Soc Indic Res (2016) 126:1065–1087
DOI 10.1007/s11205-015-0932-7

A Measure of the Cultural Distance Between Countries

Gustavo De Santis[1] · **Mauro Maltagliati**[1] · **Silvana Salvini**[1]

Accepted: 6 March 2015 / Published online: 17 March 2015

Abstract We present a new method for evaluating the relative distance between any two countries, among several, using individual data. We form clusters of respondents and we calculate the proportions of each country's respondents who belong to the various clusters. Assuming that respondents in the same cluster are similar to one another and that two countries are close to each other when their nationals distribute similarly among clusters, the dissimilarity between countries can be expressed in terms of Euclidean distances between the observed distributions (the square root of the sum of the squared differences between the proportions of nationals in the same cluster). We test the method on the World Value Survey (WVS) dataset for the years 1994–2007, first separately, by "domain" (opinions and attitudes on, e.g., religion, politics, and family), and then on all of the (selected) variables together. Groups of assumedly similar countries (the Baltic, the Nordic and the Mediterranean countries) turn out to be closer to each other than do, on average, any two European countries picked at random, which lends credibility to the method. Its pros and cons are discussed in the final section.

Keywords Clusters · Cultural distance · Multidimensional scaling · European countries

✉ Gustavo De Santis
desantis@disia.unifi.it

Mauro Maltagliati
mmaltag@disia.unifi.it

Silvana Salvini
salvini@disia.unifi.it

[1] DiSIA - Department of Statistics, Informatics, Application "G. Parenti", University of Florence, Viale Morgagni 59, 50134 Florence, Italy

1 Introduction: The Purpose of this Paper

This paper is about how to assess the relative distance between any two (or subgroup of) countries within a larger set of countries when individual data are available. The method requires data at the individual or household level for a (preferably large) sample of respondents, but is very flexible in terms of what data may be used: in this application, for instance, we had qualitative data on values and attitudes and we treated them as if they were all nominal, which is the worst possible situation. With quantitative data, as explained below, the method is easier to apply.

The results of our empirical application are not only methodologically encouraging (the proposed approach seems to work well, all in all) but also, we believe, interesting in themselves. If the method does what it is expected to do, that is, if it actually identifies similarities (in our case: cultural similarities) between countries, it permits scholars to create "objective" taxonomies, or to evaluate critically the existing ones. Indeed, creating typologies of countries has become so common that in most cases the statistical procedure (clustering) on which it is based does not stand out clearly, despite the fact that it may not be neutral and may bias both the results and their interpretation. For instance, the UN customarily speak of more, less and least developed countries (UN 2013); Esping-Andersen (1990, 1996) introduced the idea of "the three worlds of welfare capitalism"; and numberless country rankings exist, based on gender gap (World Economic Forum 2005), economic freedom (http://www.heritage.org/index/), democracy (http://www.economist.com/node/8908438); military strength (http://www.globalfirepower.com/countries-listing.asp), etc. Ranking is very close to clustering: both are normally based on the combination of a set of elementary indicators. Besides, the normal use of rankings is to take the first N countries of a list and see how they differ from, say, the following M, or the last Z of the same list, or the first N some time before, or taken from a different list. In these cases, the result of the (implicit) clustering procedure has an interest in itself.

In other instances scholars may be interested in a specific phenomenon (for instance, fertility or couple formation) or in the effect of a policy (for instance, a fiscal or a family policy), and may want to compare this phenomenon or this policy in two or more countries that are assumed to be otherwise similar, on the basis of their geographical proximity, or of one or more of the above-mentioned taxonomies (e.g., Aassve et al. 2007; Hoem et al. 2010; Reigner-Loilier and Vignoli 2011; Kapitány and Spéder 2012; Christiansen and Keilman 2013).

These taxonomies are typically based on objective indicators of various types: economic (e.g., income or unemployment), demographic (e.g., mortality), institutional (e.g., form of government), etc. If several indicators are available, and especially if the indicators are heterogeneous, the need arises to provide a synthesis, with multidimensional analysis. To this end scholars normally use the averages of the selected indicators (average income, proportions old, etc.): this is the case of Cattell (1950), for instance, a forerunner in the "effort to identify clusters of societies using the analysis of international-level [cultural] data" (Guptaa et al. 2002:11). Refinements of this basic idea have led to principal component analysis or other forms of linear combinations of indicators, assumedly homogeneous and capable of measuring some latent social dimension, e.g. social cohesion (Dickes and Valentova 2013).

This procedure has a few shortcomings, however: comparable averages can derive from widely differing distributions (think for instance of income, or the number of children born to each woman), hiding the fact that the compared countries may be less similar than the

simple consideration of averages would suggest. Besides, by definition, qualitative indicators cannot yield averages, a problem that is frequently solved (but also, perhaps, oversimplified) by creating dichotomies: e.g., "proportion of people reporting poor or very poor health conditions."

The method that we present here overcomes both limitations: it uses distributions of individuals, not average values, and can be applied to qualitative date. Its basic steps are the following:

1. On the basis of a set of relevant variables, form N clusters of individuals, where those who belong to the same cluster are *assumed* to be similar among themselves.
2. For each country, calculate the proportions of individuals in the various clusters K_1, K_2,…, K_n.
3. Use these proportions to assess how close, or how far apart, each country is from each other: the closer the proportions, the greater the proximity.

The proposed approach has several points of strength: it is very simple, it combines two consolidated statistical techniques (cluster analysis, plus Euclidean distances and multidimensional scaling to assess distances) in an original way, and it uses clusters for what they are: groups of supposedly homogeneous entities. However, the method has some weaknesses, too: the choice of the variables is crucial and is often guided by data availability more than by anything else; very many clustering criteria can be applied, each normally leading to a different outcome; and the results may depend on the number of clusters, the choice of which is, unfortunately, in large part arbitrary.

Since the method is new, we did not simply stress its formal properties and its rationale, but we also "tested" its results against a few assumedly sensible expectations. On our case the expectation was that some countries (the Mediterranean, the Nordic and the Baltic ones) would turn up to be culturally homogeneous, that is, with our metric, closer to one another than any two European countries picked at random. As we will see shortly this is what generally, although not always, happened depending on the cultural domain under examination, and on the number of clusters that we formed. To be sure, this empirical result does not prove (or, for that matter, disprove) the validity of the method, but it can be considered as a mild corroboration to our claim that the method is worth including among the tools of social scientists.

The paper is structured as follows: in Sect. 2 we give the details of the method and we present our data. In Sect. 3 we illustrate our results. Section 4 draws the conclusions, discussing the limitations, the advantages and the possible further uses of the proposed approach.

2 Methodology: Using Distributions to Compare Countries on Individual (Ideational) Variables

2.1 Overview of the Method

The proposed procedure uses individual (not aggregate) data for assessing the similarity between countries, and is based on the idea that two (or more) countries can be considered close to each other when the *distribution* of their citizens among the various modalities of a given variable is similar, and not just when the average or any other synthetic indicator, is comparable.

However, we did not look at each distribution separately, for two main reasons. The first is that it is possible to obtain as many distributions as there are variables, which still leaves the problem of the synthesis unresolved. The second reason is that several manifest variables often pertain to the same domain (or latent variable),[1] and there may be some sort of compensation within domains. For instance, an individual who does not have a savings account may possess government bonds, and since both are assets, looking at the two items together gives us a better idea of the person's economic standing. This "nuanced" approach becomes unavoidable when the variable of interest is by definition a latent (not directly observable) one: e.g., involvement in social activities or gender attitudes.

The approach that we followed here can be broken down into the following steps:

1. Focus on a few (in our case, ten), presumably relevant domains (or latent variables), such as religious attitudes or (the importance attributed to) the family;
2. select a few observable (manifest) indicators of each domain;
3. for each domain, form clusters of respondents based on the corresponding observable variables (answers given to questions that relate to that domain) but not on the respondents' nationality;
4. for each country in the dataset, calculate the distributions of nationals by cluster; and
5. use these distributions to assess how close or how far apart any two countries are (Euclidean distances between distributions).

The final step assumes that:

(a) the individuals who belong to the same cluster can be considered similar to each other with respect to that specific latent variable (e.g., politics), and different from the individuals who belong to a different cluster, which is true almost by definition; and
(b) the distribution of nationals among the clusters reveals something about the relative standing of (the citizens of) a country on that dimension, and, most importantly for our purposes, the comparison of these country-specific distributions of nationals ("profiles") indicates how close or how far apart any two countries are.

One of the advantages of the proposed method is that it does *not* demand that these clusters be "labelled", which constitutes a remarkable improvement over the standard use of multidimensional techniques. Normally, when clusters are formed, it is useful, if not essential, to understand what each cluster represents, which introduces the need for "labels." With reference to, for instance, religion, it is hard to see the utility of forming clusters if one cannot come up with descriptions like: "In cluster 2 there are very religious people," or "Cluster 6 groups individuals who formally adhere to religious precepts, but are in practice relatively detached." Our method, conversely, merely assumes that individuals in the same clusters are similar to each other, with respect to religious attitudes, and different from those who belong to other clusters, without the need of describing, or even understanding, their basic characteristics.

[1] In this paper we speak of indicators (or manifest variables) to denote variables that can be observed (i.e. the answers to the questions of the VWS—see Sect. 2.4) and domains (or latent variables, or dimensions) to denote variables that cannot be directly observed or measured, but that we assume exist and influence the corresponding indicators. For instance "religious attitude" is a latent (not directly observable) variable (or domain, or dimension), about which something can be said by observing a few empirical indicators (answers to such questions as "How often do you go to mess?" or "Is religion important in the education of a child?").

Another relevant point is that we did not introduce any assumptions regarding the (relative) distance between clusters: the respondents belonged to either the same or another cluster. The relative distance between clusters was implicitly accounted for by varying the number of clusters, which we did as a robustness check (see Sect. 3): our conclusions are based solely on the domains that yielded (basically) the same results, irrespective of the number of clusters considered.

2.2 A Practical Application (to Better Understand How the Method Works)

For the practical application of the proposed approach we exploited the "ideational" variables of the World Value Survey (WVS), described in Sect. 2.4. A few scholars argue that these are the most important variables in a post-industrial, post-materialistic world (Inglehart 1971, 2008), in the middle of its second demographic transition (Sobotka 2008; Lesthaeghe 2011). According to this line of thinking, several choices and decisions, including the demographic ones (e.g., fertility, couple formation and dissolution, family ties, support, and obligations), are driven by value orientations, personal feelings and subjective opinions more than by "hard facts" (e.g., levels of income and education). Those who are more skeptical about the relevance and comparability of this type of data may simply consider the present application as an example of how the proposed method works.

As the number of variables in the WVS dataset is very large, about a thousand, a few preliminary strategic decisions proved necessary to make our analysis more manageable. We decided to focus only on *observed variables* that could be linked to a relatively well-identifiable domain, or *latent variable*. After careful examination of the characteristics and purposes of the dataset (see Sect. 2.4), we identified ten such domains:

1. Ethics
2. Family (traditional view of)
3. Friends and associations (importance attributed to)
4. Gender attitudes (ideas and attitudes regarding gender roles)
5. Happiness/health (subjective feeling/evaluation of)
6. Politics (importance of)
7. Politics (left/right)
8. Religion
9. Self (care of, and reliance on, one's self)
10. Work

The choice and even the labelling of these domains was, in our case, data driven, in several senses. First, we wanted to keep all of the originally available 21 European countries (listed below), and since not all the questions were asked in all of these countries, we selected only the variables that were surveyed everywhere in Europe (at least once). Secondly, we retained only the combination of manifest and latent variables that better reflected the purpose and the nature of the World Value Survey, as explained in the online documentation of the survey itself: secularism and religion (Li and Bond 2010), life satisfaction (Eichhorn 2012), materialism, post-materialism and happiness (Delhey 2009), etc. Further, we checked that the empirical indicators that we had identified had sufficiently few missing values and a sufficient internal consistency (through Cronbach's alpha) to be considered indicators of a common latent domain.

Since the method is new, to the best of our knowledge, we felt the need for some corroboration for our results. Therefore, we decided to focus particularly on three sets of countries:

(a) the Mediterranean (although, in practice, only Italy and Spain were available),
(b) the Nordic (Finland, Norway, and Sweden), and
(c) the Baltic (Lithuania, Latvia, and Estonia).

These blocks of countries are generally supposed be homogeneous: for their geographical proximity, because they belong to the same welfare typologies (Esping-Andersen 1990, 1996), have comparable family systems and values (Reher 1998) and, in the case of the Baltic countries (group c), because they have been politically very close until recently, although not necessarily with mutual satisfaction. The expectation was that the measures of cultural distance that our method would produce for the countries within these groups (both separately by domain, and globally) would normally be below the European average, thus signalling proximity and (relative) homogeneity. Of course, this outcome would not *prove* that the approach is valid: it would merely corroborate our claim that the approach has some merit.

Separately from the proposed methodology, we also tried to provide an idea of the relative positions within Europe of each of the countries in each particular domain. For example, in the domain of religion, as we will see shortly, Sweden and Norway turned out to be close to each other. Does this mean that they are, by European standards, very religious, scarcely religious, or just average? Since all of the questions of the WVS were qualitative, it was impossible to arrive at an objective summary measure. We therefore resorted to providing a rather rough approximation, which we obtained by (arbitrarily) associating each answer with a numerical value that was meant to "measure" how high or low each respondent (and therefore, on average, each country) scored on each latent dimension; in this case, religion. We remind our readers, however, that we did *not* use this metric in clustering and in measuring distances. We simply tried, at the end of the procedure, to give an idea of what we had found, with the goal of guiding our readers through the multidimensional world that we created.

2.3 Between Jaccard and Ward

As the next session will show, our manifest variables are qualitative: in some cases ordinal, in others nominal. We decided to treat them all as nominal, because any other choice would have introduced a metric. For instance question A001 ("How important is each of the following in your life? Family") admitted four possible answers

1. Very important
2. Rather important
3. Not very important
4. Not at all important.

These answers are clearly ordinal, and "1" (Very important) is closer to '2' (important) than it is to '3' or '4'. Ignoring this ordering implies that part of the information gets lost. However, since it is impossible to say by how much '1' is closer to '2' than it is to '3', and since it is also arbitrary to assume that '2' is as far from '1' as it is from '3', we eventually decided to treat them as nominal: each answer is simply different, and equally different, from each other. To do this, we dichotomized the answers, as in the example below.[2]

[2] We also tried a few alternative specifications, introducing some arbitrary metrics. The results, not shown here, are in line with (actually, even slightly better than) those that will be presented shorty. All the elaborations were run with the SAS software, release 9.3.

The standard way of transforming nominal variables into a measure of distance is to start from the Jaccard index, or coefficient, of similarity J ($0 \leq J \leq 1$), which is of the type

$$J_{A,B} = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where the meaning of the symbols of intersection ($\cap$) and union ($\cup$) will be explained shortly.

When only one answer is admitted among several (the standard case with our dataset) the Jaccard index for every manifest variable is either 1 or 0. For instance, take the question on the importance of the family and imagine two individuals A and B giving different answers, e.g. '2' (Rather important) and '3' (Not very important), respectively.

After dichotomization (0 = answer not chosen; 1 = answer chosen), the information on A and B appears as follows

(A) 2 => 0 1 0 0
(B) 3 => 0 0 1 0

where, for both A and B, the first column, full of 0′s, says that neither answered "very important"; the second column tells us that A answered "rather important" but B did not; etc. The denominator of the Jaccard index is 2 (the two columns with underlined figures, i.e. the union $\cup$ of the answers actually chosen in this case—those selected by at least one of the respondents) while the numerator is 0 (the intersection $\cap$ is an empty set: there are no common answers between the two respondents).

In a couple of cases, however, the possible answers where two out of four. (E.g. "What should the two most important aims of this country be for the next 10 years?" '1—A high level of economic growth'; '2—Strong defence forces'; '3—People have more say about things'; '4—Trying to make our cities and countryside more beautiful').[3]

In this case, the answers can be either exactly the same (J = 1), or totally non coincident (J = 0), or partly coincident, as in the example for the two imaginary respondents C and D below

(A) C) 2,3 => 0 1 1 0;
(B) D) 3,4 => 0 0 1 1.

The $J_{C,D}$ (the Jaccard proximity between C and D) has 3 in the denominator (the union $\cup$ of the three columns with the underlined figures, where at least one of the respondents has "1"), and 1 in the numerator (the intersection $\cap$ is, in this case, the only case when the answers coincide), and so $J_{C,D} = 1/3$.

The Jaccard proximity index must be transformed into a (Jaccard) measure of distance for the clustering procedure to work. This is customarily done by simply taking its complement to 1, as follows

$$D_J = 1 - J \tag{2}$$

The application of this procedure generated a matrix of distances for each domain, on which we constructed clusters. This could have been done in various ways (minimum

[3] In the survey these were two separate questions, on the first and the second most important aim (questions E001 and E002). Since Jaccard multiplies the possible combinations of cases, which rapidly exceeds computing capacity, we had to circumvent this limitation in various ways: in this case we merged the answers to the two separate questions, ignoring the difference between the first and the second most important aim of the country, in the respondent's opinion.

distance, maximum distance, centroid, etc.), not necessarily leading to the same results. We eventually adopted Ward's (1963) criterion of minimum variance, both for its intrinsic logic and because it tends to distribute observations more evenly among clusters, which constitutes an advantage in general, and especially for our method, which is based on frequencies.

As mentioned, the Jaccard method leads in principle to the creation of a square matrix of distances, of each observation from each other. Since we have about 39 thousand respondents, we should have created a matrix of 39.000^2 elements—far beyond our computing capacity. When working on each single domain (with relatively few manifest variables and a limited range of possible answers), we circumvented this difficulty by creating a grid of all possible cases, with their corresponding distances (of each from each other), and used this grid to create our clusters, to which we subsequently assigned our observations.[4]

However, even this solution has its limitations, because the grid must be kept within certain limits. This is manageable, sometimes with difficulty, for each domain separately, but not for all of the variables and domains together.

Since we also wanted a general view of the distances between countries (with all the domains considered together), we also followed an alternative, somewhat less correct, path: we constructed clusters (with the Ward criterion) directly on the dichotomous variables that we had created, first separately, by domain, and then on all of the variables together—although this is not the recommended use of Ward's minimum variance criterion for clustering.

We first checked that the results by single domain obtained in the two ways (with and without Jaccard) were comparable. Since this was in fact the case (not shown here), we *assumed* that the general summary measure obtained with the Ward criterion (directly applied to the 58 original variables, all transformed into dichotomous, but without the Jaccard measure of distance) could be considered a reasonable proxy for the impossible-to-obtain correct summary measure (i.e. with clusters based on the Jaccard distance applied to all of the 58 manifest variables together). Readers should keep this assumption in mind when they get to the summary measures of Sect. 3.4.

2.4 The World Value Survey (WVS) Dataset

Our data comes from the dataset of the World Value Survey (WVS; http://www.worldvaluessurvey.org/). It contains five waves of (basically) the same survey for the years 1981–2007. We focused on the European countries, of which we found 21, not all of them present in all waves.

However, we eventually decided to drop the first two waves because they referred to years that were too remote (1981–1993), when the socioeconomic, political, and cultural contexts were considerably different from those of today, especially in Eastern Europe. Moreover, only a few of the countries were surveyed during these years. Eventually, we retained slightly more than 39,000 observations from 21 countries in the years 1994–2007 (Table 1).

As for the selection of the variables (both the latent and the manifest ones), as mentioned, our choice was largely data-driven, since we wanted to keep all of the 21 European countries, and we wanted only latent domains that could be associated with several (at least four) manifest variables. We also kept into consideration the spirit of the WVS and its

[4] Take, for instance, the domain "Gender attitude" (Sect. 2.4). It has five corresponding manifest variables, the first with 3 possible answers, the others with 4. This makes $3 \times 4^4 = 768$ theoretical different typologies (some of which empty, to be sure), and a corresponding matrix of $768^2 = 589{,}824$ (theoretical) distances, which is very large, but still manageable.

Table 1 Number of observations in the WVS file, 3 waves (1994–2007)

Label	Country	Wave			Total
		1994–1998	1999–2004	2005–2007	
Mediterranean					
IT	Italy	–	–	1012	1012
ES	Spain	1211	1209	1200	3620
Nordic					
FI	Finland	987	–	1014	2001
NO	Norway	1127	–	1025	2152
SE	Sweden	1009	1015	1003	3027
Baltic					
EE	Estonia	1021	–	–	1021
LV	Latvia	1200	–	–	1200
LT	Lithuania	1009	–	–	1009
All others					
BG	Bulgaria	1072	–	1001	2073
HR	Croatia	1196	–	–	1196
CZ	Czech Rep.	1147	–	–	1147
FR	France	–	–	1001	1001
DE	Germany	2026	–	2064	4090
HU	Hungary	650	–	–	650
NL	Netherlands	–	–	1050	1050
PL	Poland	1153	–	1000	2153
RO	Romania	1239	–	1776	3015
SK	Slovakia	1095	–	–	1095
SI	Slovenia	1007	–	1037	2044
CH	Switzerland	1212	–	1241	2453
GB	Great Britain	1093	–	1041	2134
	All	20,454	2224	16,465	39,143

Source: own elaborations on data from the WVS website (http://www.worldvaluessurvey.org/), as of April 2013

focus: on secularism (Li 2010), life satisfaction (Eichhorn 2012), etc. Finally, we checked the internal consistency of the variables that we had selected as manifest indicators of the latent domains that we were interested in (with Cronbach's alpha). After a process of trials and errors, we eventually retained the ten domains that are listed in Table 2, with their corresponding manifest variables.

3 Analysis by Domain

3.1 Subjective Assessment of Happiness and Health

In the following, we outline in detail how we analyzed a specific domain, the subjective assessment of happiness and health of the respondents. All other domains were treated in the same way.

Table 2 Latent variables (domains) and their corresponding manifest variables

Domain 1 ethics (4 MF, 0)	*Domain 2 family (traditional view of)* (5 MF, 0*)
F114. Claiming government benefits to which you are not entitled. Is this justifiable?	A001. How important is the family in your life?
F115. Avoiding a fare on public transport. Is this justifiable?	A042. Should children be encouraged to be obedient?
F116. Cheating on taxes. Is this justifiable?	D023. Do you approve of women becoming single parents?
F117. Someone accepting a bribe. Is this justifiable?	D054. Is making your parents proud one of the main goals in your life?
	E019. In the future, should more emphasis be put on family life?
Domain 3 friends and associations (3 MF, 1*)	*Domain 4 gender attitudes* (5 MF, 0*)
A002. How important are friends in your life?	C001. When jobs are scarce men have more right to a job than women. Agree?
D055. Do you try to live up to what your friends expect?	D057. Being a housewife is just as fulfilling as working for pay. Agree?
ACTIVE*[(a)]. Are you an *active* member of one or more of the following: Church or religious organization, Sport or recreational organization, Art, music or educational organization, Labor Union, Political party, Environmental organization, Professional association, Humanitarian or charitable organization, Any other organization (please specify): …	D059. Men make better political leaders. Agree?
	D060. University is more important for a boy than for a girl. Agree?
	E069_15. How much confidence do you have in the women's movement?
Domain 5 Happiness and Health (subj.) (4 MF, 0*)	*Domain 6 Politics (importance of)* (5 MF, 1*)
A008. Would you say you are feeling happy?	A004. How important is politics in your life?
A009. How would you describe your state of health?	A165. Most people can be trusted. Agree?
A170. How satisfied are you with your life?	E023. Are you interested in politics?
C006. How satisfied are you with the financial situation of your household?	E025. Political action: would you sign a petition?
	E069*.[(b)] How much confidence do you have in: Labor unions, Justice system, Government, Political parties, Parliament, The European Union, The United Nations.
Domain 7 Politics (Left/right orientation) (8 MF, 3*)	*Domain 8 Religion* (7 MF, 0*)
C002. Employers should give priority to nationals	A006. How important is religion in your life?
E001_2.* Aims of country: 1st & 2nd choice	A040. Should children be encouraged to be religious?
E003_4.* Aims of respondent: 1st & 2nd choice	A098. Are you a member of a religious organization or church?
E005_6.* Most imp. things to do for society: 1st & 2nd	E069_01. How much confidence do you have in the churches?
E012. Are you willing to fight in war for your country?	F028: How often do you attend religious services?
E033. Self positioning in political scale (left/right)	F034. Would you define yourself as a religious person?
E035. Should incomes be more or less equal?	F063. How important is God in your life?
E041. How close do you feel to each of these sentences "People can only get rich at the expense of others" vs. "Wealth can grow so there's enough for everyone"	

Table 2 continued

Domain 9 Self[c] (9 MF, 0*)	*Domain 10 Work* (4 MF, 0*)
A003. How important is leisure time in your life?	A005. [d] How important is work in your life?
A005.[d] How important is work in your life?	C009. First choice, if looking for a job
A029. Importance of independence as a child quality?	C010. Second choice, if looking for a job
A030. How important is hard work as a child quality?	E040: Does hard work bring success or is it a matter of luck and connections?
A032. How important is the feeling of responsibility as a child quality?	
A034. How important is imagination as a child quality?	
A039. How important are determination and perseverance as a child quality?	
A173. How much freedom do you feel?	
E039. Is competition good or harmful for individuals and societies?	

Notes: The exact wording of the questions (of which we show the original number), the possible answers to them and the corresponding frequency distributions are shown in De Santis et al. (2014)

MF Manifest variables (of which * = complex variables, obtained by combining other elementary variables)

[a] "ACTIVE" measures *active* participation in social life. It ranges between 0, meaning no active participation at all, and 9, meaning maximum possible active participation

[b] Sum of all answers, each ranging between 1 (greatest confidence) and 4 (complete distrust). The variable "E069" (Confidence), thereby, ranges between 7 (greatest confidence) and 28 (complete distrust), but answers have been grouped into 4 classes, based on quartiles

[c] Care of, and reliance on, one's self

[d] used for two domains: "Self" and "Work"

Source: http://www.worldvaluessurvey.org/, as of April 2013

In the survey there were four elementary manifest variables that could be associated with the latent variable "subjective assessment of happiness and health" (Table 3):

The rank correlations between the answers given to these questions have the expected sign, and are extremely significant (not shown here); besides, the corresponding standardized Cronbach's alpha is high (0.76). In short, these statistical measures corroborate our claim that the selected variables tend to measure more or less the same (latent) domain.

Based on their answers, the respondents could be classified into groups (clusters), as explained in Sect. 3. The respondents who belonged to the same group (cluster) were considered similar to one another and equally distant (and therefore equally different) from the members of the other clusters. As for the number of clusters to be formed, we tried several alternatives (5, 10, 15, 20, 30, and 50 clusters), and we decided to show how the outcome differed depending on this choice. When the outcome varied (significantly) with the number of clusters, we accepted that the results were not robust and that no firm conclusion could be reached regarding the distance between countries in that domain.

For any given number of clusters, each individual belonged to a specific cluster, and nations could be characterized by their "profile;" that is, by the proportions of their nationals who belonged to each of the N (= 5, 10,…, 50) existing clusters. The sum of these proportions for all of the countries must of course be one. With 10 clusters, for instance, we obtained the distribution of Table 4, where one can note that 1.1 % of the Italian respondents belonged to the first cluster, 17.3 % belonged to the second cluster, etc. Meanwhile, 1.2 % of the Spanish respondents belonged to the first cluster, 15.6 % belonged to the second cluster, etc.

Table 3 Happiness and health (subjective assessment). Four variables

A008. Feeling of happiness Taking all things together, would you say you are (read out and code one answer): 1. 'Very happy' 2. 'Quite happy' 3. 'Not very happy' 4. 'Not at all happy'	*A009. State of health (subjective)* All in all, how would you describe your state of health these days? Would you say it is: 1. 'Very good' 2. 'Good' 3. 'Fair' 4. 'Poor'
A170. How satisfied are you with your life All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied" where would you put your satisfaction with your life as a whole? (Code one number): 1. 'Dissatisfied' … 10. 'Satisfied'	*C006. Satisfaction with the financial situation of household* How satisfied are you with the financial situation of your household? Please use this card again to help with your answer (code one number): 1. 'Dissatisfied' … 10. 'Satisfied'

Source: WVS website (http://www.worldvaluessurvey.org/), as of April 2013

Assuming that two countries with similar proportions of individuals in clusters 1, 2, etc., were also similar to each other, we considered for each country the shares of people belonging to the same cluster and interpreted these shares as coordinates of a point (country) in an N-dimensional space, where N is the number of clusters. In the case of Table 4 (Happiness, ten clusters), Italy, for instance, is a point in a 10-dimensional space with coordinates (0.011, 0.173,…, 0.077). If each country is a point, the distance between points can be computed as a Euclidean distance, which is simply the multidimensional extension of the Pythagorean theorem (Leti 1983).

In a bi-dimensional space this is easy to see: let d_1 and d_2 be the difference between, respectively, the abscissae and the ordinates of any two points A and B. Their distance D is the length of the third side (hypotenuse) of the right triangle with sides d_1 and d_2, and therefore $D = \sqrt{d_1^2 + d_2^2}$. In an N-dimensional space the idea is the same, and the formula becomes $D = \sqrt{d_1^2 + d_2^2 + \cdots + d_n^2}$. In this case (Happiness, ten clusters) our D's turned out to be those of Table 5. Of course, these distances do not mean much in themselves: they must be interpreted in relative terms, for instance in comparison to the general average, and keeping into account the general variability of the phenomenon (Table 6).

Table 6 tells us that the average Euclidean distance between European countries is 0.221. The average distance between the Mediterranean countries is much lower (0.044), which indicates very strong proximity. Proximity, albeit not equally strong, characterizes also the Nordic countries (average distance = 0.120) and the Baltic ones (0.127). In all of these cases, the average within-group distance is markedly smaller than the average distance between any two European countries (0.221). In short, the countries of these groups appear to be close to each other, and therefore homogeneous in this domain (happiness).

In general, the number of clusters that one forms may affect the results, both in terms of absolute distances (Fig. 1) and standardized distances (Fig. 2).[5] In our case, fortunately, the almost horizontal lines of the two figures, and especially of Fig. 2 (standardized

[5] A standardized variable z can be obtained from an original variable x with the following transformation $z = (x - A)/\sigma$, where A is the average of x, and σ its standard deviation. By construction, a standardized variable z has average $A_z = 0$ and standard deviation $\sigma_z = 1$. Ideally, with results independent of the number of clusters, one should obtain (roughly) straight lines z for each group.

Table 4 Distribution of respondents by country and cluster (21 European countries, Domain = Happiness, Clusters = 10)

	Cluster										All
	1	2	3	4	5	6	7	8	9	10	
Mediterranean											
IT	0.011	0.173	0.081	0.525	0.015	0.108	0.004	0.001	0.006	0.077	1.0
ES	0.012	0.156	0.098	0.516	0.025	0.102	0.003	0.003	0.013	0.072	1.0
Nordic											
FI	0.011	0.218	0.067	0.405	0.032	0.152	0.005		0.001	0.110	1.0
NO	0.005	0.138	0.041	0.373	0.027	0.213	0.010	0.001	0.002	0.190	1.0
SE	0.006	0.158	0.036	0.395	0.023	0.159	0.005	0.002	0.003	0.213	1.0
Baltic											
EE	0.050	0.289	0.285	0.276	0.045	0.038	0.002	0.001	0.008	0.007	1.0
LV	0.032	0.376	0.227	0.274	0.050	0.025		0.003	0.004	0.010	1.0
LT	0.045	0.207	0.341	0.297	0.038	0.055	0.001	0.002	0.002	0.011	1.0
All others											
BG	0.083	0.141	0.314	0.269	0.033	0.103	0.004	0.006	0.016	0.031	1.0
HR	0.031	0.217	0.255	0.265	0.053	0.130	0.005	0.002	0.003	0.039	1.0
CZ	0.021	0.310	0.134	0.362	0.049	0.088	0.002	0.002	0.005	0.028	1.0
FR	0.020	0.170	0.080	0.391	0.024	0.127	0.007	0.003	0.016	0.161	1.0
DE	0.025	0.216	0.164	0.367	0.032	0.109	0.007	0.002	0.008	0.071	1.0
HU	0.035	0.273	0.169	0.315	0.065	0.076	0.011	0.003	0.013	0.041	1.0
NL	0.004	0.225	0.057	0.451	0.017	0.075	0.012	0.001	0.002	0.156	1.0
PL	0.019	0.304	0.092	0.312	0.104	0.089	0.014	0.003	0.010	0.052	1.0
RO	0.064	0.163	0.337	0.301	0.025	0.079	0.003	0.001	0.006	0.020	1.0
SK	0.036	0.237	0.237	0.341	0.043	0.075	0.002	0.003	0.006	0.021	1.0
SI	0.025	0.229	0.194	0.294	0.072	0.109	0.010		0.011	0.056	1.0
CH	0.005	0.123	0.048	0.464	0.011	0.136	0.005	0.000	0.007	0.199	1.0
GB	0.014	0.166	0.052	0.375	0.040	0.094	0.027	0.001	0.010	0.221	1.0

Note: The numbers of the clusters are conventional: the only important thing is whether respondents belong to the same cluster or to different clusters

Source: Own elaborations on WVS data, 1994–2009

distances), indicate that the relative within-group distances for the Baltic, the Mediterranean, and the Nordic countries are basically independent of the number of clusters (from 5 to 50), which suggests that their internal homogeneity is not a mere statistical artefact. Standardized distances are easier to compare, both when the number of clusters differ and across domains: this is why we will concentrate especially on these distances in the following, focusing in particular on whether the average distance within our predefined regions (Mediterranean, Nordic and Baltic countries), are or are not below average, meaning below zero after standardization.

The proximity of the countries within our predefined groups is also apparent with the application of multidimensional scaling (MDS), a statistical technique that permits analysts to reduce the number of dimensions of the space, from the original N to as few as one wants, but normally the two that one needs for a graphical representation in a Cartesian

Table 5 Matrix of Euclidean distances between countries (21 European countries, Domain = Happiness, Clusters = 10)

	IT	ES	FI	NO	SE	EE	LV	LT	BG	HR	CZ	FR	DE	HU	NL	PL	RO	SK	SI	CH	GB
IT																					
ES	0.044																				
FI	0.182	0.162																			
NO	0.241	0.226	0.116																		
SE	0.262	0.246	0.151	0.091																	
EE	0.314	0.281	0.288	0.375	0.372																
LV	0.332	0.303	0.311	0.405	0.399	0.092															
LT	0.303	0.269	0.290	0.362	0.369	0.102	0.188														
BG	0.289	0.255	0.231	0.281	0.298	0.195	0.272	0.137													
HR	0.264	0.229	0.193	0.272	0.288	0.126	0.185	0.118	0.118												
CZ	0.229	0.202	0.217	0.319	0.325	0.126	0.111	0.183	0.231	0.142											
FR	0.217	0.195	0.120	0.106	0.068	0.318	0.350	0.311	0.245	0.238	0.274										
DE	0.156	0.120	0.111	0.197	0.207	0.201	0.233	0.198	0.167	0.130	0.143	0.149									
HU	0.287	0.254	0.234	0.328	0.322	0.113	0.094	0.190	0.230	0.141	0.085	0.274	0.168								
NL	0.225	0.209	0.173	0.183	0.118	0.328	0.347	0.337	0.294	0.277	0.278	0.091	0.177	0.276							
PL	0.263	0.231	0.199	0.291	0.281	0.182	0.163	0.242	0.256	0.171	0.118	0.239	0.157	0.083	0.236						
RO	0.276	0.243	0.250	0.314	0.328	0.147	0.230	0.074	0.070	0.103	0.196	0.271	0.165	0.207	0.308	0.246					
SK	0.261	0.228	0.245	0.329	0.334	0.067	0.137	0.073	0.150	0.088	0.113	0.278	0.150	0.129	0.297	0.180	0.098				
SI	0.233	0.194	0.161	0.239	0.243	0.147	0.191	0.153	0.141	0.071	0.134	0.189	0.089	0.123	0.220	0.126	0.137	0.111			
CH	0.216	0.207	0.143	0.097	0.062	0.371	0.400	0.362	0.294	0.289	0.317	0.071	0.192	0.328	0.111	0.288	0.319	0.327	0.245		
GB	0.322	0.303	0.231	0.177	0.095	0.403	0.431	0.400	0.332	0.334	0.372	0.122	0.260	0.355	0.130	0.315	0.364	0.372	0.282	0.131	

Source: Own elaborations on WVS data, 1994–2009. Average distance = 0.221 (SD = 0.089)

Table 6 Average distances between 21 European countries, and within three groups: Baltic, Mediterranean, and Nordic (Domain = Happiness, Clusters = 10)

	Average	Number of distances
All	0.221	210
Mediterranean	0.044	1
Nordic	0.120	3
Baltic	0.127	3

Source: Own elaborations on WVS data, 1994–2009. The standard deviation of the 210 distances is 0.089

plan. The application of MDS to the case with $N = 15$ clusters,[6] for instance, led to Fig. 3 where our three groups of countries appear in fact to be clustered together. Note that, in this as in other cases (e.g. Fig. 7, below) our analysis indicates that other clusters of countries too would make sense—possibly even more than the three groups that we are considering here (Mediterranean, Baltic, and Nordic). Figure 3, for instance, suggests that Switzerland, France, Great Britain and the Netherlands are very close to the Nordic countries, in terms of happiness.

3.2 Other Domains

Figure 4 shows what happened with the other domains, for which, as mentioned, we decided to display only the standardized distances, which are easier to compare. The order in which the domains appear is not random: from (a) to (j) the within-group heterogeneity tends to increase. Two blocks can be identified. In the former there are the domains with marked internal homogeneity: beyond *Happiness* (same as in Fig. 2), this is also the case

[6] With a different number of clusters results were very similar, as Fig. 2 suggests (not shown here).

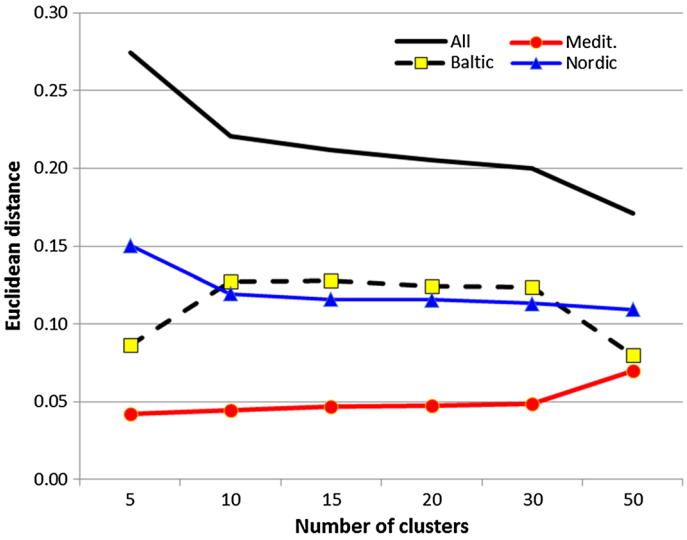

Fig. 1 Euclidean distances between 21 European countries and within three groups (Baltic, Mediterranean, and Nordic) by number of clusters (Average values. Domain = happiness). *Note* Short distances denote homogeneous groups. *Source* Own elaborations on WVS data, 1994–2009

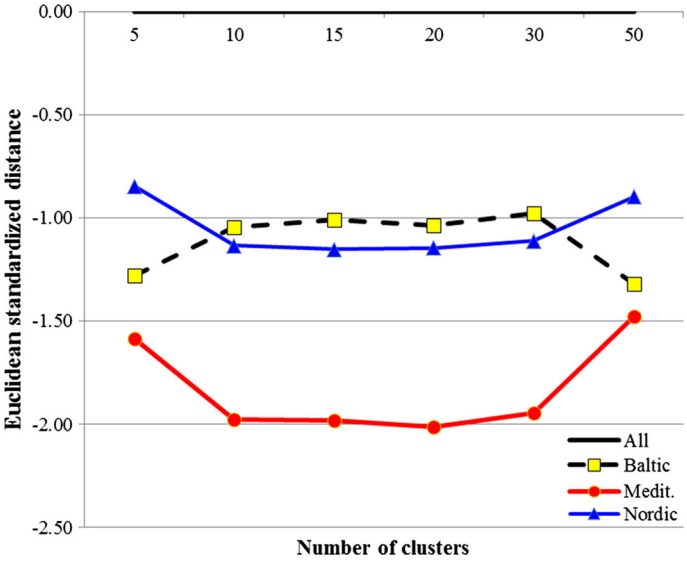

Fig. 2 Standardized distances between 21 European countries and within three groups (Baltic, Mediterranean, and Nordic) by number of clusters (Average values. Domain = happiness). *Note* Short distances denote homogeneous groups. *Source* Own elaborations on WVS data, 1994–2009

of *Friends and associations*, *Self* (care of and importance attributed to one's self) and *Work*.

Then there are several domains where internal homogeneity emerges in some groups of countries but not in others:

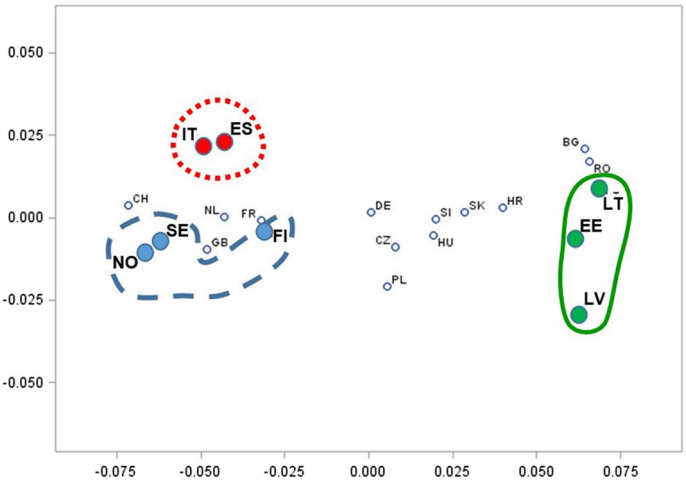

Fig. 3 Representation in two dimensions of the relative Euclidean distance between the 21 European countries of the WVS (1994–2007). Domain: happiness; No of clusters: 15. *Note* Figure obtained with MDS (SAS software, release 9.3). The figure can be rotated or flipped at wish: the only thing that matters is the distance between countries, which is not affected by these changes of perspective. The correlation between the original and the plotted distances is very high, close to 95 %, which means that the distortion introduced with MDS is minimal. The Baltic, the Mediterranean and the Nordic countries are *highlighted*. *Source* Own elaborations on WVS data, 1994–2009

- *Ethics*: the Mediterranean and the Nordic countries are internally homogeneous; not so much the Baltic countries;
- *Importance of Politics* and *Family*: vice versa;
- *Gender attitudes*: the Nordic countries are not particularly homogeneous;
- *Political orientation (left/right)*: only the Mediterranean countries appear to be homogeneous; and, finally,
- *Religion*: only the Nordic countries appear to be homogeneous (and the results depend on the number of clusters that one forms, which means that they are not particularly robust).

3.3 A Parenthesis: Scores and Relative Position of the European Countries

One of the useful aspects of our approach is that it permitted us to evaluate the relative distance between countries better and less arbitrarily than most alternatives. In order to demonstrate this more clearly, and to provide an idea of the general orientation of our set of countries in the domains that we investigated, we calculated a synthetic measure of the country's "position" in each of the domains. To do this, we assigned partly arbitrary[7] numerical scores to the various possible answers, and we computed the country average (Fig. 5). It should be noted that in no case is the absolute value of these indicators of any

[7] The scores have more or less the same range on all the manifest variables. But since we dealt with manifest variables the possible answers to which ranged between 2 -dichotomous- and 21 we adjusted the scores so as to catch their contribution to that latent domain to the best of our possibilities. With alternative scores, not qualitatively different to be sure, the results did not change appreciably (not shown here). For the details, please refer to (De Santis et al. 2014).

importance: what matters is the relative position, with respect to Europe, or to any other country or group of countries of interest.

It is interesting to note that the average country scores typically have a small range of variability, which may be taken as an indication that the European countries are generally similar to each other—or, perhaps, that stereotypes prevail. Within this limited range of variability, the countries that we deemed to be similar a priori (the Mediterranean, the Baltic and the Nordic), are rarely close to each other in Fig. 5. This happens with happiness (a) and "social life" (b- Friends and associations), but not, or at least not systematically, with other domains: the indications for forming clusters are, to say the least, unclear—and, besides, based on arbitrary scores, whereas our approach circumvents both obstacles.

Another potential problem with averages it that they may hide markedly different distributions. Consider, for instance, variable E025 (political action: signing a petition), which belongs to our domain "importance of politics." The possible answers (and the associated, arbitrary scores) are: "Have done" (= 6); "Might do" (= 3); "Would never do" (= 0). Imagine, for instance, that in country A all respondents answer "Might do;" whereas in country B 50 % of respondents answer "Have done;" while the other 50 % answer "Would never do." The average score of the two countries is identical (3, in this case), but their distributions are as far apart as they could possibly be (for the same average). Our approach, which is based on distributions and not on averages, is immune from this type of bias.

3.4 All the Domains Together: Are the Mediterranean, the Nordic and the Baltic Countries Homogeneous?

Using our method, we also wanted to arrive at a global evaluation of how homogeneous the three groups of countries are. To do this, it suffices to consider all of the elementary variables together as manifest indicators of a (very) latent domain, which may be called "country values and orientations," and to treat them in the usual way: form clusters of respondents, calculate the proportion of nationals who belong to the various clusters, and evaluate the distance between countries on the basis of these distributions.

Unfortunately, this proposition, which is true in theory, cannot be tested in practice because, as explained in Sect. 2, with the Jaccard distances the number of comparisons increases exponentially, and soon exceeds our (and probably everybody's) computing capacity. We therefore resorted to a not totally correct use of clusters, by applying Ward's minimum variance criterion directly to the (dichotomized) manifest variables, without passing through the Jaccard index.

The results are presented in Figs. 6 and 7. Both figures suggest that the Baltic and the Nordic countries do indeed form a homogenous group: the average within-group standardized distance is markedly below the average distance between European countries (Fig. 6), and the countries lie close to one another (Fig. 7). The evidence is somewhat less compelling as to the homogeneity of Italy and Spain (the only Mediterranean countries available in the WVS dataset): their average within-group distance is just slightly below the "normal" distance between any two European countries picked at random.

As for Fig. 3, one may observe that other countries lie close to those of our three groups: for instance Hungary and Slovakia appear to belong to the same group as the Baltic countries (although, as a note of caution, it is worth reminding that Fig. 7 is in just two dimensions, whereas the actual points are in a 15-dimensional space, which means that the actual distances are not exactly those that meet the eye). In some cases, these proximities can be explained with relative ease: Poland and Italy, for instance, are both Catholic countries, and they appear to be close in several respects (De Santis et al. 2014); a block of

countries from central Europe stands out (The Netherlands, Germany, France, Switzerland and Slovenia), which is more or less in line with Esping-Andersen's typology (1990, 1996); etc. Some of the proximities, instead are less obvious: this is the case of Great Britain and Spain, for instance, or of the Czech Republic and Estonia.

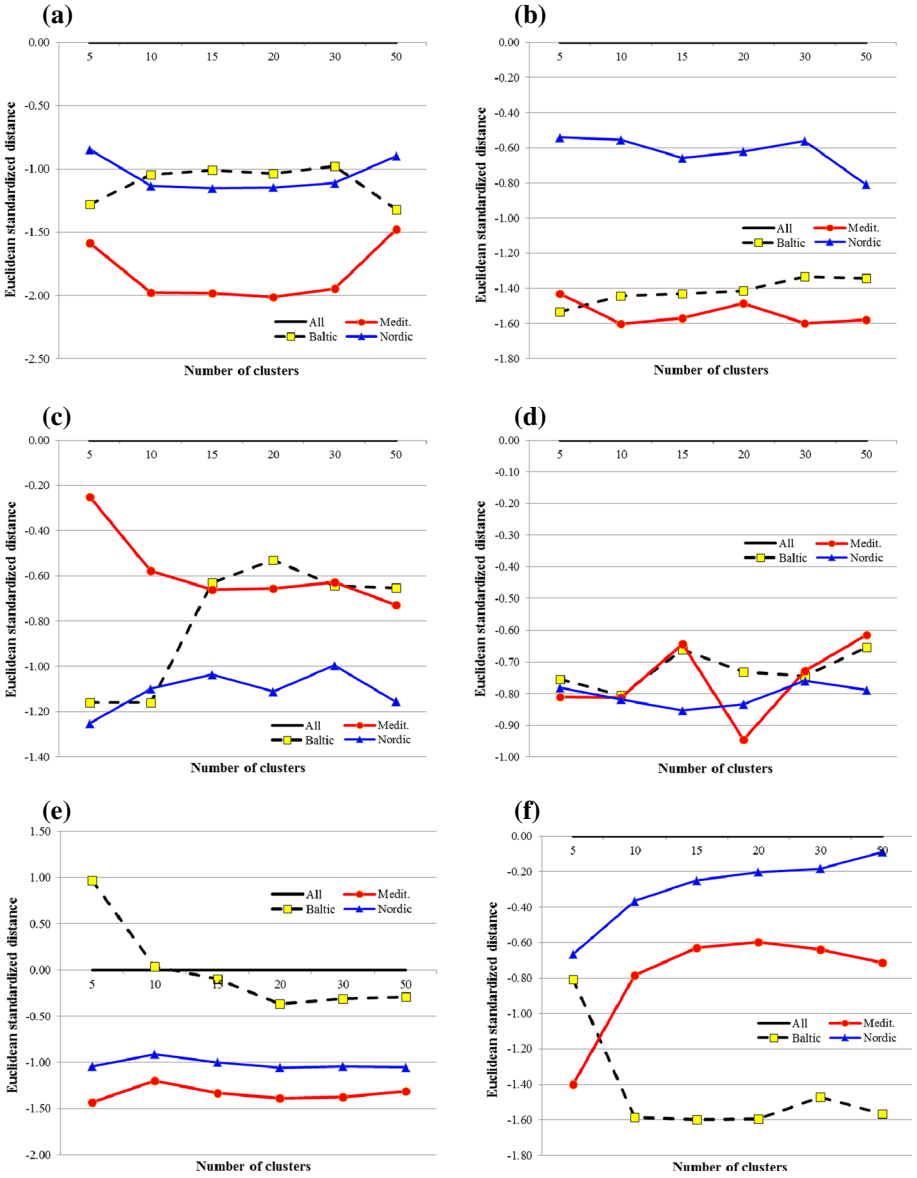

Fig. 4 Standardized distances between 21 European countries and within three groups (Baltic, Mediterranean, and Nordic) by number of clusters (Average values. Various domains), **a** Happiness **b** Friends and associations, **c** Self (care/consideration of one's self) **d** Work, e. Ethics **f** Politics (importance of), **g** Family **h** Gender attitudes, **i** Politics (*left*/*right*) **j** Religion, *Source* Own elaborations on WVS data, 1994–2009

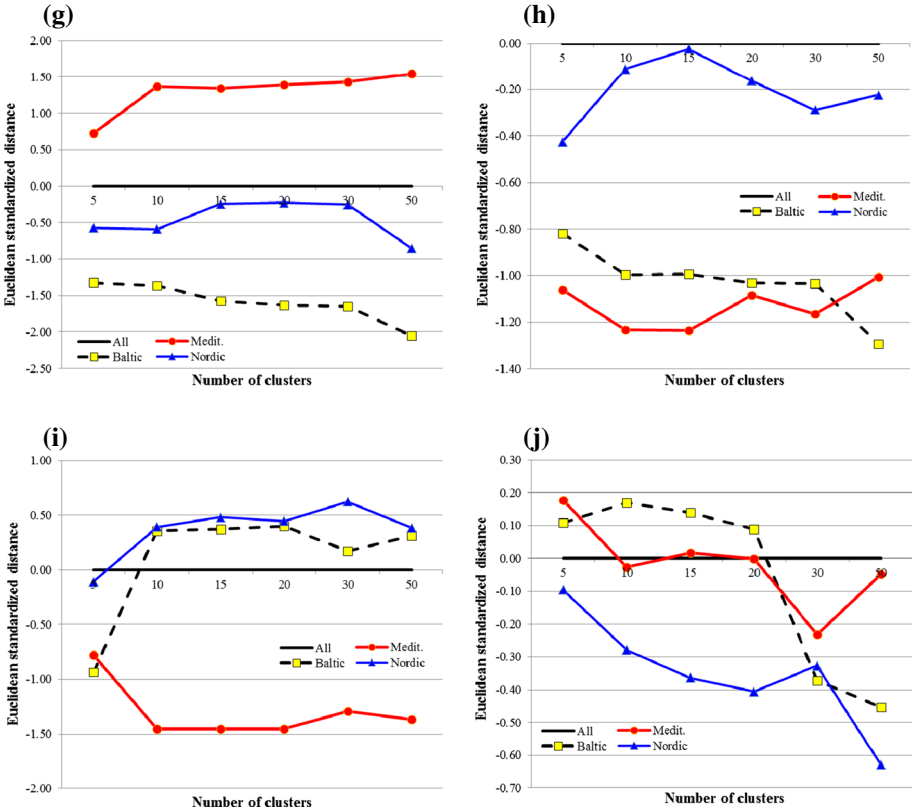

Fig. 4 continued

We remind our readers that the empirical base is rather thin here—a sample of 1000–4000 respondents, depending on the country—so that the implications of our empirical results should not be overemphasized: suffice it to say that, all in all, they do not patently rebut our claim that the method that we are proposing here may be of some help in classifying countries.

4 Discussion and Conclusions

Comparing countries with the aim of assessing how close or how far apart they are is not an easy task and in most cases such comparisons are highly subjective and arbitrary. The traditional approach is based on country averages of a set of indicators (income, unemployment, proportion of women in Parliament, etc.).

When individual data are available, which is increasingly the case as more and more internationally comparable surveys are taken, another path may be followed based on the *distribution* of national respondents among the different possible answers.

This is the (as far as we know, original) rationale of the method adopted here: based on the subjective values and opinions expressed by national respondents and collected by the World Value Surveys, the orientations of the citizens of 21 European countries are known.

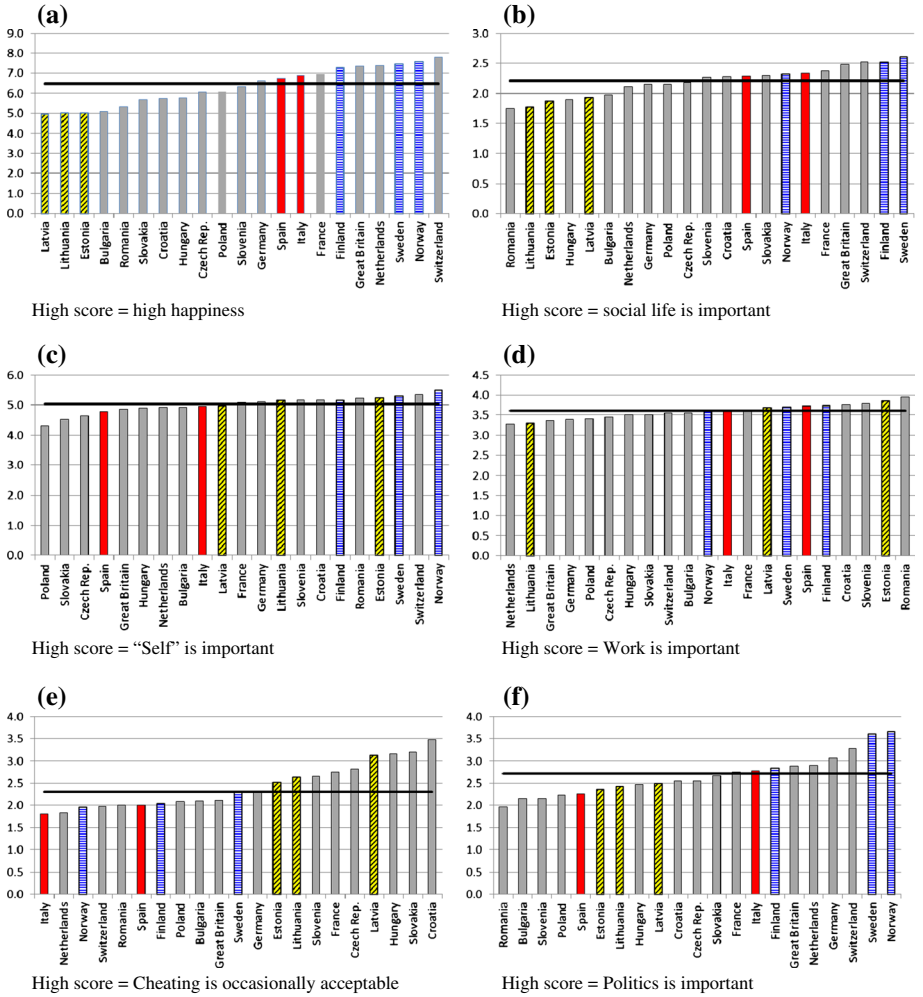

Fig. 5 Rough measures of the relative standing of each country in each domain. **a** Happiness, **b** Friends and associations, **c** Self (care/consideration of one's self) **d** Work, **e** Ethics, **f** Politics (importance of), **g** Family, **h** Gender attitudes, **i** Politics (*left*/*right*), **j** Religion, *Note* Questions and (arbitrary) scores detailed in De Santis et al. (2014). *Horizontal line* European average. The Baltic, the Nordic and the Mediterranean countries are *highlighted*. *Source* Own elaborations on WVS data, 1994–2009

On the basis of this dataset, and of a few very obvious assumptions (that the clustering method actually does what it is supposed to do: it groups together respondents with similar attitudes and beliefs), clusters of homogeneous individuals can be formed. The distribution of national respondents among the various clusters gives us an idea of how similar two given countries are: the closer their distributions, the closer the countries.

While the guiding principle of the proposed method is very simple, its practical application may be tricky: the choice of the variables is not always easy (what manifest variables can be considered valid indicators of what latent domains?), the clustering method matters (results with Ward's minimum variance method are better than those obtained with other criteria) and so does the criterion of distance that one uses (here:

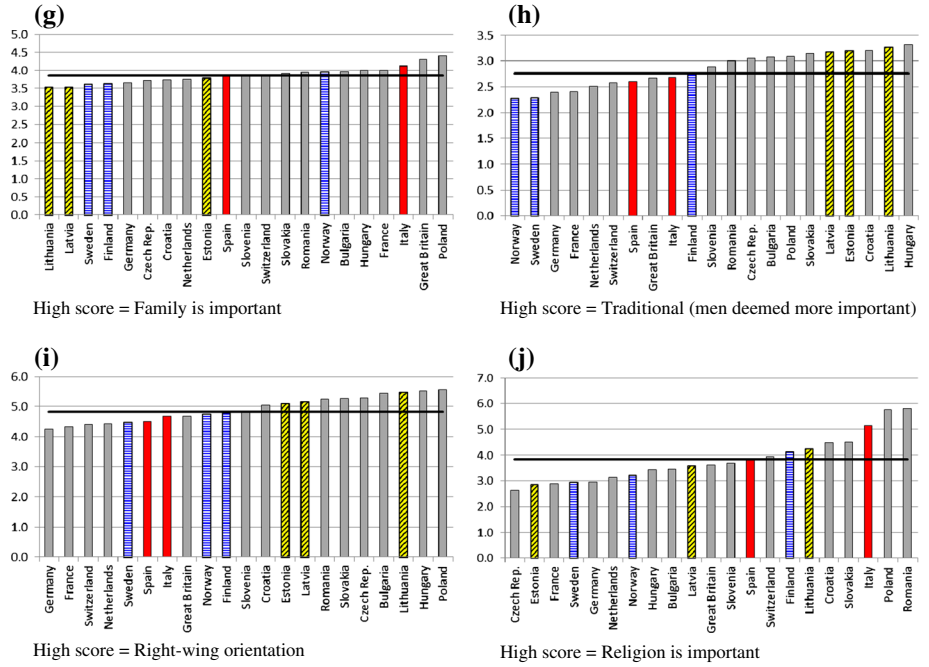

Fig. 5 continued

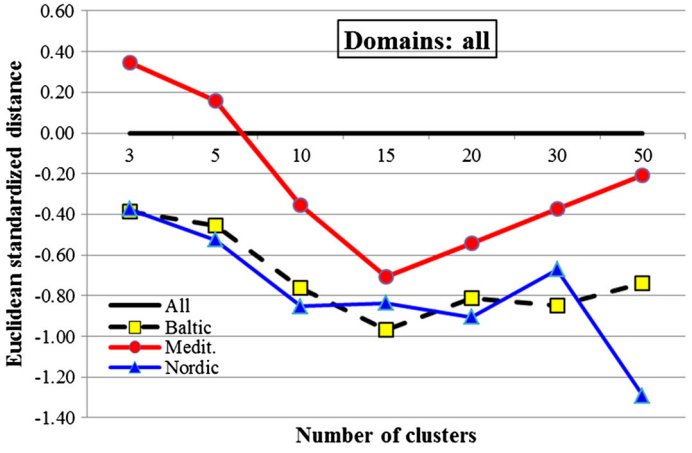

Fig. 6 Standardized distances between 21 European countries and within three groups (Baltic, Mediterranean, and Nordic) by number of clusters (Average values. Domain = All). *Source* Own elaborations on WVS data, 1994–2009

Jaccard, but it is cumbersome and severely limits the number of variables that one can keep under control), the number of clusters is not always neutral, etc.

In short, the proposed method does not seem to be fit for all possible situations, and generally does not provide analysts with a clear-cut answer: it must be tested repeatedly, as

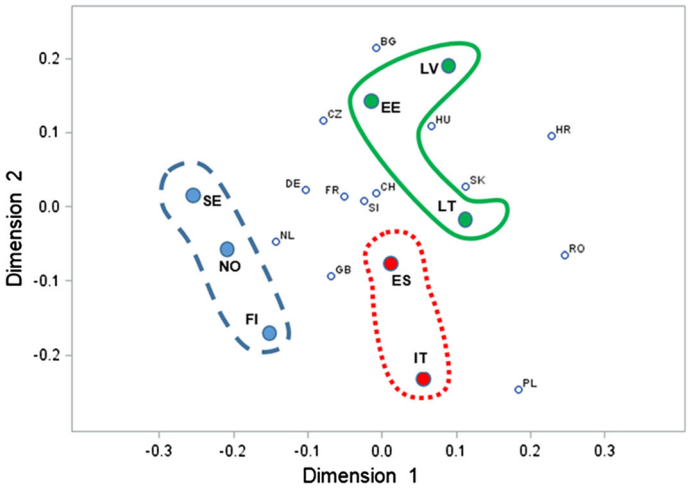

Fig. 7 Representation in two dimensions of the relative Euclidean distance between the 21 European countries of the WVS (1994–2007). Domain: all; No of clusters: 15. *Note* Figure obtained with MDS (SAS software, release 9.3). The figure can be rotated or flipped at wish: the only thing that matters (indifferent to these changes of perspective) is the distance between countries. The correlation between the original and the plotted distances is 95.9 %. The Baltic, the Mediterranean and the Nordic countries are *highlighted*. *Source* Own elaborations on WVS data, 1994–2009

we did in this case, and tried with variants and alternative specifications (of the variables, the criteria, the number of clusters, etc.). Its results can be considered (relatively) trustworthy only when they appear to be reasonably consistent, and robust to changes.

Even with these limitations in mind, we think that the proposed method fills a gap in the toolkit of social scientists. As more and more individual data become available, forming clusters of countries (or regions) can be done on more than just ecological averages, and this is precisely the direction in which the proposed method moves.

Acknowledgments We thank two anonymous referees and our colleague, Leonardo Grilli, for their useful suggestions.

References

Aassve, A., Betti, G., Mazzuco, S., & Mencarini, L. (2007). Marital disruption and economic well-being: A comparative analysis. *Journal of the Royal Statistical Society Series A, 170*(3), 781–799.

Cattell, R. B. (1950). The principle culture patterns discoverable in the syntal dimensions of existing nations. *Journal of Social Psychology, 32*(2), 215–253.

Christiansen, S., & Keilman, N. (2013). Probabilistic household forecasts based on register data- the case of Denmark and Finland. *Demographic Research, 29*(art. 43), 1263–1302.

De Santis, G., Maltagliati, M., & Salvini, S. (2014). How close? An attempt at measuring the cultural distance between countries. *Working Paper of the Institute of Statistics and Demography*, *Warsaw School of Economics no*. 38/2014, http://kolegia.sgh.waw.pl/pl/KAE/struktura/ISiD/publikacje/Documents/Working_Paper/ISID_WP_38_2014.pdf

Delhey, J. (2009). From materialist to postmaterialist happiness? *World Values Research*, 2(2): 31–55. http://www.worldvaluessurvey.org

Dickes, P., & Valentova, M. (2013). Construction, validation and application of the measurement of social cohesion in 47 European countries and regions. *Social Indicators Research, 113*(1), 827–846.

Eichhorn, J. (2012). Context matters: The effect of national-level factors on the relationship between socio-demographic characteristics of individuals on their life-satisfaction. *World Values Research, 5*(2), 26–45.

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Princeton, NJ: Princeton University Press.

Esping-Andersen, G. (1996). *Welfare states in transition: National adaptations in global economies*. Beverley Hills, CA: SAGE.

Guptaa, V., Hangesb, P. J., & Dorfmanc, P. (2002) Cultural clusters: Methodology and findings. *Journal of World Business,* 37(1), 11–15. http://www.sciencedirect.com/science/article/pii/S1090951601000700

Hoem, J. M., Gabrielli, G., Jasilioniene, A., Kostova, A., & Matysiak, A. (2010). Levels of recent union formation: Six European countries compared. *Demographic Research, 22*(9), 199–210.

Inglehart, R. (1971). The silent revolution in Europe: Intergenerational change in post-industrial societies. *The American political science review*, 65(4), 991–1017. http://costa.wustl.edu/teaching/IntroComp/Reading/inglehart1971.pdf

Inglehart, R. (2008). Changing values among western publics from 1970 to 2006. *West European Politics*, 31(1–2), 130–146. http://www.worldvaluessurvey.org/wvs/articles/folder_published/publication_559

Kapitány, B., & Spéder, Z. (2012). Réalisation et évolution des intentions de fécondité en trois ans dans quatre pays européens. *Population-F, 67*(4), 711–744.

Lesthaeghe, R. (2011). The "second demographic transition": A conceptual map for the understanding of late modern demographic developments in fertility and family formation. *Historical Social Research*, 36(2), 179–218. http://www.ssoar.info/ssoar/handle/document/34225

Leti, G. (1983). *Statistica descrittiva, Il*. Bologna: Mulino.

Li, L. M. W., & Bond M. H. (2010). Analyzing national change in citizen secularism across four time periods in the World Values Survey. *World Values Research*, 3(2):19–32. http://www.worldvaluessurvey.org

Reher, D. S. (1998). Family ties in Western Europe: Persistent contrasts. *Population and Development Review, 24*(2), 203–234.

Reigner-Loilier, A., & Vignoli, D. (2011). Intentions de fécondité et obstacles à leur réalisation en France et en Italie. *Population-F, 66*(2), 401–432.

SAS. (2011). *Online Documentation.* http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_distance_sect021.htm

Sobotka, T. (2008). The diverse faces of the second demographic transition in Europe. *Demographic Research*, 19(art. 8), 171–224. http://www.demographic-research.org/volumes/vol19/8/

United Nations. (2013). *World Population Prospects. The 2012 revision*. http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2012_Volume-I_Comprehensive-Tables.pdf

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

World Economic Forum. (2005). Women's empowerment: Measuring the global gender gap. http://www.weforum.org/pdf/Global_Competitiveness_Reports/Reports/gender_gap.pdf