Experimental Design and Participants

1 Experimental Design

We follow the classical design of the competition experiment by Niederle and Vesterlund (2007), including several measures designed to parse out other differences between men and women, such as risk preferences, altruism, and self-confidence. Our experiment consists of three rounds where the participants solve as many tasks as possible for 90 seconds and answer a survey. All participants received a show-up fee of 50 NOK, in addition to receiving payment for a randomly chosen part of the experiment (one of the three rounds or the survey part).

To reach the desired participant pool and the required sample size, we needed to recruit participants living in all parts of Norway. We therefore conducted the experiment as an incentivised online experiment, which means losing some control. For example, we cannot rule out respondents letting someone else answer in their place. However, in general, online experiments have previously been found to replicate laboratory experimental results (Edelman, 2012; Horton *et al.*, 2011). The non-standard participant pool (see Section 2 for the definition of the population) makes our experiment an artefactual field experiment in the topology of Harrison and List (2004). We recruited respondents via text messages, inviting them to follow a link to a voluntary study where they could earn money.[1] Data was collected between May and August 2019.

We modified two aspects of the classical design of Niederle and Vesterlund (2007) to conduct the experiment online: the task used and the composition of competition groups. The task in our experiment is counting the number of 1's in 5 x 5 matrices consisting of zeroes and ones (see Appendix Figure A.1). The task has two advantages: First, it reduces the possibilities of cheating in an online setting. Second, men and women have been found to perform equally well at this task in previous online experiments (Apicella *et al.*, 2017, 2020; Lezzi *et al.*, 2015).

In round 1 of the experiment, participants could earn 5 NOK (0.55 USD) for every task solved. In round 2, the competition round, payment was based on relative performance where player(s) with the highest performance within each group of 4 would receive a payment of 20 NOK per task. As in several recent studies, we informed participants that they would compete against the performance of three randomly drawn participants from an earlier pilot (e.g., Mayr *et al.* (2012), Burow *et al.* (2017), and Buser *et al.* (2020)).[2] After round 2, respondents were asked to self-assess their relative position in the group, where correct responses were incentivised with 5 NOK. This is a classic design solution assessing the importance of gender differences in self-confidence.

In round 3, participants performed the same counting task but were asked to choose between the two payment schedules: either piece-rate payment as in round 1 or competitive payment as in round 2. This choice constitutes our main dependent variable measuring WTC. After making their decision, respondents started the counting task in round 3. Importantly, respondents who chose to compete, competed against the same group as in round 2. This approach both controls for altruism considerations and benchmarks payments against the performance of all participants in the group, not only the ones with a high WTC.

---

[1] The first text message read: "The University of Oslo and the Frisch Centre invite you to a research study. Contribute to research and earn 50–400 NOK. Voluntary, approx. 20 min". Text message 2 (sent simultaneously with text message 1) read: "Read more about the research here: Link. Participate here (phone/tablet/computer): Link". The reminder sent out two days later read: "We remind you about the invitation to participate in a research study. Your response is important for the research. Participate here: Link.".

[2] We conducted a pilot of 40 students at the University of Oslo in the beginning of April 2019. We recruited participants through the mailing list of OECONLAB of students who previously have agreed to receive invitations to experiments.

After completing round 3, participants were given a survey measuring several control variables, such as beliefs about their own performance, risk aversion, and altruism (see Table ?? for all main variables and the Appendix Section A.6 for the full questionnaire). We then informed respondents about which round was selected for payment, and how much they earned.

2   Participants

Our study population includes people born in Norway between 1980 and 2000 with at least one parent born outside of Norway in one of 59 different countries. We chose the 59 ancestral countries with the highest number of individuals recorded in the Population Register in Norway. The age group has a parent generation with a fair share of immigrants from various ancestral countries, giving us enough potential participants to recruit from. In addition, restricting the age of the study population to be between 19 and 39 years old ensures that people in the sample are not too different from each other, while also avoiding adolescence and menopause, both of which affect WTC (Andersen *et al.*, 2013; Flory *et al.*, 2018). See Appendix Table A.1 for a list of countries and sample sizes in our study.

We aimed at recruiting up to 40 participants from each country background: 20 women and 20 men. We intended to invite a random draw of 200 people from each country-gender cell, but not all country backgrounds had 200 people with phone numbers registered in the Population Register. The smallest group had 71 people, and with expected response rates below 28%, our goal would not be possible. Therefore, we invited all people available from countries with fewer than 200 people registered with a phone number and a random draw of 200 people from countries with more than 200 people registered with a phone number.We oversampled individuals at the tails of the FLFP distribution to increase power (List *et al.*, 2011).[3] In total 1,943 consenting respondents completed the competition experiment (round 3).

---

[3]Our pre-registered decision rule dictates that if we were unable to obtain enough consenting participants from a country with a higher FLFP than the median (i.e., FLFP above 57), we invited additional participants from three other countries in the upper tail of the FLFP distribution with high numbers of potential participants. These three country backgrounds are Sweden, Denmark, and Vietnam. If we did not get enough consenting participants from a country with a below-median FLFP (FLFP below 57), we would invite more individuals of the same gender from the three countries in the lower tail of the FLFP distribution with most potential second-generation respondents. These three countries are Pakistan, Somalia, and Iran.

References

Andersen, S., Ertac, S., Gneezy, U., List, J.A. and Maximiano, S. (2013). 'Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society', *Review of Economics and Statistics*, vol. 95(4), pp. 1438–1443.

Apicella, C.L., Demiral, E.E. and Mollerstrom, J. (2017). 'No gender difference in willingness to compete when competing against self', *American Economic Review*, vol. 107(5), pp. 136–40.

Apicella, C.L., Demiral, E.E. and Mollerstrom, J. (2020). 'Compete with others? no, thanks. with myself? yes, please!', *Economics Letters*, vol. 187, p. 108878.

Burow, N., Beblo, M., Beninger, D. and Schröder, M. (2017). 'Why do women favor same-gender competition? evidence from a choice experiment', .

Buser, T., Cappelen, A.W., Gneezy, U., Hoffman, M. and Tungodden, B. (2020). 'Competitiveness, gender and handedness: a large-sample intercultural study', *NHH Dept. of Economics Discussion Paper*, (02).

Edelman, B. (2012). 'Using internet data for economic research', *Journal of Economic Perspectives*, vol. 26(2), pp. 189–206.

Flory, J.A., Gneezy, U., Leonard, K.L. and List, J.A. (2018). 'Gender, age, and competition: A disappearing gap?', *Journal of Economic Behavior & Organization*, vol. 150, pp. 256–276.

Harrison, G.W. and List, J.A. (2004). 'Field experiments', *Journal of Economic literature*, vol. 42(4), pp. 1009–1055.

Horton, J.J., Rand, D.G. and Zeckhauser, R.J. (2011). 'The online laboratory: Conducting experiments in a real labor market', *Experimental economics*, vol. 14(3), pp. 399–425.

Lezzi, E., Fleming, P. and Zizzo, D.J. (2015). 'Does it matter which effort task you use? a comparison of four effort tasks when agents compete for a prize', *A Comparison of Four Effort Tasks When Agents Compete for a Prize (April 15, 2015)*.

List, J.A., Sadoff, S. and Wagner, M. (2011). 'So you want to run an experiment, now what? some simple rules of thumb for optimal experimental design', *Experimental Economics*, vol. 14(4), p. 439.

Mayr, U., Wozniak, D., Davidson, C., Kuhns, D. and Harbaugh, W.T. (2012). 'Competitiveness across the life span: the feisty fifties.', *Psychology and aging*, vol. 27(2), p. 278.

Niederle, M. and Vesterlund, L. (2007). 'Do women shy away from competition? do men compete too much?', *The Quarterly Journal of Economics*, vol. 122(3), pp. 1067–1101.