As already noted, in the complex table total inertia is the average value of the 35 subtables. Table 1.5 shows the inertia of all 35 tables as well as the average inertias of the sociodemographic characteristics and of the seven national identity indicators. It can be seen that the highest inertias belong to the cross-tabulations with country, i.e., the most variation in the data is caused by country differences. Further, there are almost no sex differences for the seven items on national identity, although there are some findings that might be worthwhile to report. For example, the association between "sex" and "international sport" is much smaller than the association between "sex" and "feel ashamed of country."

1.6 Multiple correspondence analysis

In the previous examples we analyzed the relation between two variables or between two different sets of variables. In this section, we are interested in the relationships within a set of variables, for example, the interrelationships between the statements on national identity. Thus, for example, we could find out if there is an association between a "strong agreement toward international sports" and a "strong agreement toward people should support their country." In the previous analysis of stacked tables, we could only see whether these categories had the same association with sociodemographic variables.

This new case, which is reminiscent of principal component analysis, involves all the cross-tables of a set of variables, such as the national identity indicators, with themselves. Assembling all these cross-tables into a square supermatrix of cross-tables, we obtain what is known in CA literature as the *Burt matrix*, which we denote by $\mathbf{C}$. Alternatively, a data structure known as the *indicator matrix* can be constructed based on the original data. The indicator matrix, denoted by $\mathbf{Z}$, is a respondents-by-categories table with as many rows as respondents (6066 in our example) and as many columns as response categories (35 for the seven national identity indicators). The elements of $\mathbf{Z}$ are zeros apart from ones in the positions to indicate the categories of response of each respondent ($\mathbf{Z}$ is often called a matrix of dummy variables). The Burt matrix is related quite simply to the indicator matrix as follows: $\mathbf{C} = \mathbf{Z}^\mathsf{T}\mathbf{Z}$. If the usual CA algorithm is applied to an indicator matrix or to a Burt matrix, the method is called *multiple correspondence analysis* (MCA). In MCA there is no distinction between describing variables and variables to be described, as is the case in simple CA of single or stacked tables. In MCA all variables have the same status. The relationship between the analyses of **C** and **Z** in MCA is discussed in depth in Chapter 2. In the following, we illustrate the method by analyzing the 6066×35 indicator matrix **Z** that codes the responses to the seven national identity questions. The graphical solution is given in Figure 1.6.

Inspecting Figure 1.6, we see that the first dimension contrasts the strong agreements and the strong disagreements (positive part) from the middle categories (negative part). With two exceptions (statements b and g), the second dimension contrasts the positive statements from the negative ones. Therefore, it can be seen as an overall dimension toward national identity, with a relatively high national identity in the positive part and a relatively low national identity in the negative part. The variables a, c, d, e, and f form a horseshoe, a typical structure we usually find in ordered categorical data (for more details, see Chapters 2 and 4). However, there are two points to be mentioned. First, neither item b, "there are some things …," nor item g, "I am often less proud…," fulfill this structure, and maybe even worse, the most-opposite categories "b1" and "b5" as well as "g1" and "g5" are close to each other. One reason for this finding might be that a significant number of respondents misunderstood the direction of the question, which would result in such a structure (Blasius and Thiessen 2001b). Another reason is that two dimensions are not sufficient to mirror the structure of these two variables adequately. In a higher-dimensional solution "b1" and "b5" as well as g1 and g5 might be far away from each other. Second, the horseshoe belongs to the first dimension, i.e., the second dimension is more important for substantive interpretation. This might be caused by the joint analysis of the data from five countries, where respondents in different countries may understand the single questions (slightly) differently (see Chapter 20).

Notice that we have not indicated the percentages of inertia explained in Figure 1.6. There are several methods of scaling the solution in MCA that change the fit (see Chapters 2 and 3), but the overall structure of the variable categories in the space remains the same and, hence, the substantive interpretation too. Using the indicator matrix as input, the measure of fit in terms of explained inertia is heavily underestimated. Various solutions to this problem are proposed in Chapter 2 (see Section 2.3.4).

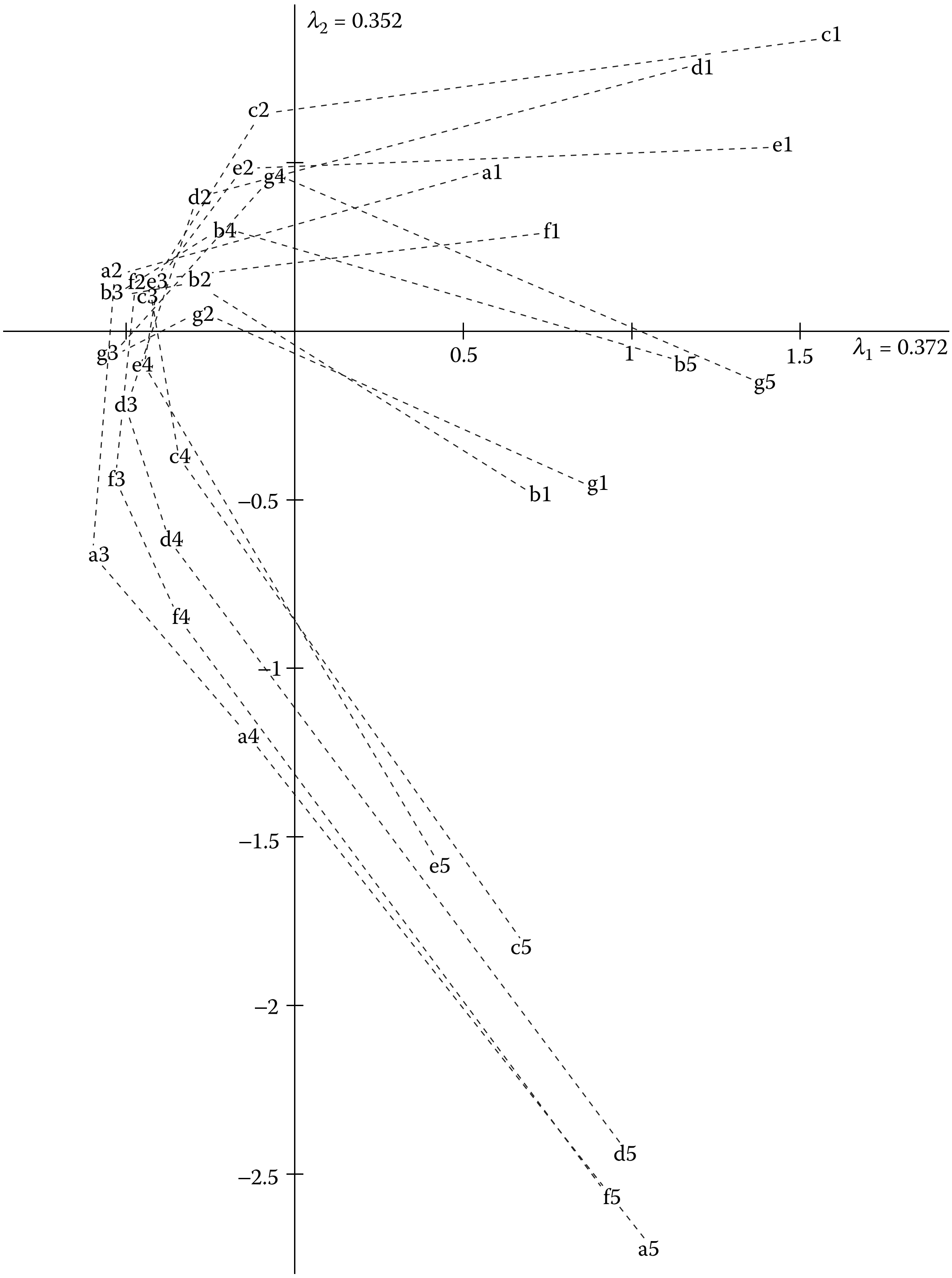

Figure 1.6 MCA map of indicators on national identity.