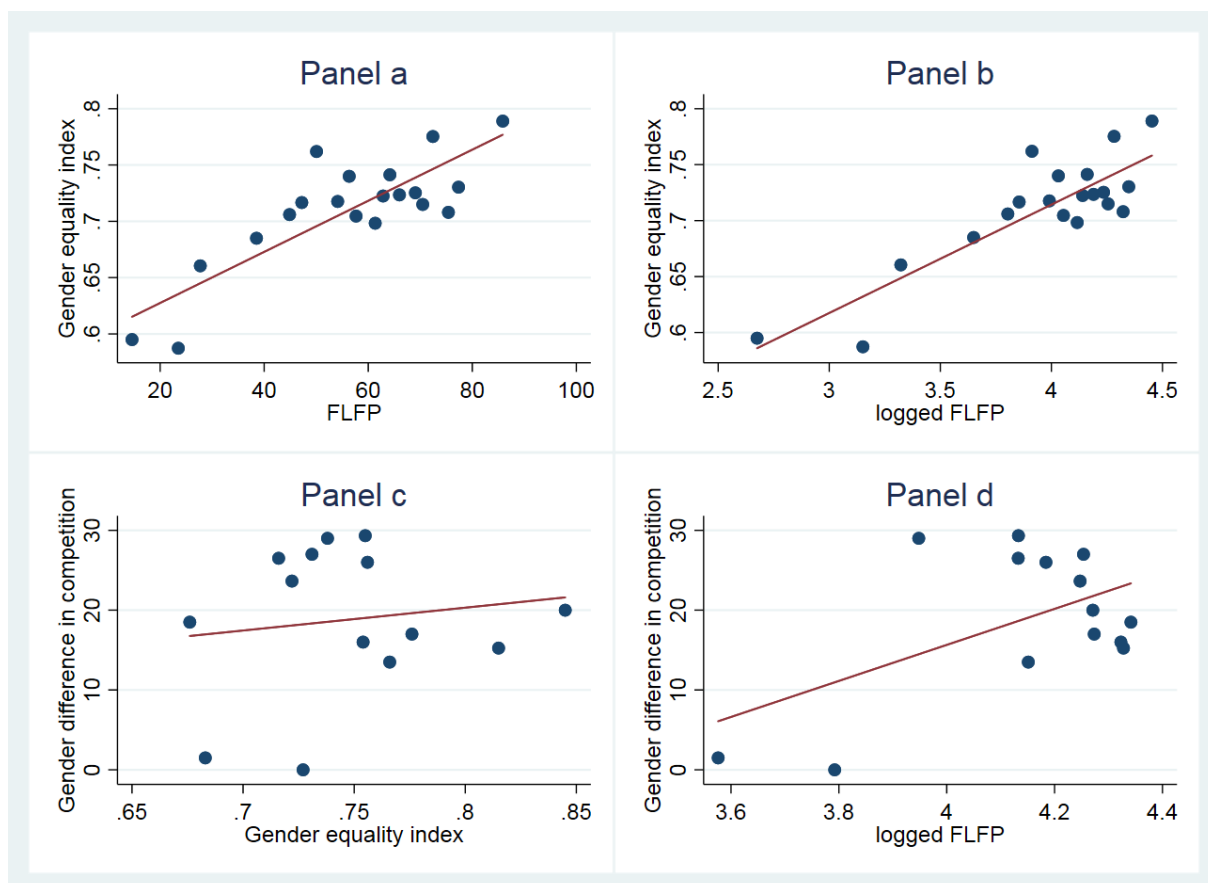


Table 1: Gender difference in competition in 43 previous studies (14 countries) and our cultural proxies.

VARIABLES	(1) Difference	(2) Difference
Gender index	25.948 (39.429)	
Logged FLFP		25.972*** (7.521)
Beta values	[0.085]	[0.457]
Observations	43	43
R-squared	0.007	0.209
Mean in sample	19.3	19.3
Robust standard errors in parentheses and standardized Beta coefficients in brackets. *** p<0.01, ** p<0.05, * p<0.1		

5. Empirical strategy and hypotheses

5.1 Main hypothesis

We will test our main variables of interest using different specifications and different samples. We have two main hypotheses, that the willingness to compete is correlated with culture for women and for men. We start by testing the gender difference in the total sample where we expect that women are less willing than men to compete.

We estimate the following regression:

$$Compete_{ip3} = \beta Female_i + \chi X_{it1} + \varepsilon_i \quad (1)$$

where i indexes individuals and p is the part of the experiment (p3 then implies part 3). The hypothesis is that β is negative (i.e., that women are less likely than men to choose to compete). X_{it} is a set of birth year fixed effects in the baseline specification and we will present results with and without these controls. To make the models fully saturated, we always partition the covariate space and add control variables as indicator variables rather than using their multi-valued codings (Athey and Imbens, 2017). If cells are too small, with less than 5 percent of the observations, adjacent cells are combined. When using interaction terms we will retain the continuous coding of the variables. If we have missing values on explanatory variables we will

code the variables as zero and include dummy variables controlling for missing status so that we do not lose observations. We use robust standard errors in all estimations.

We will present six versions of this regression in a first table: Three samples (All, Norwegian born with Norwegian born parents, and Second generation) and for all samples we will present results with and without controls.

We will move on to investigate the extent to which the cross-ancestry country differences explain variation in female and male competitiveness. We will do this first graphically and then move on to test our main hypotheses. In doing so, we first restrict the sample to either the female or the male second generation immigrants, add our measure of culture, and estimate the following regression:

$$Compete_{icp3} = \lambda Culture_c + \chi X_{ict1} + \varepsilon_{ic} \quad (2)$$

We add the subscript c for country of ancestry. The measure of culture will be standardized to have mean zero and standard deviation of 1 for presentational purposes. The error term, ε_{ic} is now clustered at the country of ancestry level. The main hypothesis is that λ is statistically significantly different from zero. The main specification is one without any controls and where the cultural proxy used is logged FLFP. Note that we treat it as an open question whether there is a smaller or larger difference for individuals with a background in more equal countries.

We will then test if the gender difference is moderated by our proxy for culture. In doing so, we use the full sample of second generation immigrants, add country of ancestry fixed effects and an interaction term between Female and our measure of culture and estimate the following regression:

$$Compete_{icp3} = \beta Female_{ic} + \partial Country_c + \lambda Culture_c * Female_{ic} + \chi X_{ict1} + \varepsilon_{ic} \quad (3)$$

The only reason for this last specification not being our main specification is that we have low power in estimating such difference in differences estimates. Hence, we focus mainly on the results for the samples of women and men. If we, for instance, find that competitiveness is correlated with culture for women and we also find a similar correlation for men we will conclude that gendered culture does not seem to be important for competitiveness (but culture in general may be so).

5.2 General tests of mechanisms

Testing for mechanisms is generally difficult as there are so many tests that can be done, which in turn are often conditional on the findings. In the analysis of the effects of culture, there is a general problem of post treatment bias as almost all other possible controls can also be affected by culture. For that reason, we are generally more confident in tests that estimate total effects of culture than in tests that control for other factors. We will nonetheless conduct tests of the latter form as well, but highlight that they may suffer from a mediation bias.

In the literature on gender differences in competition a series of standard tests of mechanisms have been used. We will conduct those tests and see how they affect our interpretation of the gender differences as well as the effects of culture on female and male competitiveness and on the gender difference.

We will first conduct the same analyses replacing *Compete* with performance and earnings. We will then test whether performance explains differences in the gender gap (equation 1) and in the culture effect (equations 2 and 3) by controlling for Part 2 performance in the baseline regressions. In equation 2 we interact part 2 performance with culture and in equation 3 we further interact Part 2 performance with the culture*female coefficient as well as with female. As culture may affect performance, the post treatment caveat is noted.

We will move on to investigate the role played by confidence and risk by first replacing *Compete* in the baseline regressions (equations 1-3) with these variables. We will then go back to the specification with *Compete* as the dependent variable and add risk and confidence as controls (interacted with the culture coefficient in equation 2 and with the culture*female coefficient in equation 3 and the female coefficient in equations 1 and 3). In addition to these standard tests performed in the literature we will also investigate the role played by our other measures of confidence (*Guess about own performance*, *Overestimation*, and *Overconfidence*) in a similar fashion.

5.3 Tests of mechanisms conditional on the results

5.3.1 If we find support for the main hypothesis

If we find that culture can explain female or male competitiveness and gender differences in competition, it is not certain that these differences are due to differences in gender culture. While specifically focusing on the difference across women and men makes this more likely, there may always be other factors that are correlated with the gender difference in behavior and cross

country differences in for example female labor force participation. One such factor may be socioeconomic status (SES) to the extent that this affects the behavior of women and men differently. Almås et al (2014) find that kids in Norway from lower socioeconomic backgrounds are less willing to compete. They furthermore find that this interacts with gender so that the gender difference is largest among kids with a higher socioeconomic background and smaller among the children with a lower socioeconomic background. In fact, there is no gender difference among the children with a lower socioeconomic background in their sample. In particular, it is the boys from the lower SES families that compete less.

As SES is a post-treatment variable (it is likely affected by culture) it is not straightforward to control for it. We will therefore first investigate if the pattern we observe is consistent with the explanation in Almås et al. (2014) by asking:

- Is there a correlation between culture and SES (measured as parents average earnings in the most recent year for which we have data)?
- Is the gender difference in competition negatively correlated with culture?
- Is the gender difference driven more by men than by women?

If the answers to these three questions are all yes, we would face a problem in differentiating our cultural argument from the argument in Almås et al. (2014). We will then also control for parental earnings in additional regressions, despite problems of post treatment bias.

We will also test if there is an effect of culture on attitudes toward gender equality. In particular, we will run the regressions (2) and (3) above but replacing *Compete* with the attitude index based on our survey questions. We will also use the variable vignette score from the vignette experiment as an outcome variable and see if culture affects discrimination of the female candidate (Ida in the vignette). As the main interest lies in whether the female candidate is given a different score and whether this discrimination differs by culture we will estimate a regression of the following form:

$$Score_{ic} = \beta Ida_{ic} + \partial Country_c + \lambda Culture_c * Ida_{ic} + \chi X_{ict1} + \varepsilon_{ic} \quad (4)$$

The standard errors will be clustered at the ancestry country level. The specification can also be augmented by adding Female and interact it with the other variables but the main test is to see if there is a difference in discrimination across cultures, which may be true for both men and women. If we find effects on the gender equality index and on discrimination as well, we will argue that culture probably matters for competition via its effects on gendered attitudes.

5.3.2 *If we do not find support for main hypothesis*

If we find that culture cannot explain female or male competitiveness or gender differences in competition, we still cannot conclude that culture is not important. There are a number of other potential explanations for a null-finding. Some of these explanations are general enough to apply to all studies using the epidemiological approach: The immigrants are not a representative sample of the population in the countries of origin and the eventual culture that they have brought may be diluted across generations.

To test if the second-generation immigrants have not internalized the cultural gender norms of their parents' home country we will use the survey questions in Part 4. First of all, we will compare the gender attitudes (the index) reported in the survey by Norwegians and second-generation immigrants. We will then test the correlation between log FLFP and gender attitudes. The questions in the attitude index are also included in the World values survey, which includes data from 42 of the countries included in the present study. We will therefore investigate if there is a correlation between ancestry country score on these variables and the scores of the second-generation immigrants with parents from those countries. These tests will give us an indication of whether second-generation immigrants have internalized the gender norms from their parents' home countries or not. If they have not, we can say that neither gender attitudes nor gendered competition behavior is affected by culture as measured using our methods.

It could also be that the second-generation immigrants have not internalized other cultural aspects of their parents' home country. We will therefore conduct the same type of tests for many variables (see Table 2).

If we find that culture is important for these other variables, but not for competition, we will argue that gender culture does not seem important for gender difference in competitive behavior. If we find that culture does not explain competition nor differences in these other variables it may be due to general concerns about the epidemiological approach (selection of immigrants and dilution of culture across generations) or due to study specific factors, such as selection or specific outcomes. As previous studies have managed to find effects of culture on the second generation in Norway (Finseraas and Kotsadam 2017), we would have to look closer into study specific limitations. A particular limitation of most lab experiments is that the respondents may not be representative and that this may bias our results. We discuss how we can use our data to think about this issue in the next section.

Table 2: Checks and tests with secondary variables.

Variable	Mean in second generation sample	Mean in sample with Norwegian born parents	Correlation with log FLFP	Correlation 2 nd gen sample/ WVS ancestor country
Gender attitudes index				
Control				
Make parents proud				
Important to self: Gender Equality				NA
Important to self: Religion				
Important to self: Competitiveness				NA
Important to self: Hard work				
Important to self: Obedience				
Important to parents: Gender Equality				NA
Important to parents: Religion				NA
Important to parents: Competitiveness				NA
Important to parents: Hard work				NA
Important to parents: Obedience				NA
Voting				NA

We will compare what the individual think are important for herself and the individuals perception of what her parents emphasized during her upbringing. We will also use our other measures of integration (language at home, language with parents) and descriptively discuss the patterns in these.

5.4 Tests of representativeness and differential selection

There are two main concerns regarding representativeness in our setting. The first is that the individuals participating in the experiment are different from the general population of second-generation immigrants from these countries in Norway. The second is the possibility of differential selection across countries threatening the internal validity of our conclusions.

As a first check we will investigate if the response rates are different across gender and across different countries. As we have different numbers of people in the pools for the different countries, we will use the smallest group as our point of departure. The smallest group consists of women from Afghanistan where we have 31 individuals. In the first week of the trial we will therefore invite 31 people from each group and we will create a variable, *Response rate*, for each group based on the percentage that replies out of these individuals.³ The variable will be top-coded for all groups with a response rate of over 64.5 percent. For these groups the response rate is “At least 64.5 percent”. In testing for differences we will code the response rate as a continuous variable where the top coded will be given the value 65. Our main test will be to regress the response rate on Culture for men and women separately as in equation 2 but we will also regress it on Female, country fixed effects, and Female*Culture as in equation 3. We will also experiment with giving the top coded group values of 70, 75, and 85 and discuss how sensitive the estimates are to these changes.

We will send out invitations to the individuals on Tuesday April 30 and send a reminder on Thursday May 2. The week after we will send new invitations to the different groups based on the response rates calculated for the first week (as there is a cost associated with sending text messages). The response rates will also be used for a calculation of how many extra phone numbers we need to buy from the countries at the tails of the FLFP distribution in order to reach our target of 2360 individuals.

We will also keep track of how easy it was to get the respondents to answer among those that finally answered. Assuming that there is a correlation between hard to reach respondents and non-respondents in their competitive behavior this may teach us something about the selection problem (Heffetz and Rabin 2013; Heffetz and Reeves 2018). We will test whether there is a difference in competitiveness between those who only received the first invitation and those who received a reminder and we will test if the gender difference and the culture difference is statistically significantly different.

We will also test if the respondents are representative by comparing them to the entire population 2nd generation immigrants with the same country backgrounds in Norway within the

³ For the three Norwegian samples (with the different incentives), the experimental samples will be of 50 people per gender within each sample and we will therefore send out 155 invitations to each of the six groups during the first week.

specific age category, controlling for age and gender. The variables we will test are listed in Table 3.

Table 3: Comparison of our sample with other Norwegians with parents from the same ancestor countries

Variable	Correlation 2 nd gen sample/general population of 2 nd gen immigrants in a regression controlling for age and gender	Mean in 2 nd generation experimental sample	Mean in 2 nd generation population
Highest education (all education variables are recoded into years)			
Mother highest education when the respondent was 16 years old			
Father highest education when the respondent was 16 years old			
Earnings (all earnings are based on “earnings for pension calculations” in the last year for which we have register data).			
Mother earnings			
Father earnings			
Married/cohabitant			
Divorced			
Widow			
Single			
Number of siblings			
Number of persons in the household			
Share foreign born in the municipality of residence			

We will also test for selection effects by using the entire population in our age groups from the 59 countries in the registry data and for each variable X in table 3 run regressions of the type:

$$X_{ics} = \alpha Sample_s + \beta Female_i + \partial Culture_c + \phi Culture_c * Sample_s + \mu Culture_c * Female_{ic} + \kappa Female_i * Sample_s + \dots (3) + \lambda Culture_c * Female_i * Sample_s + \chi Age_i + \varepsilon_{ic},$$

where *Sample* refers to our experimental sample. With, for instance, standardized Logged FLFP, this would give us the difference between our sample and the total population sample for

men from the country with average FLFP (in the coefficient α) and if the difference is different for women (κ), both of which are important for external validity. The regression result will also tell us if there is differential selection for men (ϕ) and the added differential selection for women (λ), both of which are important for the internal validity of the estimates.

We will also include a sample of individuals born in Norway to Norwegian born parents and compare our sample of “Natives” to those individuals. This will be important to test for selection in the general population and to investigate which factors affect selection into experiments more generally.

5.5. Further exploratory analyses:

We will investigate the effects of culture on several other outcomes in addition to the ones from the standard competition experiment. Running the regression specification in equations (2) and (3), we will test if there is an effect on *Negotiation* and on the index for *Personal Development Motives (PDM)*. Both the competition experiment outcomes (in particular Compete) and PDM will be correlated with earnings in the registry data to see how they correlate with labor market outcomes. We will also investigate the relationships between these variables and the variables for liking to compete, the believed gender difference in performance, and having been a leader in an organization.

Demiral, Apicella and Mollerstrom (2017) find that when competing against others, men think their results are more due to controllable factors, while women think their results are more due to uncontrollable factors. Interestingly, they find that this difference in attribution is stronger when women believed that they won a tournament. The authors argue that the gender difference in *causal attribution* can be an important mechanism behind the gender difference in competitiveness. We will test whether there is an effect of culture on causal attribution, again by running a regression of the form outlined in equations (2) and (3).

6. Power calculation

To test our main hypotheses, which are that culture is correlated with competitiveness for women and men, we have a sample of around 20 women and 20 men from 59 different countries. We calculate power by simulation and if we use a p-value of 0.05 and 80 percent power we have a minimum detectable effect (MDE) of 4 percentage points for each of the two samples. Accounting for the fact that we are testing two hypotheses using the false discovery rate method developed by Benjamini and Hochberg (1995) and adjusting the p-value to 0,025 our MDE

would be 6 percentage points. The MDE will be lower when we control for variables and as a result of having more observations in the tails of the FLFP distribution (List et al. 2011). We therefore think that our study is powered to detect relatively small differences.

7. IRB approval and consent

This study was reviewed and approved by the IRB officer at the Frisch center. Informed consent for storing and combining the data is given by the participants.

8. Archive

The pre-analysis plan is archived before any data is collected. We archive it at the registry for randomized controlled trials in economics held by The American Economic Association: <https://www.socialscienceregistry.org/> on April 29, 2019. We will start the experiment on April 30, 2019.

References

- Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. (2013). Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics*, 95(4), 1438-1443.
- Andersen, S., Bulte, E., Gneezy, U., & List, J. A. (2008). Do women supply more public goods than men? Preliminary experimental evidence from matrilineal and patriarchal societies. *The American Economic Review*, 98(2), 376-381.
- Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300.
- Bertrand, M. (2011). New perspectives on gender. *Handbook of labor economics*, 4, 1543-1590.
- Blau, F., Kahn, L., Eng, A., Meyerhofer, P., and Willén, A. (2018) Culture and Gender Allocation of Tasks: Source Country Characteristics, Housework, and Childcare Division of Labor among US Immigrants. Mimeo, 2018
- Booth, Alison, Elliott Fan, Xin Meng, and Dandan Zhang. "Gender differences in willingness to compete: The role of culture and institutions." *The Economic Journal* (2018).
- Booth, A., & Nolen, P. (2012). Choosing to compete: How different are girls and boys?. *Journal of Economic Behavior & Organization*, 81(2), 542-555.
- Boschini, Anne, Anna Dreber, Emma von Essen, Astri Muren, and Eva Ranehill. "Gender, risk preferences and willingness to compete in a random sample of the Swedish population." Mimeo (2018).
- Bowles, H. R. (2012). Psychological perspectives on gender in negotiation. *The Sage handbook of gender and psychology*.
- Bowles, H. R., Babcock, L., & McGinn, K. L. (2005). Constraints and triggers: situational mechanics of gender in negotiation. *Journal of personality and social psychology*, 89(6), 951.
- Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, 103(1), 84-103.
- Bowles, H. R., & McGinn, K. L. (2008). Gender in Job Negotiations: A Two-Level Game. *Negotiation Journal*, 24(4), 393-410.
- Bönte, Werner, Sandro Lombardo, and Diemo Urbig. "Economics meets psychology: experimental and self-reported measures of individual competitiveness." *Personality and Individual Differences* 116 (2017): 179-185.
- Charles, M. (2011). A world of difference: international trends in women's economic status. *Annual Review of Sociology*, 37, 355-371.
- Demiral, E.E., Apicella, C.L., Mollerstrom, J. (2017). Gender differences in Self-and Other-competition. Unpublished.
- Dariel, A., Kephart, C., Nikiforakis, N., & Zenker, C. (2017). Emirati women do not shy away from competition: evidence from a patriarchal society in transition. *Journal of the Economic Science Association*, 3(2), 121-136.
- England, P. (2010). The gender revolution: Uneven and stalled. *Gender & society*, 24(2), 149-166.
- Falk, A., & Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, 362(6412), eaas9899.
- Fernández, R. (2011). Does culture matter?. In *Handbook of social economics* (Vol. 1, pp. 481-510). North-Holland.
- Fernández, R., & Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics*, 1(1), 146-77.

- Finseraas, H., Johnsen Å.A., Kotsadam, A. and Torsvik, G. (2016). Exposure to female colleagues breaks the glass ceiling – Evidence from a combined vignette and field experiment. *European Economic Review*, 90, 363-374.
- Finseraas, H., & Kotsadam, A. (2017). Ancestry culture and female employment—An analysis using second-generation siblings. *European sociological review*, 33(3), 382-392.
- Fryer Jr, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-40.
- Gneezy, U., Leonard, K. L., & List, J. A. (2009). Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica*, 77(5), 1637-1664.
- Gong, B., & Yang, C. L. (2012). Gender differences in risk attitudes: Field experiments on the matrilineal Mosuo and the patriarchal Yi. *Journal of Economic Behavior & Organization*, 83(1), 59-65.
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, math, and gender. *Science*, 320(5880), 1164-1165.
- Gupta, N.D., Poulsen, A. and Villeval, M.C. (2013). Gender matching and competitiveness: Experimental evidence. *Economic Inquiry*, 51(1), 816-835.
- ILO (2018). ILOSTAT Database. Geneva, Switzerland: ILO.
- Kaiser, T. (2018) Gender equality and sex differences in personality: evidence from a large, multi-national sample.
- List, J. A., Sadoff, S., & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics*, 14(4), 439.
- Mazei, J., Hüffmeier, J., Freund, P. A., Stuhlmacher, A. F., Bilke, L., & Hertel, G. (2015). A meta-analysis on gender differences in negotiation outcomes and their moderators. *Psychological bulletin*, 141(1), 85.
- Niederle, M. (2015). Gender. *Handbook of Experimental Economics*.
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 1067-1101.
- Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1), 601-630.
- Nollenberger, N., Rodríguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5), 257-61.
- Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of personality and social psychology*, 87(2), 157.
- Rodríguez-Planas, N., & Nollenberger, N. (2018). Let the girls learn! It is not only about math... it's about gender social norms. *Economics of Education Review*, 62, 230-253.
- Schwartz, S. H., & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: Effects of gender equality. *Journal of personality and social psychology*, 97(1), 171.
- Small, D.A., Gelfand, M., Babcock, L., Gettman, H. (2007) Who Goes to the Bargaining Table? The influence of Gender and Framing on the Initiation of Negotiation. *Journal of Personality and Social Psychology*, 93(4), 600-613.
- Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological science*, 29(4), 581-593.
- Sutter, M., Bosman, R., Kocher, M. G., & van Winden, F. (2009). Gender pairing and bargaining—Beware the same sex!. *Experimental Economics*, 12(3), 318-331.
- Tungodden, Jonas, (2018). Preferences for Competition: Children Versus Parents. Mimeo (2018)

- Zhang, Y. J. (2013). Can experimental economics explain competitive behavior outside the lab? Working paper.
- Zhang, Y. J. (2018). Culture, institutions, and the gender gap in competitive inclination: evidence from the communist experiment in china. *The Economic Journal*.