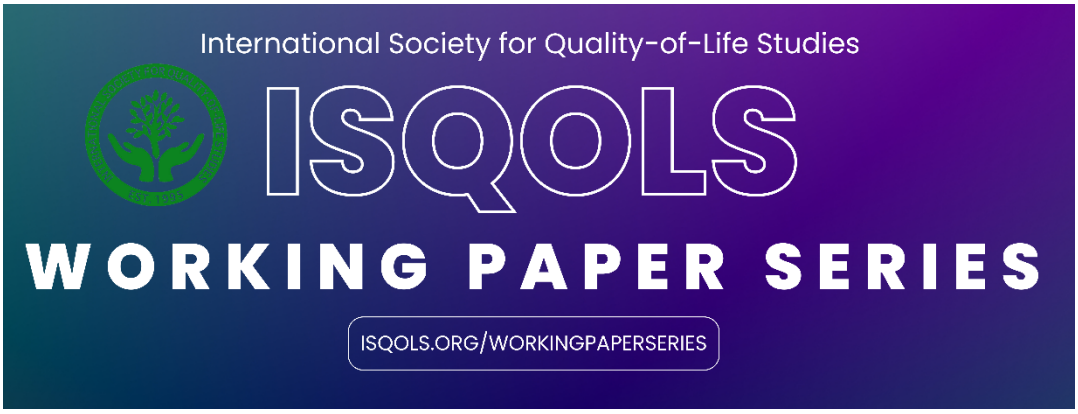


International Society for Quality-of-Life Studies
isqols.org
office@isqols.org

Working Paper No. 3

# Measuring the Gender Inequality of Education: An Information Theoretic Approach

by Georg P. Mueller

February 2024

# Measuring the Gender Inequality of Education: An Information Theoretic Approach

Georg P. Mueller
University of Fribourg (Switzerland)
Faculty of Economics and Social Sciences

Georg.Mueller_Unifr@bluewin.ch

**Abstract:** This article deals with entropy, which is a double-faced concept. On the one hand, it means uncertainty, i.e. *lack of information:* the higher the entropy of a social privilege, the lower the information about particular privilege holders. By comparing the entropy of a whole population with the conditional entropies of its subpopulations, it is possible to calculate the information contained in the membership in these subgroups. On the other hand, entropy also means variation, i.e. *inequality:* the higher the entropy of a privilege distribution, the higher the inequality among the privilege holders. Thus, it is possible to calculate the *overall inequality* of a population as well as the *particular inequalities* within its subpopulations.

The author attempts to unite the two different faces of entropy. By mathematical reasoning it is possible to show that the *total inequality* of a privilege distribution is a weighted mean of the inequalities within its subpopulations plus a weighted mean of the information contained in each of these subpopulations. The first weighted sum measures *inter-individual inequality.* The second is called *synentropy* (mutual information) and describes the *inequality of opportunities* between the subpopulations.

In order to illustrate the use of the afore-mentioned entropy-concepts, the author analyses the educational attainments in Switzerland, Turkey, and Sweden according to gender and birth-cohorts. The data source used for this purpose is the European Values Study, which allows to compare the mentioned forms of inequality between countries and different historical periods. As a major result, the inequality of educational opportunities between men and women is relatively small and decreases over time.

**Keywords:** Inequality, social change, gender, education, international comparisons, information theory, entropy.

## 1. Introduction and overview

This paper deals with the educational attainments of men and women and the related inequalities within and between the two gender groups. The discrimination of women in relation to the education of men is considered to depend on the role of women in family and society, which in turn depends on the birth cohort and national culture (Cin, 2017: 68-69). Consequently, in this article we want to compare the educational attainments of men and women at different times and in different countries. In particular we compare the situation in Switzerland with the situation in Turkey and Sweden. The latter country is known for being "advanced" with regard to gender questions, whereas Turkey is considered to be "delayed" regarding the reduction of gender inequalities.

In order to tackle this research question, the author uses concepts from information theory. As compared to other indicator systems this approach has the advantage that it integrates rather diverse theoretical ideas. *Entropy* e.g. is not only a measure of *information* but also a measure of *inequality* (Coulter, 1989: chap. 5; Hao & Naiman, 2010: 37 ff.; Mueller, 2004, 2017, 2021). Similarly, the total inequality of a privilege distribution can be represented as the population-weighted mean of the *intra-group inequalities* within its subpopulations plus the *inequality of opportunities* between these subpopulations. The second component is called *synentropy* (mutual information) and describes the mean information gains from the membership in the mentioned subpopulations.

By the aggregation of individual interview data from the European Values Study (Gesis, 2023) it is possible to get longitudinal data about gender-specific educational attainments in Switzerland, Turkey, and Sweden. As a major result of the entropy analyses, it turns out that the inequality of educational opportunities between men and women is relatively small and tends to decrease over time. Moreover, as expected, this kind of structural inequality was in Turkey nearly always higher than in Sweden and Switzerland.

## 2. A system of entropy-based indicators [1)]

In their famous book *The Mathematical Theory of Communication*, Shannon and Weaver (1962) defined the *entropy* of a random variable X with a finite number of possible values $x_1$, $x_2$, ... , $x_n$ and associated probabilities prob(X=$x_i$), prob(X=$x_2$),..., prob(X=$x_n$) by the formula

$$H(X) = - \Sigma_{i = 1,..,n} [prob(X=x_i) * \log_2(prob(X=x_i))] \quad (1)$$

It can be shown that H(X) is the expected number of binary decisions in order to identify the value of a random-element of X (Stone, 2015: chap. 1). If all elements of X have the same value, then the entropy H(X) = 0. This means no inequality and no uncertainty but also a *maximum* of information about the value of a randomly selected element of X. If X has a *rectangular* distribution and all elements of X have the same probability 1/n, then H(X) = $\log_2(n)$ (Theil 1967: 26). This is the highest possible entropy, corresponding to a maximum of uncertainty and a minimum of information about a

[1)] For readers' convenience there is at the end of the article a glossary of mathematical terms.

randomly chosen element of X. The associated *rectangular distribution* of X is in this case highly unequal. In sum, the entropy H(X) is a double-faced concept: it represents inequality as well as uncertainty, i.e. lack of information.

The entropy concept can also be applied to a subgroup with the property $Y = y_j$ by replacing in formula (1) the global by the conditional probabilities $prob(X=x_i \mid Y=y_j)$. This way, the *conditional entropy* $H(X|Y=y_j)$ of the related subgroup $y_j$ is equal to

$$H(X|Y=y_j) = - \Sigma_{i = 1,..,n} [prob(X=x_i|Y=y_j) * \log_2(prob(X=x_i|Y=y_j))] \quad (2)$$

From the sociological point of view $H(X|Y=y_j)$ is the inequality within the group $y_j$ that is primarily due to the intra-group competition for the privilege X. It has nothing to do with a possible social discrimination of the group $y_j$ since there is no direct comparison with other groups. Thus, $H(X|Y=y_j)$ measures rather the strength of *meritocratic* principles within the group $y_j$. If one is interested in the effects of the meritocratic principle on a *whole* society, one has to calculate the *weighted mean I(X) of all intra-group inequalities,* which is represented by the following formula that considers the size of the different groups by the weights $prob(Y=y_j)$:

$$I(X) = \Sigma_{j = 1,..,m} prob(Y=y_j) * H(X|Y=y_j) \quad (3)$$

The transition from the unconditional to the conditional entropy $H(X|Y=y_j)$ corresponds to the gain or loss of information about the privilege X, if we know about the group membership $Y=y_j$ of a randomly selected social actor. This positive or negative *information gain* from a group membership is defined by the difference between the general and the group-specific entropy that is given by the equation

$$G(X|Y=y_j) = H(X) - H(X|Y=y_j) \quad (4)$$

If $G(X|Y=y_j) > 0$, the knowledge of a membership in group $y_j$ yields a reduction of uncertainty or a gain of information. The group membership, e.g. of belonging to the male or female gender, becomes a *status sign* with regard to X. In the particular case of a *low* group-specific status X, this status sign can even be some kind of a social *stigma* (Goffman, 1986: chap. 2). If $G(X|Y=y_j) < 0$, the knowledge of a membership in group $y_j$ results in an increase of uncertainty, i.e. a loss of information about the group-specific value of X. In this case the group membership $Y=y_j$ is a *status mask* with regard to the privilege X.

It is easily possible to calculate the *weighted mean information gain* for all analyzed groups $y_{j = 1,..,m}$. The result is called *synentropy S* (or *mutual information)* (Mathar, 1996: 28) and is represented by the following formula that considers the size of the different groups by the weights $prob(Y=y_j)$:

$$S(X|Y) = \Sigma_{j = 1,..,m} [prob(Y=y_j) * G(X|Y=y_j)] \quad (5)$$

The higher the synentropy S(X|Y), the more important is the group membership Y for the participation in the good X. Thus, S(X|Y) is a measure of the *inter-group inequality of opportunities* due to the membership variable Y. It can be shown, that for mathematical reasons $S(X|Y) \geq 0$ (see Mathar, 1996: 33). For many types of *ascribed* membership Y like gender or race it is socially desirable that Y should have *no* influence on the distribution of the privilege X and consequently it is expected that $S(X|Y) = 0$.

By simple mathematical calculations it is possible to integrate the previously encountered types of inequality in the following formula: [2)]

$$H(X) = S(X|Y) + I(X) \quad (6)$$

Thus, the total inequality H(X) is the sum of the inter-group inequality of opportunities S(X|Y) and the mean intra-group inequality I(X).[3)]

Tab. 1 describes the decomposition (6) for three exemplary distributions, referring to three levels of privilege $x_1$, $x_2$, and $x_3$ and two analyzed groups $y_1$ and $y_2$. In *scenario 1* both groups have the same rectangular distribution. The total inequality H(X) is at the highest possible level $\log_2(3) = 1.585$. As both groups have the same privilege distribution there are *zero* information gains $G(X|Y=y_1)$ and $G(X|Y=y_2)$ and the inter-group inequality of opportunities S(X|Y) = 0. Consequently, the whole inequality H(X) = 1.585 is exclusively influenced by the mean intra-group inequality I(X) = 1.585.

Tab. 1: Three exemplary absolute frequency distributions and the resulting decomposition of inequalities.

| Level of privilege: | Scenario 1 Group $y_1$ | Scenario 1 Group $y_2$ | Scenario 2 Group $y_1$ | Scenario 2 Group $y_2$ | Scenario 3 Group $y_1$ | Scenario 3 Group $y_2$ |
|---|---|---|---|---|---|---|
| $X = x_1 = 1$ | 33 | 33 | 97 | 33 | 97 | 1 |
| $X = x_2 = 2$ | 33 | 33 | 1 | 33 | 1 | 1 |
| $X = x_3 = 3$ | 33 | 33 | 1 | 33 | 1 | 97 |
| N of all X | 99 | 99 | 99 | 99 | 99 | 99 |
| Info-gain $G(X\|Y=y_1)$ | 0.000 | | +1.109 | | +0.909 | |
| Info-gain $G(X\|Y=y_2)$ | 0.000 | | - 0.313 | | +0.909 | |
| Inter-gr. ineq. S(X\|Y) | 0.000 | | 0.398 | | 0.909 | |
| Intra-gr. ineq. I(X) | 1.585 | | 0.874 | | 0.163 | |
| Total ineq. H(X) | 1.585 | | 1.272 | | 1.071 | |

*Scenario 2* differs from the previous scenario 1 by a changed privilege distribution of group $y_1$, which now focusses on the lowest privilege level $x_1$. The result is a positive information gain $G(X|Y=y_1) = 1.109$ from the membership in group $y_1$, which is a social *stigma* due to the dominance of the low status $x_1=1$. As a consequence, there is an increased inter-group inequality of opportunities S(X|Y) = 0.398, pointing to the discrimination of group $y_1$. The privileges of the other group $y_2$ have a much broader

[2)] From the equations (3) to (5) follows: **S(X|Y)** $= \Sigma_{j=1,..,m} [prob(Y=y_j) * G(X \mid Y=y_j)] =$
$= \Sigma_{j=1,..,m} [prob(Y=y_j) * (H(X) - H(X \mid Y=y_j))] = 1 * H(X) - \Sigma_{j=1,..,m} [prob(Y=y_j) * H(X \mid Y=y_j)] =$
$=$ **H(X) - I(X)** --> **S(X|Y) + I(X) = H(X)** (see equation (6)).

[3)] This is *not* the famous entropy decomposition of Theil (1972: chap. 1.5), which refers to univariate and not to bivariate distributions like equation (6).

distribution than those of group $y_1$. This relatively high uncertainty results in a *negative* information gain $G(X|Y=y_2) = -0.313$ such that the membership in group $y_2$ becomes a *status mask.*

In *scenario 3* the inter-group inequality $S(X|Y) = 0.909$ increases further, as compared to the previously discussed scenario 2. This is due to the fact that group $y_2$ is in scenario 3 very privileged and nearly always attains the privilege level $x_3 = 3$. Consequently, its information gain $G(X|Y=y_2) = 0.909$ is a *positive status sign.* Due to the nearly bi-polar privilege distribution of the joint group populations, the total inequality $H(X) = 1.070$ is smaller than in scenario 1 and 2. The difference between H(X) and the rather high inequality of opportunities S(X|Y) is according to equation (6) equal to the mean intra-group inequality $I(X) = 0.163$, which is obviously smaller than in scenario 1 and 2.

## 3. Educational inequalities among and between men and women

### 3.1 The data

Comparisons between countries require standardized data. If in addition there is an interest in different birth cohorts, individual interviews collected in international surveys are a relatively good data source. Thus we used for the present analyses the European Values Study 2008, published by Gesis (2023) as dataset ZA4800. In particular, we analyzed for *Sweden* (Breen & Jonsson, 2020: 70-71), *Turkey* (Cin, 2017: 60-70), and *Switzerland* (Falcon, 2020: 151-153) the following variables:

a) *Educational attainment* (V336, recoded): primary, secondary, tertiary education.

b) *Gender* (V302): male, female.

c) *Year of birth* (V303, recoded): 1930-39, 1940-49, 1950-59, 1960-69, 1970-79.[4)]

*Switzerland* was selected as the domicile of the author. It was compared with two contrasting societies: *Sweden* as a more advanced and *Turkey* as a less advanced country with regard to gender questions.

### 3.2 The different forms of educational inequality

According to equation (6), total inequality is the *sum* of two components:

i) The population-weighted mean of the *intra-group inequalities* within the male and the female subpopulation. This is generally the result of meritocratic competition *within* the two gender-groups.

ii) The inter-group *inequality of opportunities* between the male and the female subpopulation. This is an indicator of possible gender-based discrimination.

Thus, the above mentioned forms of inequality are not totally independent. However, there are enough degrees of freedom such that they may have developed with different dynamics, which will be analyzed and displayed in the next paragraphs.

[4)] In order to ensure that at the moment of the interview (2008) all persons were old enough to have finished a tertiary education we omitted respondents born after 1979. In addition, by omitting persons born before 1930, no one was at time of the interview older than 78.

Figure 1 displays the dynamics of the total inequality of educational attainments. The educational inequality of Switzerland and Sweden are rather similar and systematically decrease over time. Initially in the 1930s there was in both countries a heterogeneous mix of different educational attainments. Over the years, the distribution became more and more equal, probably due to the growing importance of tertiary education and the disappearance of people with only primary education. The situation in Turkey is quite the reverse: in the 1930s nearly everyone finished the educational curriculum at the primary level. Consequently, educational inequality was very low. Over the decades, more and more pupils got the opportunity to acquire also secondary or tertiary education and the total inequality increased. If Turkey follows the

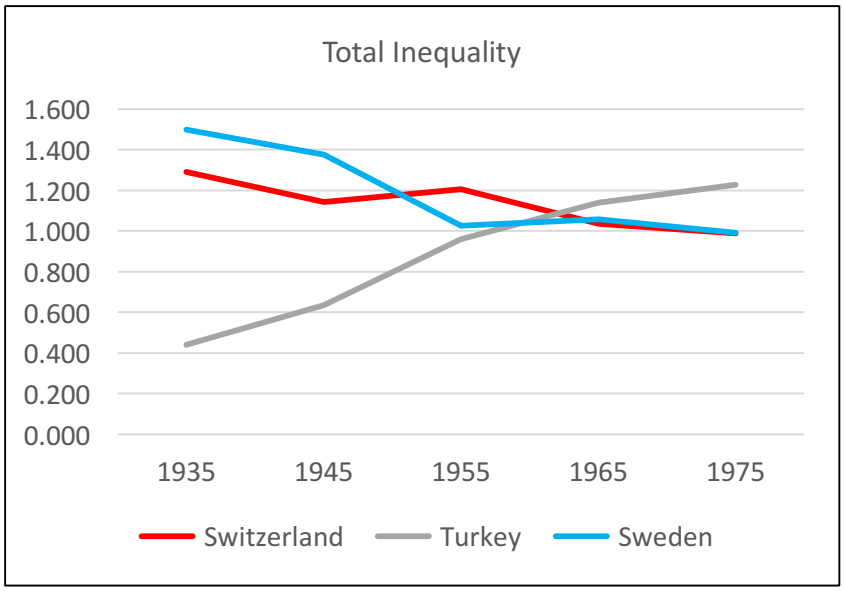


*Fig. 1: Total inequality by country and birth cohort.* [5)]

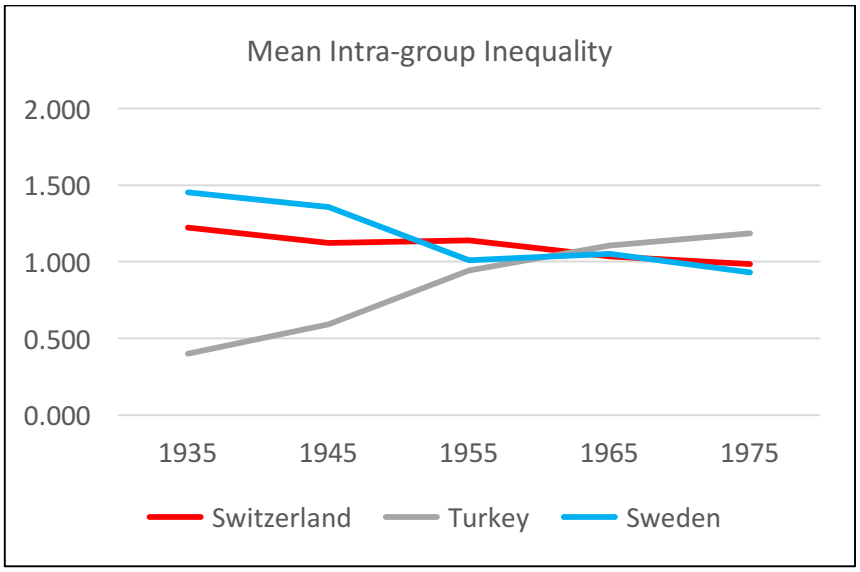


*Fig. 2: Mean intra-group inequality by country and birth cohort.*

[5)] The years y on the horizontal axis refer to births in the time-interval [ y-5 , y+5 [.

development path of Sweden and Switzerland, its educational system will in the *future* become more and more tertiary and the inequality will drop. In sum, the relation between the development of education and its inequality is an inverse U-function, as proposed by Lenski (1984: 437). Thus, Fig. 1 displays for Turkey the *first* and for Sweden and Switzerland the *second* phase of this relation.

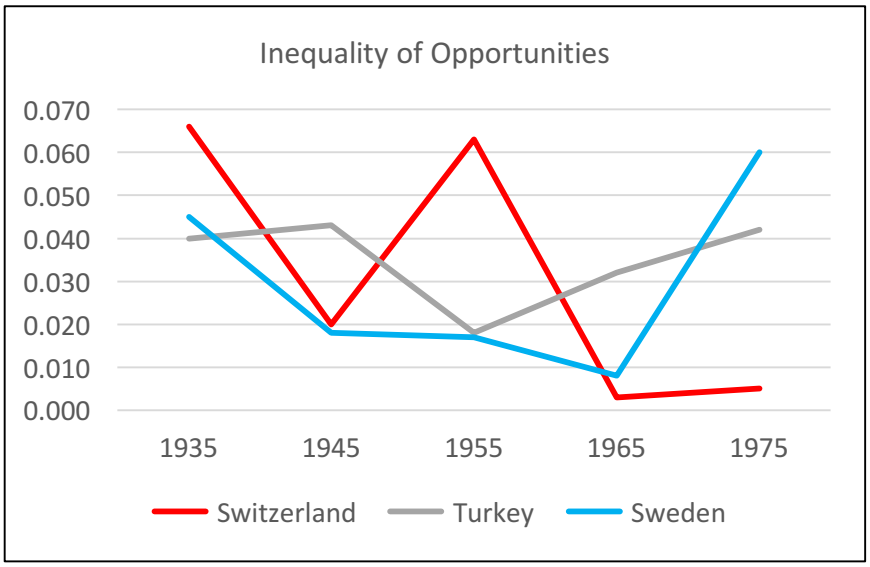


*Fig. 3a: Inequality of opportunities by country and birth cohort: Absolute values.*

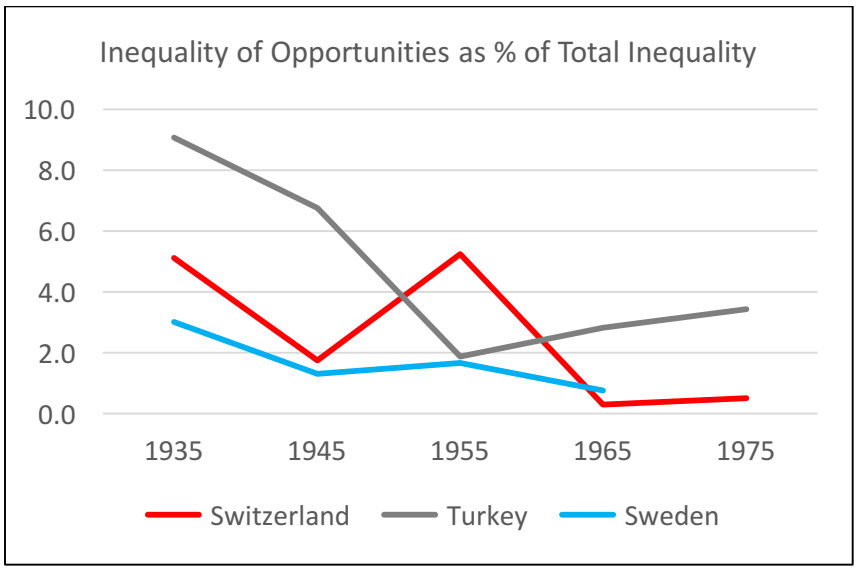


*Fig. 3b: Inequality of opportunities by country and birth cohort: Relative values.* [6)]

[6)] Excluding data-point 1975 of Sweden, due to absence of persons with only primary education: see section *Raw data* at the end of the paper.

Figure 2 is very similar to Fig. 1, in qualitative as well as in quantitative terms. So we suppose that the *intra-group inequalities* within the male and the female subpopulation are important drivers of the *total inequality* of education. They are probably the result of inter-individual competition, which is typical for meritocratic societies (Arrow et al., 2000). The dominance of this factor holds (more or less) for the whole analyzed period from the 1930s to the 1970s. As a consequence, the *inequality of opportunities* between the male and the female subpopulation should be of minor importance. The respective Fig. 3a seems to confirm this assumption: here, the highest values (≈ 0.07) are far below the lowest values (≈ 0.45) observed for total inequality (see Fig. 1).

Apart from the generally low *absolute* inter-group inequality of opportunities, Fig. 3a does not display regularities that can easily be interpreted. However, this situation improves if one relates in Fig. 3b the inequality of opportunities to total inequality such that it corresponds to Theil's U (Wikipedia, 2023) and omits some technically problematic Swedish data (see footnote 6). As expected there is a secular decline of structural inequality that discriminates women. Sweden has the smallest inequality of educational opportunities and Turkey the highest. However, even for the latter country, the respective values are relatively small and point to a relatively universalistic access to education.

### 3.3 Information gains from gender about education

In Turkey and Switzerland information gains from *female* gender are always *positive* and from *male* gender predominantly *negative* (see Figs 4a,b). This is probably the result of the secular expansion of the educational systems. It created new educational opportunities in the secondary and tertiary sector, which were sized by some but not all men, whereas women continued to stop their educational careers at the traditional levels, i.e. primary in Turkey and secondary in Switzerland. As a result, the behavior of women is more standardized and predictable than the behavior of men that displays

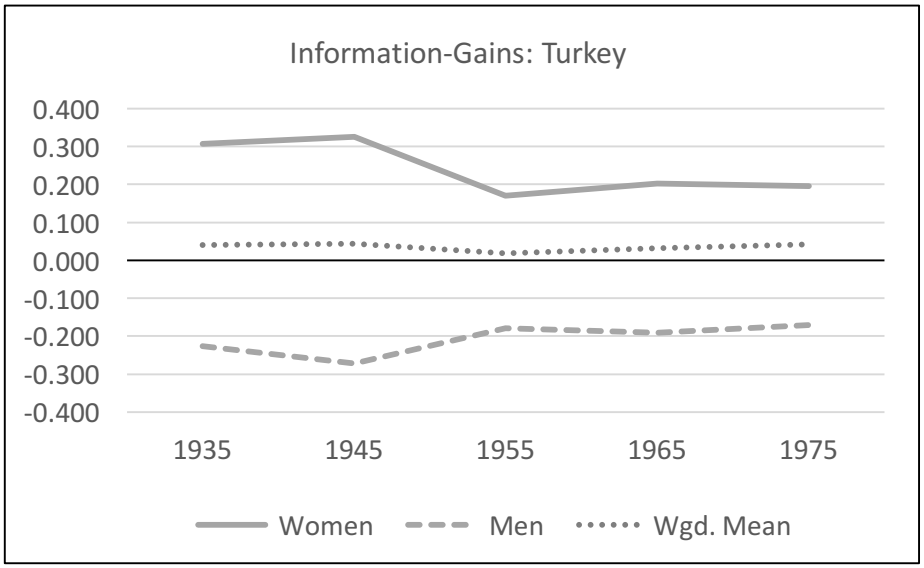


*Fig. 4a: Gains from male and female gender in Turkey.*

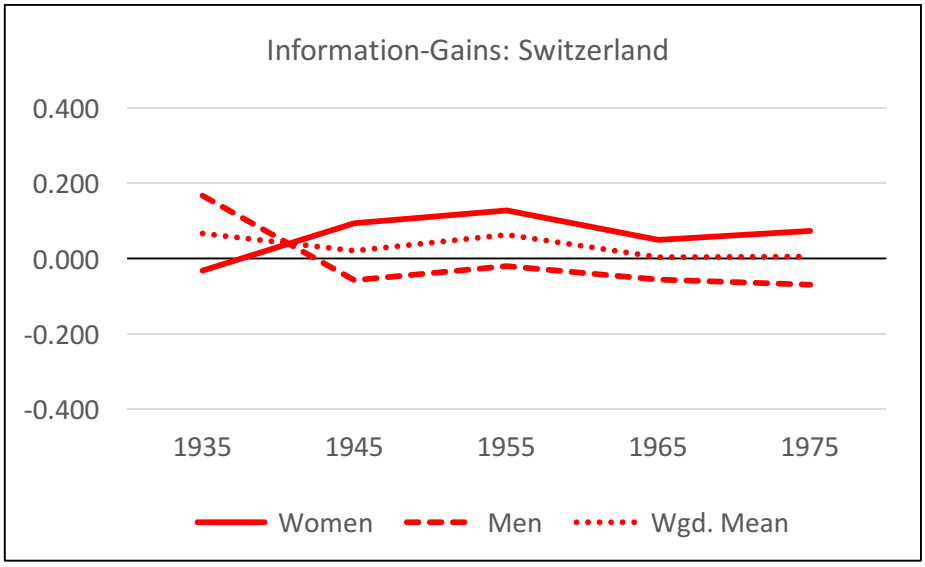


*Fig. 4b: Gains from male and female gender in Switzerland.*

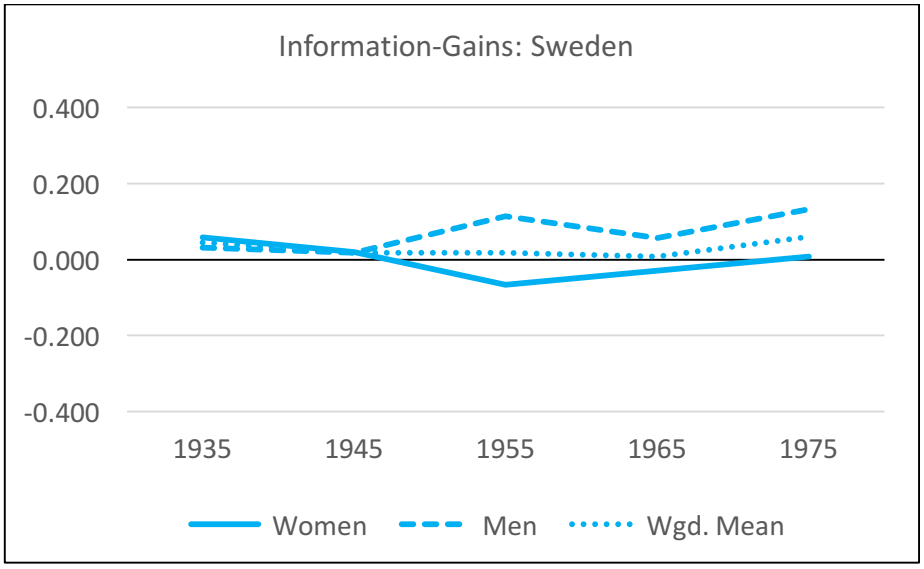


*Fig. 4c: Gains from male and female gender in Sweden.*

more variation. Thus, with regard to educational attainments gender is for men a *status mask,* whereas for women it is a (negative) *status sign.* The related mix of positive and negative gains from the knowledge of gender gives in Figs 4a-c only small weighted mean values (Wgd. Mean). They are identical with the inter-group *inequality of opportunities* and correspond to the unexpectedly small values, which we encountered in Fig. 3a.

In Sweden, the information gains are just the reverse of Turkey and Switzerland. As Fig. 4c demonstrates, the respective gains are for *women* always positive and slightly higher than for men. The Swedish university system expanded between the 1930s and the 1980s like in Switzerland and Turkey. In Sweden, however, the new opportunities in the tertiary educational sector were mainly taken by *women,* whereas

men were more traditional and stopped education much earlier.[7] Thus, in terms of information gains men are more predictable. The reason for women's early "race" to higher education may have to do with an early start of gender policies and/or particularities of the Swedish labor market for women.

## 4. Summary and outlook

In this paper we wanted to present an entropy-based system of indicators for measuring different forms of social inequalities. For this purpose, we made use of the double nature of entropy, which means statistical diversity as well as lack of information. The resulting indicators are highly interdependent due to their mathematical definitions: according to equation (6) e.g., total inequality is the sum of intra-group and inter-group inequalities, where the latter concept can be partitioned into the information contained in the two gender groups (see equation (5)). The application of the proposed indicators to Switzerland, Turkey, and Sweden have led to plausible results, however with the exception of the inequality of opportunities: the original Fig. 3a was hard to interpret and required a standardization by the total inequality (see Fig. 3b). If needed, this standardization can also be applied to most other concepts without losing the advantage of their mathematical coherence. The expected advantage of facilitating the interpretation of empirical results has of course to be confirmed by enlarging the set of analyzed countries.

## References

Arrow, K., et al. (2000). Introduction. In Arrow, K., S. Bowles & S. Durlauf (Eds.). *Meritocracy and Economic Inequality* (ix–xv). Princeton: Princeton University Press.

Breen, R. & Jonsson, J. O. (2020). Sweden, the Middle Way? Trends and Patterns in Social Mobility and Educational Inequality. In Breen, R. & W. Müller (Eds.). *Education and Intergenerational Inequality in Europe and the United States* (chap. 4). Stanford: Stanford University Press.

Cin, F. M. (2017). *Gender Justice, Education and Equality.* Cham: Palgrave MacMillan.

Coulter, Ph. B. (1989). *Measuring Inequality: A Methodological Handbook.* Boulder: Westview Press.

Falcon, J. (2020). The Swiss El Dorado? Education and Social Mobility in Twentieth-Century Switzerland. In Breen, R. & W. Müller (Eds.). *Education and Intergenerational Inequality in Europe and the United States* (chap. 7). Stanford: Stanford University Press.

Gesis (2023). *ZA4800: European Values Study 2008: Integrated Dataset.* (https://search.gesis.org/research_data/ZA4800). Cologne: Gesis.

Goffmann, E. (1986). *Stigma.* New York: Simon & Schuster.

Hao, L. & Naiman, D. Q. (2010). *Assessing Inequality.* Thousand Oaks: Sage Publications.

[7] See section *Raw data* at the end of the paper.

Lenski, G. (1984). *Power and Privilege: A Theory of Social Stratification.* Chapel Hill: University of North Carolina Press.

Mathar, R. (1996). *Informationstheorie (Information Theory)*. Stuttgart: Teubner.

Mueller, G. P. (2004). Entropy Measures for Comparative Analyses of Social Justice and Discrimination. In Grözinger, G. & A. van Aaken (Eds.). *Inequality: New Analytical Approaches* (pp. 15–31). Marburg: Metropolis-Verlag.

Mueller, G. P. (2017). Gender Inequality Under Different National Welfare Regimes: An Empirical Evaluation with Entropy Measures from Information Theory. *The Senshu Social Well-being Review 4,* 39–51.

Mueller, G. P. (2021). Entropy measures of social mobility: The example of the intergenerational transmission of education. *FORS Working Paper Series*, paper 2021-3. Lausanne: FORS. DOI: 10.24440/FWP-2021-03.

Shannon, C. & Weaver, W. (1962, [1949]). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Stone, J. V. (2015). *Information Theory.* Sheffield: Sebtel Press.

Theil, H. (1967). *Economics and Information Theory.* Amsterdam: North-Holland Publishing Company.

Theil, H. (1972). *Statistical Decomposition Analysis.* Amsterdam: North-Holland Publishing Company.

Wikipedia (2023). *Uncertainty Coefficient (https://en.wikipedia.org/wiki/Uncertainty_coefficient)*. Accessed on August, 25, 2023.

**Glossary of mathematical terms**

| | |
|---|---|
| $G(X\|Y=y_j)$: | Information gain about the privilege X from knowledge of the group membership $Y=y_j$. |
| H(X): | Total inequality (entropy) of the privilege X. |
| $H(X\|Y=y_j)$: | Conditional entropy of the privilege X for the group $Y=y_j$. |
| I(X): | Mean intra-group inequality with regard to the privilege X. |
| $\log_2(x)$: | Binary logarithm of value x. |
| m: | Number of population categories of the group Y. |
| n: | Number of status categories of the privilege X. |
| prob(E): | Probability of the event E. |
| prob(E\|C=c): | Conditional probability of the event E, if C=c. |
| S(X,Y): | Inter-group inequality of opportunities (synentropy) between the groups $y_j$ of Y, with regard to the privilege X. |
| X: | Privilege variable. |
| Y: | Group variable. |
| $\Sigma_{i=1,..,n}(x_i)$: | Sum $x_1 + x_2 + ... + x_n$ |
| [a,b[: | Open interval, including the lower limit a, but excluding the upper limit b. |

**Raw data: Number of persons by highest educational attainments**

**Turkey**

| Year of birth: | Primary educ. Men | Primary educ. Women | Secondary educ. Men | Secondary educ. Women | Tertiary educ. Men | Tertiary educ. Women |
|---|---|---|---|---|---|---|
| 1930-39 | 47 | 53 | 5 | 1 | 2 | 0 |
| 1940-49 | 75 | 100 | 13 | 4 | 6 | 1 |
| 1950-59 | 89 | 137 | 27 | 19 | 11 | 8 |
| 1960-69 | 111 | 193 | 59 | 46 | 21 | 11 |
| 1970-79 | 123 | 253 | 96 | 76 | 30 | 18 |

**Sweden**

| Year of birth: | Primary educ. Men | Primary educ. Women | Secondary educ. Men | Secondary educ. Women | Tertiary educ. Men | Tertiary educ. Women |
|---|---|---|---|---|---|---|
| 1930-39 | 20 | 9 | 23 | 27 | 8 | 14 |
| 1940-49 | 23 | 13 | 77 | 64 | 27 | 40 |
| 1950-59 | 0 | 2 | 72 | 68 | 35 | 53 |
| 1960-69 | 1 | 2 | 66 | 74 | 35 | 60 |
| 1970-79 | 0 | 0 | 53 | 42 | 21 | 56 |

**Switzerland**

| Year of birth: | Primary educ. Men | Primary educ. Women | Secondary educ. Men | Secondary educ. Women | Tertiary educ. Men | Tertiary educ. Women |
|---|---|---|---|---|---|---|
| 1930-39 | 2 | 12 | 37 | 39 | 21 | 11 |
| 1940-49 | 7 | 10 | 65 | 81 | 30 | 17 |
| 1950-59 | 4 | 10 | 43 | 81 | 37 | 19 |
| 1960-69 | 4 | 3 | 77 | 105 | 37 | 42 |
| 1970-79 | 3 | 2 | 70 | 86 | 32 | 28 |