Methods

ESS data

The data were acquired from the ESS. Then, the relevant variables were selected and cleaned, including removal of missing values and transformation of round to corresponding year. After the data cleaning, the opinion metrics were calculated, and a separate PCA was performed for each round and country. The loadings and explained variance metrics derived from the first principal component were then considered for further interpretation.

The data were obtained from the website of the European Social Survey Data Portal (ESS Data Portal, 2024) using the inbuilt datafile builder wizard tool which allows to specific selection of the variables, rounds, and countries of interest and extracts the resulting data set as a CSV file. The data for our analysis comprised the following 39 European countries Albania, Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Kosovo, Latvia, Lithuania, Luxembourg, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine and The United Kingdom. The data surveyed by the ESS are on a typical Likert scale. This type of scale measures attitudes, opinions, or perceptions and provides a range of options for respondents to choose from. It often ranges from strongly disagree (0) to neutral (5) to strongly agree (10), with shades in between (e.g., Jamieson, 2004). The Likert scale is often used in social science in order to quantify subjective data such as attitudes and satisfaction levels (Journal of Social Sciences, 2022). The focus on was on three variables related to opinions about migration, which ranged from 0 to 10. Thus, 0 would indicate strong resentment, 5 a neutral sentiment, and 10 a strong approval of immigration. Other values encoding questionnaire answers such as 77 = ‘Refusal’, 88 = ‘Don’t know’ and 99 = ‘No answer’ were present in the data as well. Those values were re-labeled as missing data points and subsequently removed them from the data set. The three relevant variables were imwbcnt(‘Immigrants make country worse (0) or better (10) place to live’), imueclt(‘Country's cultural life is undermined (0) or enriched (10) by immigrants’), and imgbeco(‘Immigration is bad (0) or good (10) for country's economy’).

Weights are crucial in survey data as they help to ensure that the survey results accurately reflect the population being studied. Surveys aim to gather information from a sample of the population, but it's often difficult to get a perfectly representative sample. Some groups within the population may be over-represented or under-represented in the sample due to factors like sampling design and responsiveness (some people chosen for the survey don't participate). Weights adjust the data to compensate for these imbalances, giving more influence to the responses of under-represented groups and less influence to the responses of over-represented groups (Pfeffermann, 1996; Ciol et al., 2006). The ESS provides several survey weight variables. The analysis weight (variable name anweight)corrects for differential selection probabilities within each country as specified by sample design, for nonresponse, for noncoverage, and for sampling error related to the four post-stratification variables, and takes into account differences in population size across countries. It is constructed by first deriving the design weight, then applying a post-stratification adjustment, and then a population size adjustment. Starting from Round 9, anweight is provided in the integrated data file (ESS weighting variables, 2024). For data from earlier ESS rounds, anweight was derived by multiplying *pspwght* with pweight in accordance with instructions provided by the ESS (Guide to Using Weights and Sample Design Indicators with ESS Data, 2024).

The data were analysed using the R programming language in version 4.3.2 (R Core Team, 2023). PCA was done using the inbuilt stats (R Core Team, 2023) package. Data manipulation was done using the dplyr (Wickham et al., 2023) and the glue (Hester & Bryan, 2024) packages. The 2-letter country codes were transformed to the full country names and vice versa using the countrycode (Arel-Bundock, Enevoldsen & Yetman, 2018) package. Missing data were handled using the naniar package (Tierney & Cook, 2023). Visualizations were created using the ggplot2 (Wickham, 2016) and the ggrepel (Slowikowsi, 2024) packages and composed using the Patchwork (Pedersen, 2024) package. A complementary web application was created in order to create a visual interface for data exploration using the shiny (Chang et al., 2023), shinyWidgets (Perrier et al., 2025) and shinydashboard (Chang & Borges Ribeiro, 2021) packages.

**Polarization metrics**

*imbgeco*, *imueclt* and *imwbcnt* all range from 0 (extremely negative) to 10 (extremely positive), with 5 being a neutral stance. However, for many of the following metrics, the fractions denoted as had to be calculated first as the relative proportions of (valid) answers for the options zero to ten regarding the survey variable in question. A proportion compares a part to the whole. It indicates what fraction of the total a particular part represents and thus ranges from 0 to 1. The weighted proportions of each level of *imbgeco*, *imueclt* and *imwbcnt* were calculated, taking into account the aforementioned weighting variable anweight.

Non-neutrality

The fraction of individuals who do not use the central (neutral) answer option (five on the eleven-point scale from zero to ten). This assesses polarization in a very basic sense by counting the fraction of people who do not take a neutral stance. For public opinion on immigration increased non-neutrality means that more individuals take a decision to either lean towards acceptance or objection instead of being impartial. Most polarized in that sense would be a society without any impartial opinions. It captures loss of central mass.

Average deviation from neutrality

where

Shows resemblance to the concept of group polarization in psychology: How far is the average attitude away from the scale’s midpoint? This is operationalized by capturing the absolute value of the distance of the mean of the opinion distribution. For public opinion on immigration an increasing average deviation from neutrality means that individuals become either increasingly accepting or increasingly rejecting in their views. Most polarized in that sense would be a society where everyone has an extreme stance, be it "0" or "10". It also relates to the concept of group polarization from social psychology as it captures a shift of opinion positions “toward a more extreme point” (cf. Sunstein, 2003, p. 81). However, the most polarized society in this notion is also consensual in its extremity, and thus least polarized in the following notion.

Dispersion

Is measured by its mean absolute deviation. This is a good basic measure of polarization for typical survey response distributions on bounded scales (0-10 in the ESS). The maximum dispersion is when half of the individuals are on both extremes while the minimal dispersion appears for any consensus with all individuals having the same response. This measure is exactly as proposed by Bramson et al. (2016).

Moderate divergence

with

and

and

and

Assessed by the absolute difference of group means of the moderate accepting group and the moderate opposing group, as described in Bramson et al. (2016). Lorenz (2017) analysed the typical characteristics of ESS opinion distributions and pointed to the existence of five endogenous groups in ESS opinion distributions: The extreme left, the moderate left, the neutrals, the moderate right & the extreme right. Per item, we operationalize similar groups, consisting of: The "full-on acceptors" (individuals with opinion 0), the "moderate accepting group" (individuals with opinion 1 - 4), the "neutrals" (individuals with opinion 5), the "moderate opposing group (individuals with opinion 6 - 9) & the "full-on opponents" (individuals with opinion 10).

The factor normalizes the measure to range from 0 to 1 (0 = maximally similar, 1 = maximally divergent). Starting from here, p1 to p4 and p6 to p9 MUST be present, because those are the moderate groups

Moderate group consensus

where

and

Based on the mean absolute deviation (MAD) of the two moderate groups. In contrast to dispersion, which we assess as MAD of the entire opinion distribution, the measurement of group consensus increases with decreasing dispersion in the two groups.

Moderate size parity

}

The relative size of the smaller group of moderates compared to the larger group. Hereby the mass of the smaller group is divided by the mass of the larger group. A size parity of 1 indicates equal size of moderate groups and thereby the maximum possible polarization in the sense of parity. This is a simplified version of the size parity measure proposed by Bramson et al. (2016). For public opinion on immigration, increasing moderate size parity means that moderately accepting views and moderately opposing views become more similar in numbers. Most polarized in that sense would be a society where both groups are of equal size.

Multi-dimensional approach from Stefano's capstone project

Principal Component Analysis (PCA) transforms a dataset with potentially correlated variables into a new set of uncorrelated variables – the so-called principal components. These principal components are ordered so that the first few retain most of the variation present in all of the original variables, allowing to simplify complex datasets while preserving crucial information.

As the first step of performing PCA, the covariance matrix is calculated to understand the relationships between the considered variables. Which were *imbgeco*, *imueclt* and *imwbcnt* in the case of this analysis. The weights were incorporated by creating the correlation matrix using the analysis weight variable (anweight), resulting in the weighted correlation matrix based on which the following steps of the PCA were done. Next, the eigenvectors and eigenvalues of the covariance matrix are computed using methods from linear algebra. The eigenvectors become the directions of the principal components, and the eigenvalues indicate the amount of variance explained by each component. By selecting only the top few principal components - those with the highest explained variance -, it is possible to reduce the dimensionality of the data (e.g., Abdi & Williams, 2010; Gewers et al., 2018). In essence, PCA is a way to find the most important patterns in a complex dataset by finding the directions of greatest variance, and then using those directions to represent the data in a simpler way. To use an analogy, one could think of a shadow: Imagine shining a light on a cloud of dots representing data points. The shadow on the wall is a simplified version of the cloud. PCA is like finding the best angle to shine the light, so the shadow captures as much of the original shape as possible. Now, imagine stretching the cloud along its longest stretch and squeezing it along its shortest stretch. PCA is like finding the right stretches and squeezes to simplify the cloud.

As mentioned before, each principal component (PC) has a certain amount of explained variance. Assuming the that first PC would reflect a general migration attitude, we were now able to identify trends in the form of a slope coefficient by assessing the explained variance of PC1 over time using linear regression.

Linear regression attempts to find a linear relationship between a dependent variable (the variable one wants to predict or model) and one or more independent variables (the variables use to make the prediction). In simple terms, it tries to fit a straight line to a set of data points. The "best fit" line is determined by minimizing the differences between the predicted values (from the line) and the actual values of the dependent variable. A common method for doing this is the "least squares" method (e.g., Su, Yan & Tsai, 2012; Kumari & Yadev, 2018). For only one independent variable, the line follows the form where m is the slope coefficient describing the average change in y for a 1-unit change in x.

For the research question at hand, linear regression was used to model the explained variance of the first principal component over time. This, in turn, made it possible to use the corresponding slope coefficient as an indicator for an ongoing trend (or the lack thereof) of issue alignment or heterogenization regarding the three migration-opinion variables.