Appendix

Gender differences in preferences and gender equality are yet unrelated: a replication of Falk and Hermle (*Science*, 2018)

Sara Cerioli and Andrey Formozov

Independent researchers, Hamburg, Germany
Correspondence: sara.cerioli@outlook.com, formozoff@gmail.com

Methods

Overview

We replicate the results using the R programming language version 4.0.3 (2020-10-10), and its open-source IDE RStudio. The following packages with respective versions are used:

Package	Version
data.table	1.13.2
bit64	4.0.5
bit	4.0.4
plyr	1.8.6
dplyr	1.0.7
haven	2.4.1
ggplot2	3.3.2
missMDA	1.18
MASS	7.3-53

Data Collection, Cleaning, and Standardization

Global Preferences Survey and Gallup World Poll datasets

To download the GPS dataset, one can go to the website of the Global Preferences Survey briq - Institute on Behavior & Inequality in the section "downloads". There, choose the "Dataset" form and after filling it, one can download the dataset. Some sociodemographic variables (for instance, education level or income quintile) are not part of the Global Preference Survey, but of the Gallup World Poll dataset that is not openly available. This data is protected by copyright and can't be given to third parties. Check the website of the briq - Institute on Behavior & Inequality for more information on it.

GDP and gender equality indexes

From the website of the World Bank, one can access the data about the GDP per capita on a certain set of years. The data for log GDP p/c calculated in 2005 US dollars was already archived. We used log GDP p/c in 2010 US dollars, instead. To build an estimator for log GDP p/c, we averaged the data from 2003 until

2012 for all the available countries, as done in the original article, and matched the names of the countries with the ones from the GPS dataset.

The Gender Equality Index used in the original article was composed of four main datasets as the first principle component of the PCA of them:

- **WEF Global Gender Gap:** WEF Global Gender Gap Index Taken from the World Economic Forum Global Gender Gap Report 2015. For countries where data was missing, data was added from the World Economic Forum Global Gender Gap Report 2006. Note that we modified some of the country names directly on the csv file, that is why we provide this as an input file in the GitHub repository.

- **UN Gender Inequality Index:** Taken from the Human Development Report 2015. We kept only the table called "Gender Inequality Index".

- **Ratio of female and male labour force participation:** Average International Labour Organization estimates from 2003 to 2012 taken from the World Bank database (http://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS). Values were inverted to create an index of equality. We took the average for the period between 2004 and 2013.

- **Time since women's suffrage:** Taken from the Inter-Parliamentary Union Website. For several countries more than one date where provided (for example, the right to be elected and the right to vote). We use the last date when both vote and stand for election rights were granted, with no other restrictions commented. Some countries were colonies or within a union of the countries (for instance, Kazakhstan in the Soviet Union). For these countries, the rights to vote and be elected might be technically granted two times within a union and as an independent state. In this case, we kept the first date. It was difficult to decide on South Africa because its history shows the racism part very entangled with women's rights (Walker 1990). We kept the latest date when also Black women could vote. For Nigeria, considering the distinctions between North and South, we decided to keep only the North data because, again, it was showing the completeness of the country and it was the last date. Note: USA data doesn't take into account that also up to 1964 black women weren't granted the right to vote (in general, Blacks were not granted that right up to that year). We didn't keep this date, because it was not explicitly mentioned in the original dataset.

In this work we additionally involve the data GDI index:

- **Gender Development Index** taken from Human Development Reports 2020. Note that we have downloaded the two tables of the Human Development Index for males and females, and used the ratio of the two as described in the report.

Missing Data and Imputation

The procedure for imputation and cleaning for each dataset is described in the corresponding section below. We standardized the names of the countries and merged the datasets into one. An additional issue that we faced while trying to reproduce the results of the article has been the missing data: we will treat this specific issue later on because it requires a bit of background.

During the reproduction of the article, we found that the authors didn't describe in detail how they handled missing data in the indicators. They mention on page 14 of the Supplementary Material, that (quoting): "For countries where data were missing data were added from the World Economic Forum Global Gender Gap Report 2006 (http://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf)."

However, regarding the year when women received the right to vote in a specific country, the missing values are the ones coming from the United Arab Emirates and Saudi Arabia, that neither in 2006 (when the WEF Global Gender Gap Report that the authors quote as a reference for the missing values) nor now (in 2022) have guaranteed the right to vote for women yet. There is missing data also in the other sources that the authors quote. So a quick search for the missing countries of the WEF report of 2015, shows us that these countries can't be found in the report of 2006 either. Missing data and imputation, in general, are not crucial for the replication of the analysis, although not desirable. The problem of missing data for a given country

often does not influence much the overall trends of found correlations. However, it becomes very relevant for checking the implications of the study concerning a specific country of interest.

Pure replication and comparison to the Original Article

In this section, we describe how to reproduce the plots and compare the results in terms of z-scores. A z-score threshold value of 2 (two-tailed p-value = 0.0455) is taken as a reference to indicate if one has to reject the hypothesis that two correlation terms are equal. The number of observations for each country-level indicator is shown in the table below, and it is used in the z-score calculation.

Table 2: Number of observations for each country-level indicator. Note that the difference in the GEI number of observations is due to the imputation that we performed in our analysis.

Indicator	Original	Replication	Extended
Log GDP p/c	76	76	76
GEI	71	76	76
WEF GGGI	72	72	72
UNDP GII	75	75	75
F/M LFP	76	76	76
TSWS	76	74	74
UNDP GDI	–	76	76

Reproducing the Plots of the Main Article

To reproduce the plot of Figure 1A, we grouped the countries in quartiles based on the logarithm of their average GDP p/c, extracted the mean of each preference from the gender coefficients (the β_1^c) of the countries for each quartile, after standardizing them. The same method was applied to the GEI in correlation to the gender differences for each economic preference, to reproduce the plot in Figure 1C.

Then, we related the magnitude of the summarised gender difference coefficients (the first component of the PCA) with the logarithm of the average GDP per capita to see the effect of the economic development. This reproduced Figure 1B of the original article. We used a linear model to fit the correlation and extract the p-value, and for the plot the variables on the y-axis were additionally transformed as $(y - y_{min})/(y_{max} - y_{min})$. We applied the same method to extract the correlation between the GEI and the summarised gender preference, to see the effect of gender equality in the countries (Figure 1D). Note that here also the GEI is transformed to be on a scale between 0 and 1.

In Table 3, we compare the results of our replication (using linear regression) and extended (with robust linear regression) analysis to the one from the original paper Falk & Hermle (2018a) (Figure 1B and 1D). Starting with the summarised gender preferences to economic development and gender equality, we see that our analysis brings us to very similar results in terms of correlation coefficients. The p-values are all indicating a statistically significant correlation, as in the original paper, and when calculating the z-scores thanks to Fisher's r to z transformation, we see that each one is below 2 (which is usually taken as a threshold to be statistically significant). This means that our correlations were not statistically significantly different from the ones in the original article. Very similar results appear also from the robust linear regression used in the extended analysis.

Table 3: Correlation between PCA-summarised gender differences in economic preferences vs log GDP p/c, aggregated Gender Equality Index, and Gender Development Index. Significance ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*)

		Log GDP p/c	Gender Equality Index
Original		0.6685***	0.5580***
Replication		0.6830***	0.6079***
	z-score	-0.161	-0.449
Extended		0.6733***	0.5905***
	z-score	-0.053	-0.288

We reproduced the plots in Figure 2A-F using the variable conditioning analysis. This has been done for the economic development, for the GEI, and for each of the four indicators building the GEI. The variable used on the y-axis is the first Principal Component of the PCA made on the gender differences on the six preferences. All the variables used have been standardized to have a mean at 0 and a standard deviation of 1 before applying the conditional analysis. Using the residuals, we performed a linear regression on the data points, and we extracted correlation coefficients and p-values. In Table 4, we compare the results obtained from our replication analysis and the extended analysis to the results found in Falk & Hermle (2018a), Figure 2A-F.

Table 4: Comparison of the conditional analysis results for original, replicated and extended study. Reported are the **slopes** of the linear regressions and the corresponding p-value. Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

Variable	Residualized on	Original	Replication	Extended
Log GDP p/c	Gender Equality Index	0.5258***	0.5003***	0.4862***
Gender Equality Index	Log GDP p/c	0.3192**	0.3358***	0.3432**
WEF Global Gender Gap	Log GDP p/c	0.2327**	0.2234*	0.2106*
UN Gender Equality Index	Log GDP p/c	0.2911	0.3180	0.3017
F/M in Labor Force Participation	Log GDP p/c	0.2453*	0.2206*	0.2034*
Time since Women Suffrage	Log GDP p/c	0.2988**	0.1879*	0.1929*

Most of the time, we find very similar slope coefficient and statistical significance between the variables of our study (both replicated and extended) and the results from the original study, with only two exceptions: One is related to the indicator WEF Global Gender Gap, and the other to the time since women suffrage. Regarding the last one, the differences can be explained by the imputation of the data used. The dataset from which this indicator is coming from, as a matter of fact, can be interpreted non unanimously (see the paragraph "Gender Equality Index" in the Appendix and discussion sections) and we can't draw any conclusion from it. Regarding the WEF Global Gender Gap Index, the difference of the p-value significance is of factor 2, and it is not clear why such difference is exhibited. The coefficients are anyway in good agreement with each other.

In Table 5, we summarise the results of the correlation of single preferences to the economic development of the countries, to be compared to Figure S2 in Falk & Hermle (2018b), while in Table 6, we show the results for the single preferences gender differences regressed on the Gender Equality Index, as done in Figure S3 in Falk & Hermle (2018b).

Table 5: Correlation coefficients for country-level gender differences in economic preferences vs log GDP p/c obtained in the original article, present replication and extended analysis. Significance $\leq$ 0.001 (***), $\leq$ 0.01 (**), $\leq$ 0.05 (*)

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.5918***	0.5847***	0.3086**	0.3542**	0.3685**	0.3837***
Replication		0.5761***	0.5882***	0.3216**	0.3683**	0.3649**	0.4415***
	z-score	0.144	-0.032	-0.087	-0.098	0.025	-0.421
Extended		0.5926***	0.5488***	0.3240**	0.3929***	0.3579**	0.4253***
	z-score	-0.007	0.32	-0.103	-0.272	0.074	-0.301

Table 6: Comparison of the correlations between Gender Equality Index and country-level gender differences in economic preferences. The correlations are calculated using the OLS method. Significance $\leq$ 0.001 (***), $\leq$ 0.01 (**), $\leq$ 0.05 (*)

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.4050***	0.5073***	0.1280	0.4035***	0.3412**	0.4257***
Replication		0.4829***	0.5030***	0.2132	0.3751***	0.3441**	0.4415***
	z-score	-0.576	0.034	-0.521	0.199	-0.019	-0.115
Extended		0.4921***	0.4686***	0.2156*	0.3714**	0.3414***	0.4434***
	z-score	-0.648	0.301	-0.536	0.224	-0.001	-0.129

For the comparison of the results from Figure S4 in Falk & Hermle (2018b), we refer to Table 7, showing the correlation between the average gender differences to the single-gender equality indicators.

Table 7: Single indicators for the gender equality at the country level correlated with gender differences. Significance $\leq$ 0.001 (***), $\leq$ 0.01 (**), $\leq$ 0.05 (*)

		WEF GGGI	UNDP GII	F/M LFP	Time Since Women Suffrage
Original		0.4097***	0.6482***	0.2661*	0.5139***
Replication		0.4075***	0.6680***	0.2860*	0.4517***
	z-score	0.016	-0.210	-0.130	0.487
Extended		0.3937**	0.6586***	0.2645*	0.4535***
	z-score	0.112	-0.109	0.01	0.473

For the comparison of the results of the Figure S5 and S6 of Falk & Hermle (2018b) to ours, refer to Table 8 and Table 9.

Table 8: Log GDP p/c conditional on Gender Equality Index, regressed on each single economic preference gender difference coefficient. Here reported are the correlation terms for the original analysis, the replication and the extended analysis. Significance levels: $\leq$ 0.001 (***), $\leq$ 0.01 (**), $\leq$ 0.05 (*).

	Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original	0.4574***	0.4751***	0.2771*	0.2444*	0.2868*	0.2621*

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Replication		0.4265***	0.4338***	0.2509	0.2111	0.2256*	0.2288*
	z-score	0.228	0.309	0.167	0.208	0.389	0.210
Extended		0.4450***	0.3959***	0.2524	0.2451*	0.2191*	0.2444*
	z-score	0.092	0.581	0.158	-0.0044	0.429	0.112

Table 9: Gender Equality Index, conditional on log GDP p/c, regressed on each single economic preference gender difference coefficient. Here reported are the correlation terms for the original analysis, the replication and the extended analysis. Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.2050	0.3304**	-0.0115	0.2788*	0.1973	0.2967*
Replication		0.2472*	0.2696*	0.0481	0.2240*	0.1863*	0.2841*
	z-score	-0.264	0.397	-0.354	0.347	0.068	0.082
Extended		0.2521*	0.2401	0.051	0.2037*	0.1868*	0.2781*
	z-score	-0.295	0.584	-0.371	0.473	0.065	0.120

Lastly, we have reproduced the results from Figures S8 and S9 of Falk & Hermle (2018b) in Tables 10 and 11. Here we report only the results of the replication analysis because to our extended analysis this was seen as useless.

Table 10: Preferences standardized at global level for log GDP p/c.
Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.5787***	0.5505**	0.2819*	0.2980**	0.2974**	0.4391***
Replication		0.5761***	0.5882***	0.3216**	0.3683**	0.3649**	0.4124***
	z-score	0.024	-0.337	-0.264	-0.478	-0.458	0.197

Table 11: Gender differences and economic development by preference and country without controls. Significance levels: ≤ 0.001 (***), ≤ 0.01 (**), ≤ 0.05 (*).

		Trust	Altruism	Pos. Rec.	Neg. Rec.	Risk Tak.	Patience
Original		0.5434***	0.5808***	0.2748*	0.4038***	0.3860***	0.4830***
Replication		0.5462***	0.5881***	0.2826*	0.3864***	0.3886***	0.4804***
	z-score	-0.024	-0.067	-0.051	0.125	-0.018	0.020

Further notes on the replication

For the replication of the Figure S7 of Falk & Hermle (2018b), we couldn't replicate it because we had no access to such granularity of the data.

For the replication of the tables in the supplementary material, we were not able to get the datasets as indicated from the sources.

References

Falk, A. & Hermle, J. (2018a). Relationship of gender differences in preferences to economic development and gender equality. *Science*, **362**.

Falk, A. & Hermle, J. (2018b). Supplementary materials of relationship of gender differences in preferences to economic development and gender equality.

Walker, C. (1990). Women and gender in southern africa to 1945. Retrieved from https://www.sahistory.org.za/archive/womens-suffrage-movement-politics-gender-race-and-class-cheryl-walker