Centre de recherche en économie et statistique

Série des Documents de Travail

n° 2016-38

Travelling with Albert Gifi Comparing Nominal, Ordinal and Interval Approaches in Comparative Studies of Social and Cultural Spaces

D.Joye[1]
G-E Birkelund[2]
Y.Lemel[3]

Travelling with Albert Gifi
Comparing Nominal, Ordinal and Interval Approaches in Comparative Studies of Social and Cultural Spaces

Dominique Joye Gunn Elisabeth Birkelund Yannick Lemel

2016-08-02

Earlier versions of this paper were presented at different conferences, and we thank participants and discussants for their comments. We would also like to thank Jan de Leeuw, Julie Falcon, John Eriksen and Amal Tawfik for discussions and comments. The content of this paper is, however, our own responsibility.

Abstract

In this paper we re-introduce the Gifi-framework which allows us to compare different methodological designs applied to the same data. Our empirical case is based on Pierre Bourdieu's homology hypothesis, developed in "La Distinction", postulating a correspondence between a "social space" (defined by variables describing economic and cultural capital) and a "cultural space" (defined by variables describing cultural preferences and tastes).Using data from three countries, France, as in "La Distinction," and Norway and Switzerland as contrasting countries, we evaluate the strength of the relationship between social and cultural spaces.. Substantially, we find that the structures in the three countries share a lot of similarities. Methodologically, we test the impact of the level of measurement attributed to the variables, from interval to nominal, to see which measurement properties are the most important in the models used. We also question the established research praxis on the direction of the analyses of the relationship between cultural and social spaces: Instead of regarding one of these spaces as the "target" – we prefer, in line with the idea of homology, to consider symmetry between these spaces, using canonical correlation analysis. In other words, we discuss how far methodology and measurement matter when discussing social and cultural spaces.

1 Introduction

In the social sciences, researchers, who address the same topic of investigation, even using the same data, might prefer to apply different methods when analysing these data. The methodological consequences of the choice of research design are most often not evaluated in detail. Obviously, these choices have consequences, at least in two aspects: a) Different research designs are often associated with different methods, and, b) these methods are often based on different assumptions about the measurements of the (same) variables. In this article we evaluate in detail the consequences of these methodological choices. We will introduce the Gifi framework (Gifi, 1990), which is rarely used in sociology, in order to compare different methodological designs applied to the same data.

Our empirical topic of investigation is studying social space and cultural activities in modern societies, and we will compare outcomes in three countries: France, Norway and Switzerland. Inspired by the homology thesis of Bourdieu[1], some researchers in the field of social stratification and cultural consumption generate latent dimensions by means of multiple correspondence analyses (MCA) (Lebaron and Le Roux, 2015), whereas other researchers generate latent dimensions by means of other multivariate methods, such as principal component analysis (PCA) or multi-dimensional scaling (Chan, 2010). These methodological choices are associated with different ways to consider the data: MCA refers to nominal variables, PCA or multi-dimensional scaling are based on higher level measurements, such as interval or ordinal level variables.

Therefore, even when researchers use the same original data, the choice of method, and associated assumptions about the level of measurement of the variables, might have implications for how well the models fit the data, as well as – perhaps – for the substantial outcomes of the analyses. Perhaps surprisingly, to our knowledge there is no study discussing these issues - comparing different methods and measurements applied to the same data. In addition, there is no consensus on how we might best compare different models in terms of how well they fit the data. In this paper, we begin to undertake this extensive task, by testing empirically three different methods, with associated assumptions about variable measurements and single versus multiple scaling, applied to the same data. Comparative studies of methodological choices have been few, even more in this field of research, so it is often unclear if differences in outcomes are driven by methodology or are meaningful in terms of substantial results.

On a substantial point of view, the main research question addressed in this literature is how cultural consumption is associated with social position.

[1] First Edition 1979, éd. de Minuit, Paris. English translation 1984, Harvard University Press, Harvard.

In Bourdieu's tradition, social space can be defined as a two-dimensional space, where the first dimension captures the overall volume of capital and the second dimension captures the composition of capital (Bourdieu, 1985: 724). This is also in line with the "Japanese lesson" when Bourdieu was writing "Social space is constructed in such a way that agents or groups are distributed in it according to their position in the statistical distribution based on two differentiation principles which, in the most advanced societies, such as the United States, Japan, or France, are undoubtedly the most efficient: economic capital and cultural capital" (Bourdieu, 1991: 631).[2] Such a multi-dimensional representation of the social structure is rather common in the analysis of social stratification; see, for instance, Wright (1985) or Oesch (2006). [3]

Using comparable data from national representative surveys on cultural consumption in France, Norway and Switzerland, we will compare the association between a set of variables measuring cultural consumption and a set of variables measuring social position. In particular, we will discuss the methodological aspects of these comparisons.

Theoretically, according to Bourdieu's proposition, the social and cultural spaces symmetrically shape each other, yet the most common approach to this type of investigation treats the two sets of variables in a pre-defined order: first, a cultural space is constructed, and then the social position variables are "added".[4] There are a number of examples of such a strategy in the literature, such as in Coulangeaon and Duval (2015), Lebaron and Le Roux (2015), Rosenlund (2015) and Bennett et al. (2015) to mention a few. In our case, in order to find the best association between the two sets of variables, we will apply a canonical correlation design, which generates latent dimensions that are the best possible representation of the two sets of manifest variables under the constraint of maximum correlation between the two-by-two sets of latent dimensions. We apply the canonical correlation methods, giving the same weight to both sets of variables, within a Gifi framework (Gifi, 1990), which allows flexibility in terms of the levels of measurements for all the variables included, such as nominal, ordinal and interval. Thus, "traveling with Gifi" allows the researcher to experiment with the levels of variable measurements in order to maximize specific parameters in the data analysis process, such as correlations or eigenvalues (de Leeuw, 2014).

Thus, our paper has three intertwined contributions, methodological as well as substantial. First, we believe our discussion contributes to the theory

[2]Some authors use the concept of "Social space" as comprising both "social positions and particular dispositions in matters of taste" (Melldahl and Börjesson 2015). We prefer to keep social space as referring to the structure and composition of capital.

[3]The omnivore thesis, associated with the publication of Petersen (1992) and later the one of Chan and Goldthorpe (2007), is less explicit about the number of stratification dimensions, even if these analyses mainly operate with one principal dimension.

[4]Or the reverse as for example: Prieur, Rosenlund and Skjott-Larsen (2008).

of measurement by demonstrating the usefulness of explicit considerations about different levels of measurements. We also believe re-introducing the canonical correlation design to generate latent dimensions is useful, in particular for this topic of investigation. Third, by comparing three countries we will provide a robust test of our findings to see if the associations between cultural consumption and social position are affected by different national contexts. Although our paper addresses a particular substantive topic, the homology thesis, we hope that our discussions and analyses might be helpful for other topics of investigation as well, where one of the main purposes is to identify meaningful latent dimensions in different social contexts.

This article is organized in the following way: first we present the methodological framework of our analyses, which includes a discussion on the scaling of variables and the canonical correlation design. We then present the results of our analysis, where we report both single scaling models and (one) multiple scaling model. To be able to adjudicate between these models in a way that balances their complexity with the number of parameters to estimate, we propose a measure of the models' explanatory power, taking into account its complexity. The implications of the substantive results as well as the methodological consequences are discussed in a last conclusive part.

2 Methods

The Gifi framework was developed by Jan de Leeuw and colleagues at the University of Leiden, building on the 1990-book (Gifi 1990; de Leeuw 1984). The Albert Gifi team[5] developed a system of nonlinear multivariate analysis that extends various techniques, such as principal component and canonical correlation analysis.[6] The Gifi framework is relevant to us because it is a general way to present different multivariate techniques. For example, multiple correspondence analyses, often used in the French tradition of "Analyse des données", as well as principal component analysis, could be seen as particular cases of the Gifi framework . The Gifi framework considers a continuum of restrictions ranging from the least constrained, where each variable category is independent of the others, as it is in correspondence analysis, to the principal component, where we have, by construction, a metric order between the categories of each variable. This range of methods with more and more constraints includes of course analyses of ordinal variables by trying to define an "optimal" scaling between the categories. In other words, the Gifi framework allows us to consider variables as interval, ordinal and nominal

[5]Albert Gifi was the servant of Galton, the latent helper, whose contribution was large, but mainly unknown to the outside world. The team at University of Leiden, working with scaling and categorical data decided to publish their main contribution using his name, to honor "... his loyalty and devotion" (Gifi, 1990, p. X) and to show the feelings of equality among the authors (van der Heijden & Sijtsma, 1996).

[6]Cox & Cox (1994), see also http://gifi.stat.ucla.edu.

and to appreciate the loss of fit that such situation implies.[7] In other words, we will find an optimal coding of the variables, respecting the constraints, in order to maximize an indicator of quality of the model. This is what me mean by "scaling".

2.1 Scaling

The nominal, ordinal or interval characteristic of variables is a complicated topic. For instance, consider the following question, as translated from one of our surveys: "How often do you go to an art exhibition?" The questionnaire includes the following answers: 1) never; 2) not often; 3) sometimes and 4) often. The variables could be treated at face value, which means that the codes 1, 2, 3 and 4 would be used directly, with 1 measuring the distance between the responses (constant intervals between the categories). This is in fact an identity transformation, which in our paper is called the "single interval" solution. This praxis is often unproblematic, in particular if, for instance, the original variables represent quantities, like the number of times a year a person goes to an art exhibition, a formulation used by another of the countries considered here.[8] This strategy is, however, not necessarily the best solution. For example, as long as we have erguments, we might argue that the distance between "never" and "not often" is bigger than that between "not often" and "sometimes." Therefore, a better coding might be something totally different while still respecting the original order of the categories.

As an example, assume that a coding like 0) never; 5) not often; 7) sometimes and 12) often will be better according to some criteria—which we would of course need to specify. . The PCA uses as measure of quality the proportion of the variance explained. That means that we can search in the case the numerical value of the categories that maximises this criteria. In other types of analysis, such as regression analyses, R-square measure how well the model fits the data and could be used as a criteria to maximise. In some instances, such criteria could also be a correlation measure or a regression coefficient. For example, in a very well-known example, when Ganzeboom and colleagues (1992) proposed the International Socio-Economic Index, ISEI, the model was based on the minimization of a regression coefficient. In the same logic, Clogg (1984) used a maximizing strategy linked to an external variable in order to estimate the "value" of missing answers. More generally, there are many measures that could be used as external criteria in order to find the "best" scaling.[9] In the case of canonical correlation analysis, the

[7] More precisely, every multivariate analysis has some measure of quality of the fit of the model to the data, most often based on explained variance.

[8] By the way, this last form is recommended also for other reason in the literature about question design. See for example Conrad, Brown and Cashman, 1998.

[9] See, for instance, the R-package Aspect (Mair and de Leeuw 2010).

sum of the squares of the canonical correlations will be the criterion .

In line with the discussion on the measurement level of variables, there are different forms of scaling, related to single and multiple solutions.

The "single ordinal" solution preserves the ordinal measurements of the variables, applying an external criterion to find an optimal transformation of the scaling of the variable. Thus, the "single ordinal" solution builds interval variables from ordinal ones, with the help of additional information.

We can, of course, also imagine a more complex transformation, allowing not only the values to change, but also the order of the categories. This can be seen as consuming more "degrees of freedom," but this kind of transformation can give a better fit. Such a case is called the "single nominal" one.

In the terminology used here, the examples given so far are "single" scaling models, as only one transformation is used for each variable. This will necessarily be the case if the solution of the multivariate analysis considered is uni-dimensional. However, if, as in our case (and many others), the analysis implies generating more than one latent dimension, then the manifest variables involved might be scaled differently when estimating each latent dimension. This is the "multiple" scaling solution, which is in fact an alternative way to present multiple correspondence analysis.

We will return to the difference between "single" and "multiple" scaling models when we present the results of the analysis. It is nevertheless important to underline that if we have only discussed until now a model composed by a single set of variables like in PCA or MCA, this logic can of course be generalized to two or more sets of variables, like in the family of canonical analysis, which we will address soon.

Such a discussion on the scaling is even more important in a context of comparative studies, for example cross-national ones. Just take an example. Even if income is measured as a continuous variable (interval level) it might make sense to recode the variable into an ordinal variable, income by quantiles . Income quantiles will distinguish respondents according to their position in the income hierarchy within each country, independently of GDP, exchange rates and living costs. This could also be true for other variables where the meaning depends on the context. In this perspective, some indicators may offer "better" measurement, from a comparative point of view, if they are considered as ordinal rather than interval. The discussion about measurement level is therefore not only a theoretical or statistical topic, but also an empirical one.

In other words, the measurements of the variables, nominal, ordinal and interval, which is often discussed in statistical textbooks for the social sciences in a rather "mechanistic perspective", could sometimes be seen as partially misleading. This is even more problematic if particular statistical tools are directly and automatically linked to what is supposed to be the measurement level: for example mean and standard deviation for interval variables,

median and interquartile range for ordinal variables. However, this does not mean there is complete freedom in the choices of measurements and scaling. Adopting a nominal strategy, for example, adds complexity by implying the need to estimate more parameters. In other words, to use more "degrees of freedom" means that there is a price to pay.

Adjudicating between methods and associated models is analytically very challenging. In order to choose between different models, we apply the scientific principle of "Occam's Razor" (Lex parsimoniae), a heuristic device of parsimony (i.e., simplification), which is used in logic, research, problem-solving and even in artistic productions. If two models are alike, in terms of how well they describe the data, but one is more complicated than the other, we should prefer the most simple/least complicated model. We will consider what this principle means after having presented canonical correlations in more details.

2.2 Canonical correlation

The canonical correlation method was initiated by Hotteling as early as 1938 (Gittins, 1985). Canonical correlation measures the relationship between two sets of variables by searching for the best linear combinations of variables in the first set that could be related to the best linear combinations of the variables in the other set (Levine 1977). The canonical correlation method can be seen as one variant of classical factorial models. The underlying idea is the following: Suppose two sets of manifest variables are measured for the same observations, the dimensionality of the first and second set being p and q, respectively. Then, p and q latent variables (factors) can be calculated (subject to various assumptions and standardizations) respecting the following constraints: The p latent variables are uncorrelated, and the q latent variables are uncorrelated too; the first of the p latent variables is maximally correlated to the first of the q variables, the second of the p latent variables is maximally correlated to the second of the q variables, and this is the same for the third dimension of each set, and so on. To summarize, the constructed latent dimensions are explanatory factorial dimensions of the two respective original manifest variable sets under the specification of having the strongest possible correlations between each set of latent dimensions. Of course, as in factor analysis, we expect that the first dimension (here: the first pair of dimensions) will carry most of the meaningful information and that we can neglect the last dimensions. In our case, see Figure 1, the manifest variables on the left side measure social position, and the manifest variables on the right side measure cultural activities. Thus, L1 / L2 refer to the latent dimensions of social space, and R1 / R2 refers to the latent dimensions of cultural consumption.

Canonical correlation will generate as many pairs of latent dimensions as the minimum number of manifest variables (i.e., with three variables for

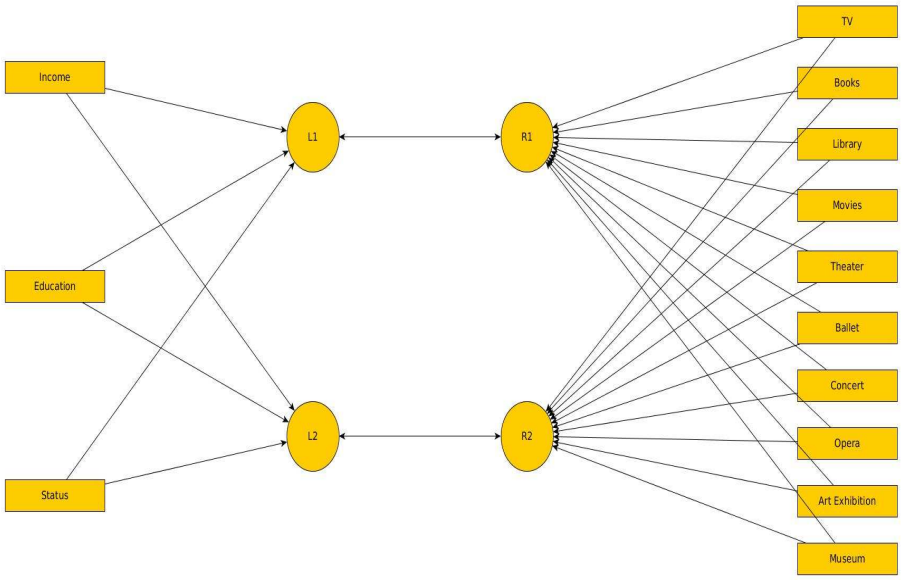

Figure 1: Canonical correlation model as used in this chapter

social position and 10 variables for cultural consumption, canonical correlation will generate three pairs of latent variables: L1 and R1; L2 and R2; L3 and R3). We only present the first two pairs here.[10]

The canonical correlation analysis takes into account the central hypothesis proposed by Bourdieu, namely that social positions cannot be defined without simultaneously considering the lifestyles that may be associated with them. By correlating two sets of variables—one measuring social space (such as education, income, status, etc.) and the other measuring cultural activities (such as attending the opera, ballet, etc.) the canonical correlation method offers a fairly direct way to assess the relevance of the theory of structural homology. The magnitude of the correlation between the two latent dimensions/canonical components (built on each of the two sets of manifest variables) is a direct evaluation of the homology hypothesis. Despite all these advantages, we are aware of only one text (Frie & Janssen, 2009) applying the canonical correlation method in work inspired by "La Distinction."). Here, we will use this method systematically, varying the measurement level attributed to the variables. In the Gifi framework, this means recalculating the scaling of the variables under the assumption that the two sets of variables, measuring respectively social and cultural space, should be as highly correlated as possible (in line with the homology hypothesis of Bourdieu).

[10]In fact the criteria to chose for considering the number of dimension is analogous with the one use in PCA, looking to the change of explained variance from one solution to the other.

To summarize, within the Gifi framework, we will combine two tasks in our analysis:

- A statistical model to study the structure and relations of the variables. In our case this means that we will apply the canonical correlation model with the usual criterion of maximizing the strength of the correlation between the latent dimensions (pairs of latent variables).[11]
- An optimal scaling of the variables, according to the criterion mentioned above, distinguishing from single interval to multiple nominal.

Using the software developed in the Gifi tradition (Canals, Homals, and Overals),[12] we can consider both tasks within the same frame (Gifi, 1990; see also http://gifi.stat.ucla.edu).

3 Data and variables

The three countries we compare are similar enough to make comparisons meaningful, yet different enough to provide contextual variation. These countries are all located in Europe and are relatively wealthy; they all have a long tradition of democratic institutions, and all three countries have high scores on UN's well-being indexes.[13]. Looking at differences, Norway and Switzerland are both outside the EU and have smaller populations, whereas France is one of the founders of the EU and has a large population. In addition to geographical, historical and demographic diversity, the three countries also differ in cultural terms. According to Hofstede's model on cultural dimensions, France, Norway and Switzerland are similar in some respects (individualism and long-term orientation), but different in other dimensions (power distance, uncertainty avoidance and masculinity).[14] An important reason for comparing these three countries is the available survey data, which allows for a detailed comparative analysis. As we will see, even if these surveys are fairly comparable, there are still some variations between them. This challenge is frequently occurring in comparative research, and our analyses will therefore also contribute to a discussion of the importance of different levels of measurements in a comparative context as well as their impact on the results.

[11] More precisely, this is the sum of the squared canonical correlation coefficients.

[12] Homals is available in R, Overals corresponds to the module Categories in SPSS while Canals is still available as a standalone Fortran Program (http://gifi.stat.ucla.edu). For technical reasons we have mostly used Canals in our analysis.

[13] Furthermore, these three countries have also been more investigated in this research perspective than other European countries: outside the bourdieusian tradition of France. See for example to the work of Rosenlund (2010, 2015) in Norway and Tawfik (2013) in Switzerland.

[14] http://geert-hofstede.com/dimensions.html

3.1 Surveys

The French survey, Participation Culturelle et Sportive, was part of the May 2003 issue of the "The Continuous Survey of Living Conditions" carried out three times a year by the French Statistical Office. The random sample, with a response rate of 67 percent, is representative of individuals aged 15 and over who live in private households in metropolitan France. The Norwegian survey, Kultur og mediebruksundersøkelsen, was carried out in 2004 by Statistics Norway. The sample is representative of the Norwegian resident population aged 19-79 years. The response rate is 70 percent. The Swiss survey, Pratiques culturelles en Suisse, conducted in 2008 by the Swiss Statistical Office with a CATI procedure, obtaining 66% response rate. For comparative purposes, we limit the samples in all three countries to include only occupational active individuals between 20-64 years of age, which leaves us with a sample of 3,744 individuals in France, 1,005 in Norway and 2,442 in Switzerland.

For France and Norway, more recent surveys on cultural consumption are available, but not for Switzerland. We therefore decided to use surveys similar in time to the Swiss survey. Although these data are not entirely fresh, we note that Tawfik (2013) has concluded that the relation between social and cultural spaces is probably stable on the long run.

3.2 Variables

The variables we include, as well as the characteristics of these variables, are important for testing differences in methods and measurements. Can we plausibly argue a priori that variables pertinent for analysis of social space should be seen as nominal, ordinal or interval alternatives? Some components of social position, such as education, are sometimes treated as interval level variables, yet they often are measured at less precise levels in surveys. This may be problematic in countries with a vocational training system and an academic track at the same level, yet with different implications for human capital. Nominal variables, such as respondents' occupations, are often grouped into categories such as social classes, or they are ranked according to an interval-level socio-economic scale or other scales (such as Treiman's prestige scale or status scales). Furthermore, these variables can be linked to the individual or the household, leaving place for a very vivid debate, on the 'proper' unit of analysis in social stratification, which we will not engage in here, except to note that if we use individuals as the basic unit of analysis, we also include a measure of economic capital based on the household's income.

In our analysis, in order to describe the social space, three variables are used: education, household income and occupation. Education is measured in Norway by the number of years necessary to get the highest diploma, with

10 categories. In the Swiss and French data, education has 6 categories; and in Switzerland these categories also include information on vocational training. In all three countries household income is measured as the logarithm of household income per capita. Occupation is measured according to the ISEI scale in Switzerland (Joye & Chevillard, 2011) and the Chan and Goldthorpe social status scale in France and Norway (France: Cousteaux & Lemel, 2004; Norway: Chan et al., 2011).[15] Ideally we would have wanted to supplement these variables with other indicators of social structure, such as home or stock ownership.. Unfortunately, this information is not available, at least in a comparative way, in our three surveys. However, it is clear that income, education and occupation basically are the three most important indicators used by stratification and social mobility researchers, which also correlate with other measures of social position. We are therefore confident that this set of manifest variables, can fruitfully be used to describe the social structure and form the basis of constructing the social space.[16]

The three surveys include a variety of questions about cultural practices, such as reading, listening to music, visiting museums, going to the opera and other cultural outings, television viewing, artistic hobbies and sports activities, allowing us to form a set of variables measuring cultural consumption. These questions are retrospective in nature, asking about the respondents activity over the past 12 months (excluding professional and school obligations). Ten cultural activities were similarly fielded in the three countries: watching TV, reading books, visiting library, movies, theatre, ballet, concert, opera, art exhibition and museum.[17] The cultural consumption variables are usually measured at an ordinal level, using a scale of ranked values. Nevertheless such variables are often analysed as interval ones, with the justification that they measure an unobserved latent continuous dimension.

Our survey data include rather small samples. In addition, we have limited our analyses to occupationally active respondents between 20 and 64 years of age. To avoid excluding too many individuals, we decided, for variables where there were many individuals with missing values, to impute new values for the missing cases using auxiliary information.[18]

[15]The Chan and Goldthorpe social status scale was never computed for Switzerland. We could have used the CAMSIS score, but decided for the sake of simplicity to stay with the ISEI. The impact of this choice is probably small as bi-variate correlations between different social status scales are at the 0.90 level or more in Switzerland (Joye & Chevillard, 2013).

[16]All these variables were recoded into a smaller number of categories after considering their distribution. A table in the Annex recapitulates the number of categories used for each variable, in each country. For previous analyses using these data, please see Birkelund & Lemel (2013).

[17]This list of variables is close to the one that can be found in international surveys like the module 2007 of ISSP or the Eurobarometer (Gayo, Joye, Lemel, 2016).

[18]In the French data, the French Statistical Office (INSEE) routinely edits data before they are issued for research purposes. This also means imputing income for households without information on this variable. One Norwegian household with an annual income

4 Results

In a country by country strategy, we will start by introducing single scaling models, which assume that each manifest variable has the same transformation of the scale throughout the whole analysis. Within the single scaling models, we first introduce the interval level model, which assumes a fixed distance between the values. Second, we then assume an ordinal level model where the categories can be ranked, but the distances between them might vary. Third, we assume the manifest variables are at the nominal level, where each category of each variable can have is own numerical value. These nominal level models are the most demanding in terms of degrees of freedom whereas the other models gradually introduce more restrictions.

For each country, a table comparing the canonical correlation between the latent dimensions generated by the three single scale models, the single nominal, the single ordinal and the single interval, is computed. If two models provide a similar level of explanation, that means having similar canonical correlation coefficients, the simplest model should be preferred (confer Occam's razor). We therefore adjudicate—separately for each country—which of the single scaling models are the most efficient for maximizing the correlation between the two sets measuring social and cultural space.

We then move on to the multiple scaling models, which imply that we allow the values of each variable to be transformed differently for each dimension within the same canonical correlation analysis. In fact we will only consider a multiple nominal model, and we will compare the outcomes of this model with the "best" single scaling solution.

4.1 Single scaling models

As noted above, the main reason for generating latent dimensions is revealing systematic patterns in the data that are not easily detected when we analyse the manifest variables directly. We only consider the most meaningful latent dimensions, since keeping all latent dimensions means no reduction of information compared to the original manifest variables (with three variables in the smallest data set, there will be at most three dimensions to consider). Using a traditional criterion, like the decrease of the canonical correlations between the two variable sets, we have decided to continue with two latent dimensions.

higher than 8 million NOK was excluded. 65 Norwegian respondents had missing values on household income, and we therefore imputed the average income for these respondents. In the Swiss data, nearly 20 percent were missing values for the income of the household. Their income was then estimated with a linear regression using age, sex, education, ISEI of the occupation and household composition. More than 33 per cent of the variance was explained by this regression. This income was then imputed in case of missing information. This design implies we end up with 2,442 individuals.

Switzerland

	Correlation 1	Correlation 2	R2
Single interval	0,47	0,19	0,25
Single ordinal	0,49	0,20	0,28
single nominal	0,49	0,20	0,28

France

	Correlation 1	Correlation 2	R2
Single interval	0,60	0,14	0,38
Single ordinal	0,61	0,24	0,43
single nominal	0,61	0,24	0,43

Norway

	Correlation 1	Correlation 2	R2
Single interval	0,53	0,23	0,33
Single ordinal	0,55	0,26	0,37
single nominal	0,56	0,26	0,38

Table 1: Canonical correlations in the single interval, single ordinal and single nominal solutions

In order to evaluate the goodness of fit we have constructed a R2 like statistics, which will be used later together with the potential number of degrees of freedom to measure the models' overall fit to the data.

Each model we estimate is a replica of the stylized model given in Figure 1, with manifest social position variables on the left side and manifest cultural consumption variables on the right side of the model. Each model generates latent left side variables and latent right side variables. Table 1 shows the canonical correlations between the left and the right side latent dimensions, as well as the sum of their squared value, which gives an indication of the explained variance.

In Table 1, we have of course a better fit at the nominal level, confirming that the more degrees of freedom we use, the higher the fit will be. Still, we find that for each country, there are hardly any sensible differences between the canonical correlations (Correlation 1, Correlation 2) associated with the

	Switzerland			France			Norway		
	I-O	I-N	O-N	I-O	I-N	O-N	I-O	I-N	O-N
L1	0,98	0,98	0.999	0,96	0,96	0.999	0,99	0,97	0.999
L2	0,91	0,89	0,99	0,79	0,79	0.999	0,99	0,98	0.999
R1	0,96	0,96	0.999	0,98	0,98	0.999	0,97	0,96	0.999
R2	0,86	0,79	0,97	0,78	0,76	0,99	0,93	0,9	0,99

Table 2: Correlations between the latent variables obtained in the three simple scaling models

nominal and ordinal level of measurements, respectively. Comparing the ordinal and interval levels, we find, in each country, slightly lower canonical correlations associated with the interval level, which indicates that the ordinal level of measurement seems more appropriate: The ordinal level models are associated with higher explanatory power than the interval level models, and the ordinal level models utilize fewer degrees of freedom than the nominal level models. In addition to the canonical correlations, we note that the R2 for each country is slightly lower for the interval level models. The strength of the relation as measured by this R2 between the left and right side latent variables or social and cultural spaces respectively, is strongest in France and weakest in Switzerland, consequence of a higher canonical correlation for the first dimension.

There is an alternative way to consider the changes related to different scalings. In Table 2, we report the correlations between the latent variables obtained in the three models. L1 and L2 as well as R1 and R2 represent the first and second latent variable on the left (L) and right (R) side, respectively. I, O and N are abbreviations for single internal, ordinal and nominal scaling respectively.

According to such a criteria, the measurement levels do not seem to matter for the first dimensions where the correlations are close to 1 independent of the scaling, meaning that the nominal scaling does not add any information concerning the position of the respondent when compared to the ordinal one.

The other correlations in Table 2 are relatively weaker even if extremely high (0.76-0.98). In particular, for R2 (the second dimension on cultural consumption), we see that the scaling matters somewhat more, with correlations varying between 0.79-0.97. The lowest values are observed for France, meaning that in this country there are probably more changes after transformation of the variables. This result is not unexpected, given country differences and the way the questions have been asked : In Norway, the questions on cultural activities concerned the number of occurrences (i.e. referring to what

	Switzerland		France		Norway	
	L1	L2	L1	L2	L1	L2
Isei	-0.65	0.08	-0.85	0.52	-0.77	-0.04
Income	-0.63	0.76	-0.82	-0.25	-0.96	0.01
Education	-0.93	-0.29	-0.67	-0.52	-0.21	-0.97

Table 3: Loadings on the left set, ordinal measurement

is expected to be an interval level) whereas the same questions in France (and Switzerland) were quantified as "often" or "somewhat often" (i.e., the ordinal level), which is seen in the survey literature as a measurement of lower quality.

4.1.1 The single ordinal model in detail

The best single scaling model according the number of parameters is the ordinal one. But what is in this case the interpretation of the dimensions obtained? We first look at the two left-side dimensions of social space, then the two right-side dimensions of cultural activity.

Loadings of the left set – Social space There are four latent variables associated with two sets of the canonical correlation solutions. On each side of the canonical correlation model, the latent variables are constructed from the manifest variable sets, and the correlations between each manifest and latent variable show their relations. These correlations are called loadings, as in PCA and we have to keep in mind that the algorithm generating these latent variables is set up so that they are defined under the condition of maximizing the correlation between the right (R) and the left side (L). In our analysis, this means that the two latent variables measuring "social space" are generated simultaneously with the two latent variables measuring cultural activities. We adhere to a common practice of only emphasizing loadings above 0.4 when interpreting the results.[19]

[19]Some authors have considered a rotation in the context of canonical correlation, though this is not common. We decided not use rotation in order to keep the structure of the correlation between the two sets. Gittins (1985) as well as Levine (1977) have discussed this possibility, but we have not yet seen any empirical analysis using rotation in canonical correlation models. This could be due to the fact that rotation algorithms are not implemented in canonical correlation packages. meaning the need to use other packages for rotations. Furthermore, a rotation aimed at optimizing the solution for the left and the right set respectively, may end up loosing the symmetry in the optimization.

In the three countries, all the manifest variables have negative loadings on the first latent dimension (L1). Thus, the first left hand side latent variable appears to be an indicator of global capital, where a high score implies a low volume of capital, and vice versa.[20] This pattern is the same in all three countries.

In the second latent dimension (L2), the three variables have different loadings in the different countries. In Switzerland, income has the highest loading in the second latent variable; in France, social status is opposed to education; and in Norway, education is the only manifest variable with a noticeable loading on the second latent variable (L2). These latent variables appear to differ partly from the usual Bourdieusian interpretation of the second latent dimension of the social space (i.e., the capital composition). One reason for this result might be that we have only crude measurements of income (at the household level only), whereas education and social status are measured at the individual level. Another reason for this finding could be that the Bourdieusian interpretation of the second dimension of social space is not relevant, at least not with these data. A third reason might be that we here generate the latent dimensions of the social space simultaneously with the latent dimension of the cultural space, based on the manifest variables of cultural activities. For now, we note that the second latent variable differs between the countries. We will discuss these possible interpretations later.

Loadings of the right set – Cultural space In all the countries, nine out of ten manifest variables have negative loadings on the first latent dimension (R1). In all countries R1 captures an opposition between TV and all the other cultural activity variables. This means that this dimension could be interpreted as a kind of omnivore dimension, contrasting TV with all the other cultural activities.

The second dimension of the right set is more difficult to interpret. In Norway and Switzerland, visiting libraries have the largest loadings on R2, whereas in France, this aspect also appears, yet weaker. However, none on these cultural activities seem to have a strong discriminating power (higher loadings than .40) in France.

In other words, the first latent dimensions shows a relatively strong relation between a global capital and a cultural composition, with a relatively high canonical correlation reinforcing the homology hypothesis, whereas the second latent dimensions seems more particular with reference to the use of library with a lower canonical correlation. This implies that the homology is stronger in an uni-dimensional perspective within the single scaling models.

[20] Like in PCA, the latent variables have no a priori direction and the fact to have negative or positive correlation have no meaning in itself, as long as the analyst is careful about this in the interpretation.

	Switzerland		France		Norway	
	R1	R2	R1	R2	R1	R2
TV	0.41	0.38	0.37	-0.15	0.29	-0.26
Book	-0.7	-0.09	-0.63	0.28	-0.39	-0.19
Library	-0.23	-0.83	-0.37	0.36	-0.41	0.72
Movie	-0.55	-0.09	-0.63	0.52	-0.57	0.16
Theatre	-0.47	0.06	-0.56	-0.37	-0.52	-0.13
Ballet	-0.27	-0.16	-0.26	-0.19	-0.27	0.3
Opera	-0.55	-0.26	-0.38	-0.24	-0.39	-0.01
Concert	-0.48	0.04	-0.53	-0.18	-0.17	-0.02
Museum	-0.74	-0.17	-0.71	-0.32	-0.75	-0.2
Exhibition	-0.53	-0.02	-0.76	-0.31	-0.56	-0.15

Table 4: Loadings on the right set, ordinal measurement

4.2 Multiple scaling model

Until now, the scaling of the manifest variables have been considered as "single," meaning that each manifest variable has kept the same scaling when constructing the different latent dimensions. This restriction might be relaxed to allow for multiple transformations, one for each dimension. In fact such a multiple nominal scaling corresponds to MCA, used in Bourdieu's tradition. We will do the same in the context of canonical correlation and see whether the difference between single and multiple scaling models is important.

We first note that multiple interval models do not make sense. A multiple interval model would imply taking the scaling of the manifest variables at face value (i.e., no transformation of the variables, implying that the scaling of the manifest variables are the same as in the original data file). Subsequently, there is only one way to measure these variables, and a multiple scaling does, per definition, not exist in this case. Multiple ordinal models could be imagined, but we are not aware of any study that has used this complicated design. Thus, in the world of multiple scaling, we only consider usual multiple nominal models, as implemented in fact in the MCA.

Table 5 compares the previous canonical correlations of the single scaling models (as presented in Table 1) with the canonical correlations obtained in the multiple scaling models. We include the same R2-like statistic, defined as the sum of the square of the canonical correlations, in order to assess the

	Switzerland			France			Norway		
	Cor1	Cor2	R2	Cor1	Cor2	R2	Cor1	Cor2	R2
Single Interval	0.47	0.19	0.25	0.60	0.14	0.38	0.53	0.23	0.33
Single Ordinal	0.49	0.20	0.28	0.61	0.24	0.43	0.55	0.26	0.37
SingleNominal	0.49	0.20	0.28	0.61	0.24	0.43	0.56	0.26	0.38
Multiple Nominal	0.49	0.23	0.29	0.62	0.28	0.47	0.56	0.29	0.40

Table 5: Canonical correlations in the three countries according to scaling

global "quality" of the link between the two sets.

When comparing the models in Table 5 we first appreciate the increase in the R2 from one model to the other. As the multiple nominal model uses the maximum number of degrees of freedom, it is therefore to be expected that the R2 increases for this model, but, according Table 5, this increase in either non-existing (in Switzerland) or relatively minor (in France and Norway). The question, however, is how to balance this increase of complexity in relation to the models' explanatory power. We will propose a simple rule, based on the following reasoning. First, in single interval models, the transformation is an identity matrix (which means that the scaling of the manifest variables is the same as in the original data file). In this case, the model uses no extra degrees of freedom for the scaling of the categories. Second, in single ordinal models, the ranking of the values is based on fixing the extreme values and then letting the intermediate categories vary while respecting their order. This means that the degree of freedom in such a case can be approximated by the sum of the number of categories minus two (i.e., the two fixed extremes) for each variable. Third, in the single nominal case, the degrees of freedom, as defined here, are based on the number of categories minus one. Finally, in the multiple nominal models, the degrees of freedom will be twice that of the single nominal in the case of two dimensions, three times more in case of three dimensions and so on...

In Figure 2, the estimated R2 associated with these four models are shown by an indicator of degrees of freedom. R-square for Switzerland is in green, Norway in red and France in black. We see that for each country, the single interval model is associated with the smallest number degrees of freedom, whereas the multiple nominal models are associated with the largest degrees of freedom.

Based on this estimation, we now have a comparison of the gain in the R2 according to the increase in the degrees of freedom used when we compare the three single scaling models and the multiple nominal models. In each case, when we take into account the growth of the number of parameters, the single ordinal model seems to be the most efficient solution. This does

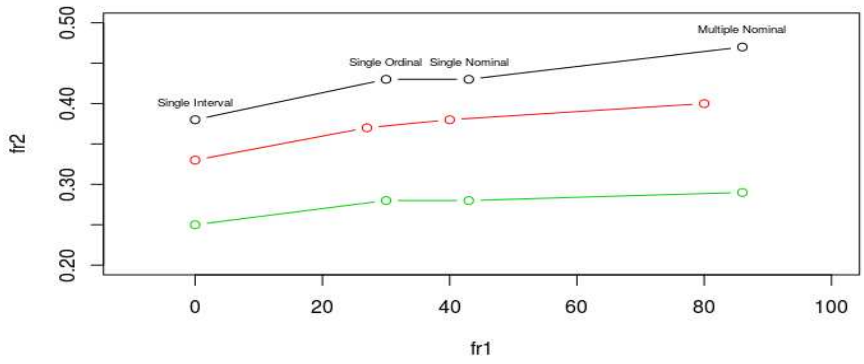

Figure 2: Estimated R2 in three countries according to an indicator of degrees of freedom

not mean that it is not interesting to consider, particularly in the case of France, the added information by the multiple nominal models.

One way to do this is to look at the rescaling proposed by the multiple nominal solution and presented in Figures 3 (France), 4 (Norway) and 5 (Switzerland), which show the original and the transformed values for each manifest variable for the first latent dimension (in black) and the second latent dimension (in red). It is interesting to note that for the cultural space (see the lowest 10 sub-figures), all transformations (rescaling) are in line with the original order of the values, confirming the idea that ordinal transformation is optimal. We also noted in Figure 2 that the differences between the scaling of the two dimensions were very small. In other words, in each country, the cultural activity variables would best be considered as ordinal variables.

When we look at the scaling of the three variables describing the social space, we see that for some categories there are more differences according to scaling. These differences are found among values/categories in the middle rank, and not in the upper rank categories. If we had found a particular pattern for the upper positions, we might argue that we need to differentiate the ranking of the elite from the other categories. This might justify a different ordering, distinguishing a cultural and economic elite. However, as seen in Figures 3, 4 and 5, the irregularities in the sorting between neighbouring categories are small and not particularly located at the top of the hierarchies as we should have expected by reference to the presentation Boudieu does in The Distinction. Nevertheless, it is interesting to underline that this more in the measurement of social positions that non-linearities are observed, showing the difficulty of measuring social position in simple linear models, while the cultural composition seems in this regard easier to measure.

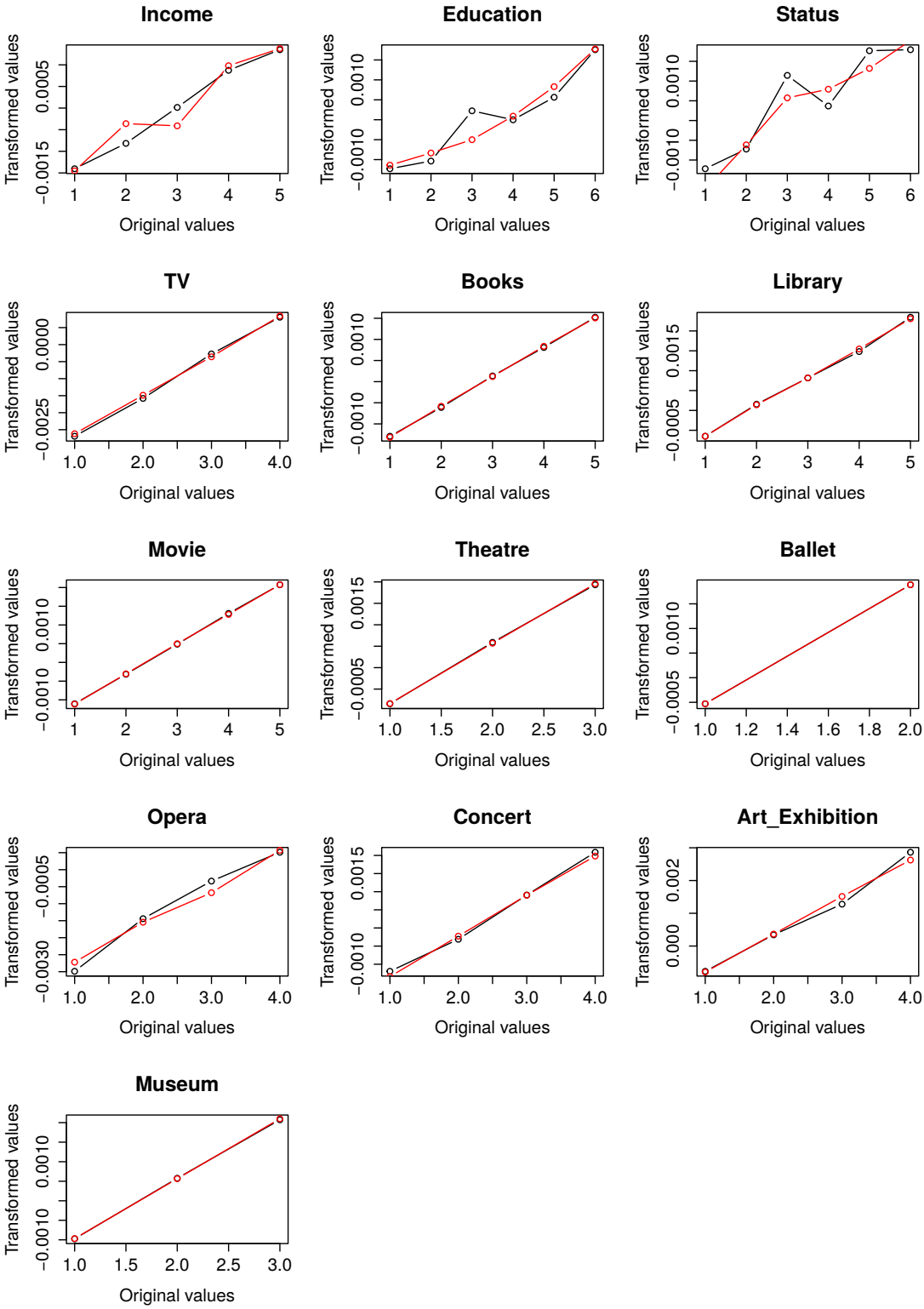

Figure 3: Transformation for Switzerland

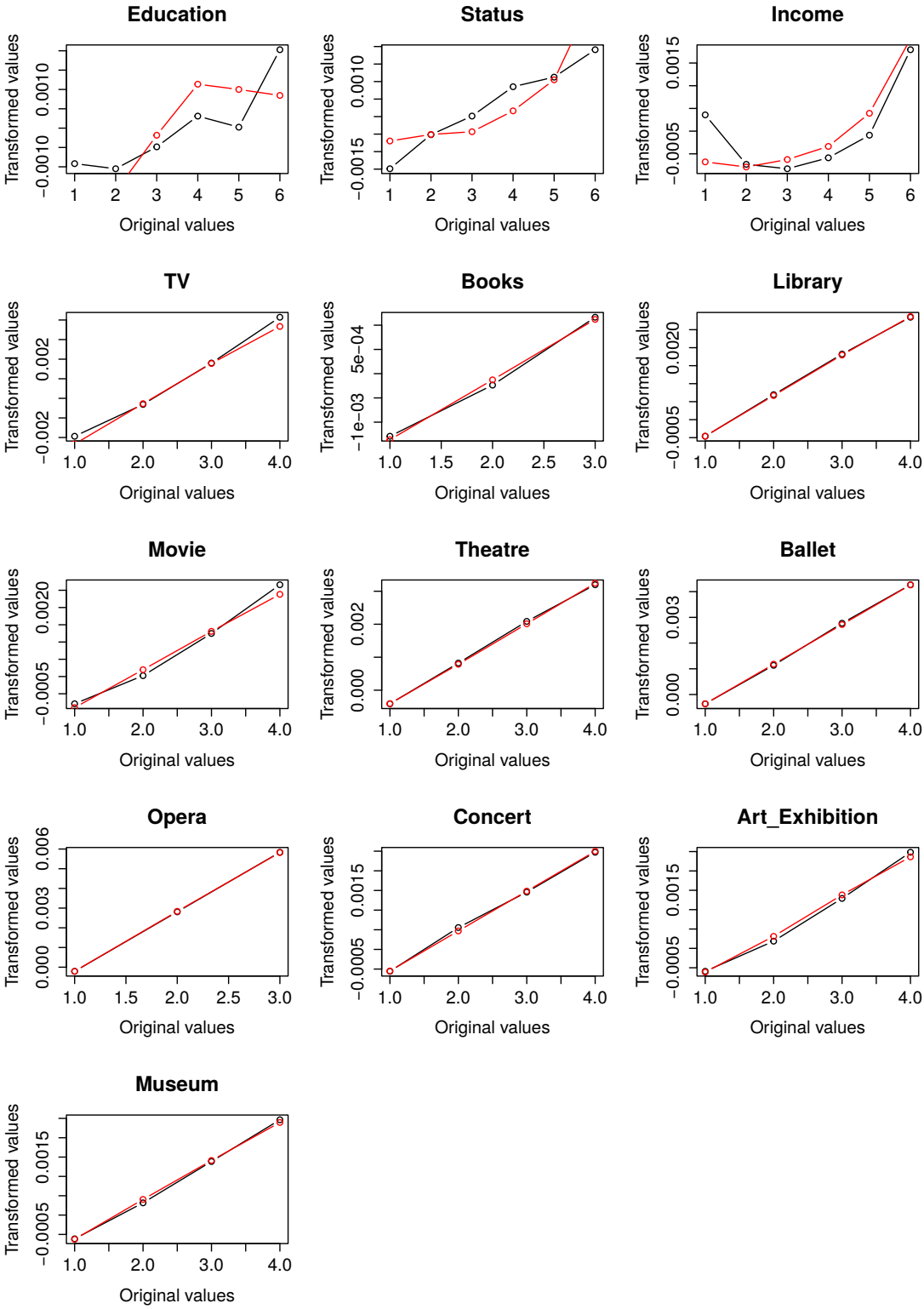

Figure 4: Transformation for France

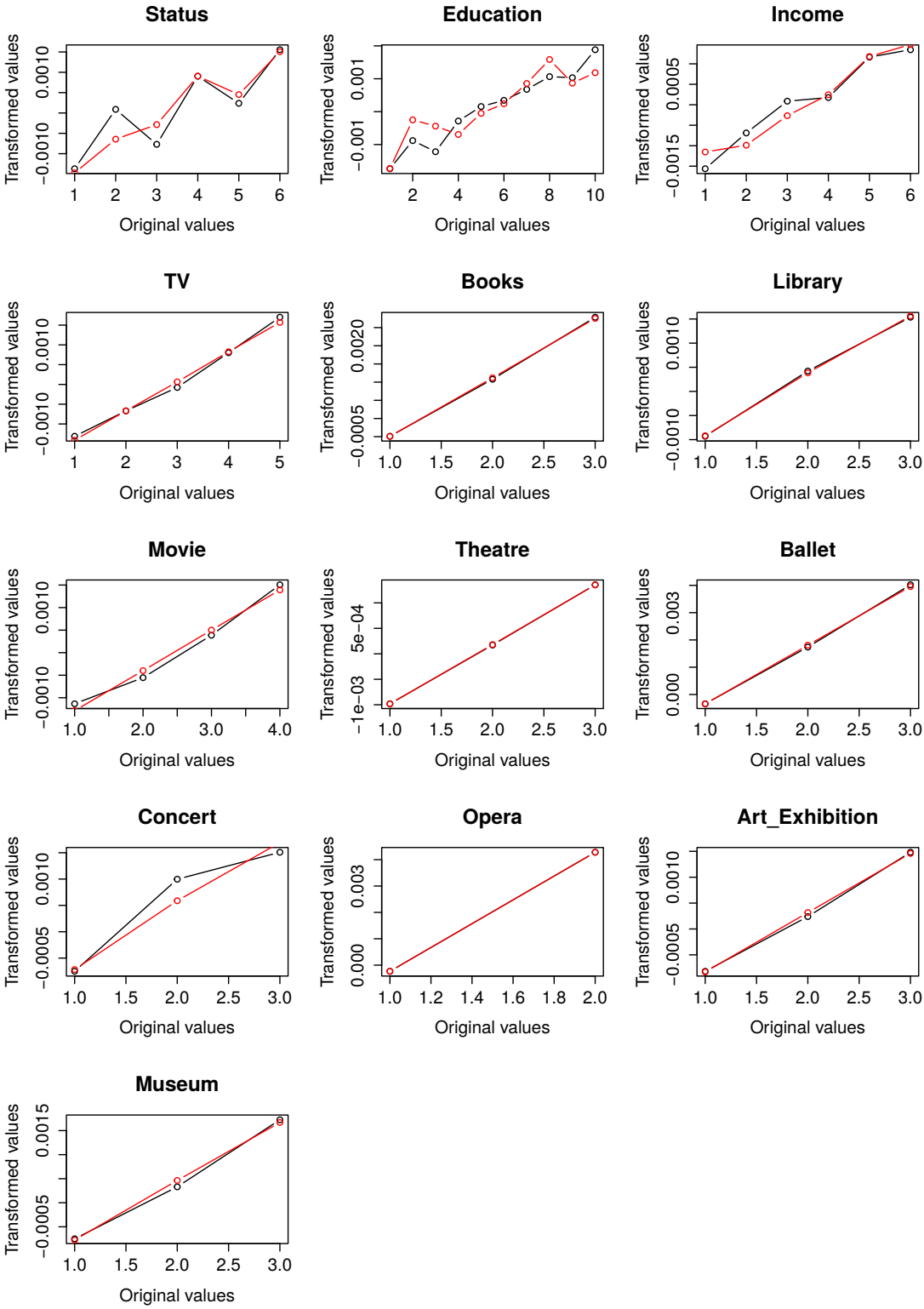

Figure 5: Transformation for Norway

Similarly, for categories at the lower rank of the variables, we did not find differences according to scaling. The only exception is found in France, for lower income, which could partly be due to the interplay between individual and household characteristics. We will not comment in further detail on the multiple nominal solution, but the examinations of the correlations between the canonical variates in the single and the multiple nominal cases show that the differences are extremely small — which is one more argument to stick to the simplest solution, the single ordinal one.

In summary, the Gifi framework applied to the canonical correlation analysis and the principle of parsimony provides us with a clear conclusion: Keep the most parsimonious models, which in our case are the single ordinal models. Second, we have seen that the outcomes of the other models are rather similar. This is an interesting conclusion as it demonstrates the robustness of the methods generally used. It also shows that, with these data, a simple ordinal model is dominating the structure even if for some particular variables, multiple scaling might be appropriate.

5 Discussion

We have analysed social and cultural spaces in three countries: France, Norway and Switzerland. In each country, we used survey data allowing us to construct the social space associated with and according to the homology thesis—theoretically defined by cultural activities and socio-economic position. Our main topic of investigation was to consider the choices of methodology related to previous analyses of this topic. In particular, we have discussed two issues: first, the overall methodology of these analyses, and second, the scaling of the variables used in these analyses.

Previous studies on social and cultural space have started by first representing a cultural "space" based on the multiple correspondence analysis of one set of manifest variables (most often the cultural consumption variables) before adding the indicators of social position as supplementary variables (based on the set of variables measuring social position). This reasons for our methodological choice are seldom explicated in detail. Within the Gifi framework we have, in line with the theoretical argument of the homology thesis, decided to treat the two spaces, social and cultural space, in a direct symmetrical way, which is in line with the homology thesis.

On the scaling of the variables, we have explored different possibilities in order to see if the established research praxis of assuming without discussion that all variables are scaled at the nominal level of measurement (as in MCA) or at the interval level of measurement (as in PCA) is justified. Thus, we have, within the canonical correlation design, explored nominal, ordinal and interval levels of measurements to see which level of scaling is most appropriate for the association of social and cultural sets of manifest variables,

given the principle of parsimony in science. These models are termed "single" models since the scaling, or the transformation of the original codes, was the same across all latent dimensions. In addition, we have also explored if our choice of scaling for the manifest variables may differ across different latent dimensions, i.e. we have compared the single scaling models with the "multiple" models (as it would be in a comparison between PCA and MCA).

In summary, we would like to underline three points:

- Whatever scaling we apply, in all analyses, two latent canonical dimensions emerged in all three countries; where the first set of dimensions were most important. In the French case, the relation between the social and cultural spaces was stronger than in Switzerland and Norway. In France, the importance of the second dimension was larger than in Switzerland and Norway. In other words, the homology thesis is sustained in all the three countries but appears stronger in France.

- In all three countries, looking at both "single" and "multiple" models, the solutions obtained were very close to each other. Thus, the idea that a particular method will induce a particular kind of representation, showing results that would not have been possible to observe with another model, is not confirmed. On the contrary, there was a great coherence of the results between the different models used in this analysis.

- The "single" ordinal model was, however, slightly more adapted to our data by comparison to the others; thus, following the idea of parsimony as stated in the Occam's Razor principles, we suggest that the "single" ordinal model is the best in terms of explanation's economy. Of course, there are also interesting details that can only be discovered by carefully looking at the transformation of the variables. In particular, we have seen that a precise description of the social space could imply close examination of the relative position of education and income for some categories.

6 Conclusions

More generally, we believe that reintroducing the canonical correlation design is useful, in particular for this topic of investigation. However, we also suggest to apply such a design for other topics where two sets of manifest variables are theoretically symmetrically associated. We also believe that "travelling with Gifi" has contributed to the theory of measurement by demonstrating the usefulness of explicit considerations of different scalings.

Furthermore, by comparing three countries, we have provided a more robust test of our findings We have seen that the associations between cultural consumption and social position arerelatively similar across these countries. This result is well in-line with the conclusion of Falk and Katz-Gerro (2015) when they write about cultural behaviour: "There is surprisingly little variation in the influence of education and income across countries." It is also in-line with other comparative studies mentioned in Coulangeon and Duval (2015) and in Lebaron and Le Roux (2015), which also shows fairly similar strength of homology between social and cultural spaces across different national contexts.

The usual approach to this topic of investigation has been first to represent a cultural "space" based on the multiple correspondence analysis of one set of the variables (the cultural consumption variables) and then adding the indicators of social position as supplementary variables (based on the set of variables measuring social position) which introduces an idea of asymmetry between social and cultural spaces. Our paper improves this research strategy by generating the latent dimensions of the social and cultural space simultaneously by applying canonical correlation analysis. Although not vital for our argument, as the conclusion of previous studies are not contested on this point, we do in fact believe that this approach is more appropriate, given Bourdieu's homology hypothesis and the theoretical definition of social space. The canonical correlation analysis is a general analytic and methodological approach that is independent of researchers' choice of scaling.

The Gifi framework integrates different strategies for variable transformations within each model (i.e., "singular" or "multiple"). In line with the usual rules of empirical research we would emphasize this framework as flexible, allowing the researchers to decide which method to use, given the characteristics and properties of the data. This choice has to be, before all, an empirical question. It has been argued that the use of one or another methodology have implications on a more epistemological level. For example, some researchers have argued that the MCA method allows social scientists to find a particular category of results, and to "discover" otherwise invisible phenomena. In this context, Philippe Cibois (1981, our translation) recalls a very illuminating quote of Jean-Paul Benzecri, the 'father' of the correspondence analysis method: "[if we make such analyses], it is in the hope of discovering the very axes of a really existing equilibrium in the World (...), we aspire to discover the hidden properties placed higher in the natural hierarchy of causes than those that are obvious."

Contrary to this view, we have shown, based on robust analyses of survey data comprising three countries, that researchers' choice of methodology is not—and should not be—based on academic habits and commonly applied research methods associated with specific theoretical questions. Rather we would argue in favour of pragmatism when choosing which methodological design would be the most appropriate, given the theoretical ambition,

the data and their measurements. Adjudicating between methodological approaches is also interesting in this regard, as they can be seen as framed in very different scientific traditions that could be important to discuss openly.

Annex:

Variables and number of categories, by country

Switzerland		France		Norway	
N	2442	N	3744	N	1005
Income	5	Education	6	Status	6
Education	6	Status	6	Education	10
Status	6	Income	6	Income	6
TV	4	TV	4	TV	5
Books	5	Books	3	Books	3
Library	5	Library	4	Library	3
Movies	5	Movies	4	Movies	4
Theatre	3	Theatre	4	Theatre	3
Ballet	2	Ballet	4	Ballet	3
Opera	4	Opera	3	Concert	3
Concert	4	Concert	4	Opera	2
Art Exhibition	4	Art Exhibition	4	Art Exhibition	3
Museum	3	Museum	4	Museum	3

7 References

Bergman MM and Joye D (2005). *Comparing social stratification schemas: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman, and Wright*, Cambridge Studies in Social Research, No 10, SSRG Publications

Bennett, T, Savage M, Silva E, Warde A, Gayo-Cal M and Wright D (2009) *Culture, Class, Distinction*, London and New York: Routledge

Bennett T, Bustamante M, Frow J (2015) The Australian space of lifestyles in comparative perspective in Coulangeon P. and Duval J. (2015) *The Routledge Companion to Bourdieu's Distinction*, London: Routledge

Birkelund GE and Lemel Y (2013) Lifestyles and Social Stratification: An Explorative Study of France and Norway. *Comparative Social Research*, 30: 189-220.

Blau P and Duncan OD (1967) *The American Occupational Structure*, New York: Wiley.

Blasius J and Greenacre M (eds) (2014) *Visualisation and Verbalization of Data*, Boca Raton: CRC Press (Taylor and Francis Group)

Bourdieu P (1984) *Distinction: A Social Critique of the Judgement of Taste.* London: Routledge & Keagan Paul.

Bourdieu P (1985) The Social Space and the Genesis of Groups, *Theory and Society*, 14: 723-744.

Bourdieu P (1991) First Lecture. Social Space and Symbolic Space: Introduction to a Japanese Reading of Distinction, *Poetics Today*, 12: 627-638.

Cibois P (1981) Analyse des données et sociologie. *L'Année sociologique*, 1981, 31:333-348.

Chan TW and Goldthorpe JH (2007) Social stratification and cultural consumption: The visual arts in England. *Poetics*, 35: 168–190.

Chan TW (ed.) (2010) *Social Status and Cultural Consumption*, Cambridge: Cambridge University Press

Chan TW, Birkelund GE, Aas AK and Wiborg Ø (2011) Social Status in Norway, *European Sociological Review*, 27: 451-468.

Conrad FG, Brown NR, Cashman ER (1988) Strategies for Estimating Behavioural Frequency in Survey Interviews, *Memory*, 6(4): 339-366.

Coulangeon P. and Duval J. (2015) *The Routledge Companion to Bourdieu's 'Distinction*, London: Routledge

Cousteaux AS and Lemel Y (2004) *Etude de l'Homophilie Socioprofessionelle à travers l'enquête contacts*, CREST Working paper 2004-10.

Cox TF and Cox MAA (1994) *Multidimensional Scaling* London : Chapman and Hall.

de Leeuw J and Mair P (2009) Homogeneity Analysis in R: the Package homals. *Journal of Statistical Software*, 31(4): 1-21.

De Leeuw, J.The Gifi system of nonlinear multivariate analysis. *Data analysis and informatics III* (1984): 415-424.

Erikson R and Goldthorpe JH (1992) *The Constant Flux.* Oxford: Clarendon Press.

Falk M and Katz-Gerro T (2015) Cultural Participation in Europe: Can We Identify Common Determinants? *Journal of Cultural Economics* 40,2: 127-162.

Frie KG and Janssen C (2009) Social inequality, lifestyles and health – a non-linear canonical correlation analysis based on the approach of Pierre Bourdieu, *International Journal of Public Health* 54: 213–221

Ganzeboom HBG, de Graaf PM and Treiman DJ (1992) A standard international socio-economic index of occupational status. *Social Science Research*, 21: 1–56.

Gayo M, Joye D, Lemel Y (submitted) Testing the universalism of Bourdieu?s homology thesis: structuring patterns of leisure participation across 26 countries.

Gifi A (1989) *Algorithm Descriptions for ANACOR, HOMALS, PRINCIPALS and OVERALS.* Department of Data Theory, University of Leiden.

Gifi A (1990) *Nonlinear Multivariate analysis.* Hoboken: Wiley

Gittins R (1985) *Canonical analysis A Review with Applications in Ecology*, Berlin: Springer

Joye D and Chevillard J (2011) Identités, inégalités et changement social: les enjeux de la mesure. In: Burton-Jeangros C and Maeder C (eds.) *Identité et transformation des modes de vie. Identität und Wandel der Lebensformen.* Zurich: Seismo, pp. 184-207.

Joye D and Chevillard J (2013) Education, Prestige and Socioeconomic Indexes in Switzerland. In: Becker R, Bühler P and Bühler T (eds.) *Bildungsungleichheit und Gerechtigkeit.* Bern: Haupt, pp. 163-177.

Le Roux B, Rouanet, Savage M, Warde A (2008) Class and Cultural Division in the UK, *Sociology*, 42(6): 1049-1071.

Levine MS, 1977 *Canonical Analysis and Factor Comparison*, Beverly Hills, London, New Dehli: Sage

Mair P and de Leeuw J (2010) A General Framework for Multivariate Analysis with Optimal Scaling: The R Package aspect. *Journal of Statistical Software*, 32(9): 1-23

Michailidis G and de Leeuw J (1988) The Gifi system for Descriptive Multivariate Analysis. *Statistical Science*, 13:307-336

Nagel, I and Lemel Y, 2016 (submitted) The effects of parents' lifestyle on their children's status ttainment and lifestyle in the Netherlands

Oesch, D. (2006) *Redrawing the Class Map*, Basingstoke: Palgrave Macmillan,

Peterson, R A 1992. Understanding audience segmentation: From elite and mass to omnivore and univore. *Poetics*, 21: 243-258.

Prieur A, Rosenlund L and Skjott-Larsen J, 2008, Cultural capital today: A case study from Denmark, *Poetics*, 36: 45–71.

Rosenlund L (2015) Working with Distinction: Scandinavian experiences in Coulangeon P and Duval J (eds) (2015) *The Routledge Companion to Bourdieu's 'Distinction*, London: Routledge.

van der Heijden PGM and Sijtsma K (1996) Fifty Years of Measurment and Scaling in the Dutch Social Sciences, *Statistica Neerlandica*, 50, 1, 111-135

Wright, EO, 1985. *Classes.* London: Verso.