Latent Polarization*

Klaus Desmet
SMU, NBER and *CEPR*

Ignacio Ortuño-Ortín
UC3M

Romain Wacziarg
UCLA and *NBER*

May 2024

Abstract

We develop a new method to endogenously partition society into groups based on homophily in values. The between-group differentiation that results from this partition provides a novel measure of latent polarization in society. For the last forty years, the degree of latent polarization of the U.S. public has been high and relatively stable. In contrast, the degree of partisan polarization between voters of the two main political parties steadily increased since the 1990s, and is now converging toward that of underlying values-based clusters. Growing partisan polarization in the U.S. is a reflection of partisan views becoming increasingly aligned with the main values-based clusters in society.

*Desmet: Department of Economics and Cox School of Business, Southern Methodist University, 3300 Dyer, Dallas, TX 75205, kdesmet@smu.edu; Ortuño-Ortín: Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain, iortuno@eco.uc3m.es; Wacziarg: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, wacziarg@ucla.edu. This paper was previously circulated under the title "Endogenous Partitions". We thank seminar and conference participants at Bocconi University, Fundação Getulio Vargas (São Paulo), UC Berkeley, Central European University (Vienna), CERGE-EI (Prague), Stanford University and the 2023 ERINN Conference (Paris), as well as David Baron, Matthew Gentzkow, Keith Krehbiel, Enrico Spolaore, Francesco Trebbi, Clémence Tricaud and Luigi Zingales, for useful comments.

1 Introduction

There is growing concern over rising social divisions in the the United States. Many have observed the salience of the culture wars in the public arena, arguing that there is growing social animosity due to disagreements over issues like abortion, the rights of minorities, the role of religion, the extent of redistribution, and a range of other moral, cultural and policy issues. In this paper, we develop a new measurement framework to bring structure and rigor to characterizing the extent of social disagreements. We use this new framework, along with survey data on values, attitudes and beliefs, to assess whether underlying social divisions have increased in the United States over the last forty years, finding that they have not.

How then can we make sense of the perception of growing animosity? We argue that political divisions increasingly reflect the social polarization that has been latent for decades. While these latent divisions have not increased, their expression in socially visible groups has become more pronounced over the last twenty years. These groups are increasingly defined by partisan affiliations: voters of the two main parties align more and more with latent values-based clusters that we identify and characterize.

To document the evolution of latent divisions, we propose a methodology to endogenously identify the relevant values-based groups that serve as a basis to measure these divisions. We start from the premise that people prefer to interact with others who share similar values (homophily). The greater the difference between the vectors of values of two individuals, the more antagonism they experience when interacting. Individuals then choose the group that minimizes their expected antagonism. An equilibrium is a partition such that no agent prefers to join another group. It is important to emphasize that such an equilibrium does not necessarily and always manifest itself in visible social divisions such as political parties, ethnic and racial groups, regional clusters, etc. Rather, it is a latent division.

Our methodology allows us to define a novel metric of the degree of social polarization. The partition that minimizes within-group antagonism is also the one that maximizes between-group antagonism: by making groups as homogeneous as possible in terms of values, we are also making groups as different as possible from each other. As such, the between-group differentiation that results from this endogenous partition provides a measure of maximum attainable polarization in a society when people's values and norms are the primitives to form groups. This polarization is *latent*, setting it apart from polarization between actually observed groups. Only if social groups (such as political parties) differentiate according to people's values, would this latent polarization express itself as actual polarization.

We implement our methodology using data from the seven waves of the integrated World Values Survey (WVS) for the U.S., spanning the last four decades. In every wave, we consider a set of about 200 questions reflecting respondents' values on a wide range of issues, from economic redistribution and politics, to religious beliefs and moral views. To avoid the duplication of similar questions, our analysis focuses on the main principal components of these questions. Using a clustering algorithm, we find the endogenous values-based partitions corresponding to the theoretical equilibria, and compute a measure of latent polarization from these groupings.

A first finding is that latent polarization is much larger than partisan polarization. For instance, in the most recent wave of the WVS, the average reduction in antagonism is 41.2% when partitioning the U.S. into two values-based clusters. That is, 41.2% of the overall heterogeneity in values is between groups. This is our measure of the degree of latent polarization, in the case of two groups. In contrast, the corresponding share of heterogeneity in values that is between voters of the two main political parties is 14.7%. Hence, partisan polarization is almost three times smaller than latent polarization. Polarization between other types of groups, such as those based on race or gender, is even lower. Heterogeneity in values between racial groups makes up only 1.1% of overall heterogeneity, the same percentage as for gender groups. From this, we conclude that if individuals prefer to associate with other like-minded individuals, then forming groups based on race or gender is not very efficient. Clustering on political party affiliation is clearly more efficient, though still substantially less efficient than forming groups based directly on people's values.

A second finding is that latent social polarization in the U.S. has been relatively stable over time, and the two main dimensions that partition society have remained largely unchanged. Over the last four decades, the reduction in antagonism that results from forming two values-based clusters has remained stable around 40%. There is no evidence of an upward trend in the degree of latent polarization between values-based clusters. In addition, the main cleavages that drive the division between clusters have not changed either. In 1981, the first two principal components of the WVS largely captured religious/moral values and social capital; for the most recent wave of the WVS, almost four decades later, this was still the case.

A third finding is that although the most efficient partition has been mostly along the religious/moral dimension for at least four decades, in the 1980s and early 1990s there was another (less efficient) equilibrium. In this alternative equilibrium, income differences between the two clusters are substantial, whereas religious differences are absent. This suggests that in the 1980s there were two ways to partition society, one more along religious lines and another more along income lines. By the mid-1990s, the only stable way to partition the U.S. had become along religious lines.

A fourth finding is that polarization between the voters of the two main political parties in the U.S. has increased, making them more aligned with the underlying values-based clusters. Before 2000, partisan polarization, measured as the share of overall antagonism that is due to between-party differences, was consistently below 5%. Around 2000, it started to increase, reaching 10% by the mid-2000s, and further increasing to around 15% in the last decade. This indicates that partisan polarization is rising against the backdrop of a still substantially larger latent polarization of society. An alternative way of measuring the partisan divide is to compute the difference in the mean positions of Democrats and Republicans along the main principal components. When doing so, we observe the same partisan divergence. In the 1980s there was almost no difference in the average positions of both groups. Since then, the average positions of the two political clusters have gradually diverged from each other. During this process, the mean partisan positions have become much more aligned with the mean positions of the values-based clusters. In fact, by 2017, the mean position of Republican voters coincided almost exactly with the mean position of one of the two values-based clusters identified by our algorithm (the socially conservative one).

Together, these findings suggest that rising partisan polarization in the U.S. cannot simply be ascribed to people's preferences changing or to society becoming inherently more divided. The latent clusters have long been highly polarized. Instead, political parties have become more representative of these latent social clusters. In other words, politics is increasingly giving a voice to underlying values-based clusters, making them less latent and more visible.

Our paper is related most closely with the literature documenting the rise in partisan polarization and its relation to divisions within the American public. There is a broad consensus that partisan polarization has increased in recent decades. One expression of the widening partisan divide is the increased correlation between an individual's policy views and her voting behavior (Fiorina and Abrams, 2008): knowing someone's views on issues has become more predictive of political party affiliation. Another expression is the rise in affective polarization, with an increasing number of Americans expressing a dislike of voters of the other party (Iyengar et al., 2019; Boxell et al., 2022). There is less consensus about whether growing partisan polarization reflects actual deep divisions between Americans. When analyzing disagreement on different issues among the overall population, Ansolabehere et al. (2006) talk about a 'Purple America' and Fiorina et al. (2005) about the 'Myth of a Polarized America'. In contrast, Abramowitz and Saunders (2008) claim that ideological polarization has increased substantially among the mass public since the 1970s. Instead, we find that social polarization has been high and relatively stable for decades.

When analyzing the degree of polarization of the American public, the literature has taken different approaches. A first approach is to analyze either the distribution of self-reported ideology, or the distribution of opinions on a few key issues. If the distribution becomes more bimodal or more dispersed, one concludes that societal polarization has increased. As an example, Abramowitz and Saunders (2008) consider polarization over seven issues, such as aid to Blacks, abortion, and defense spending. A second approach is to analyze issue alignment: the more correlated people's opinions are on different issues, the more polarized society is. For example, Baldassarri and Gelman (2008) analyze pairwise correlations between 47 issues from 1972 to 2004, finding no evidence of increasing issue alignment.

These approaches have two drawbacks. One drawback is that it has to select what the relevant issues are for measuring polarization. An exception is DellaPosta (2020) who takes all opinion questions from the General Social Survey. While he does not find increased alignment across politicized issues between 1972 and 2016, he uncovers increasing alignment on a broader set of opinion dimensions. A second drawback is that it does not measure polarization between groups, because it does not propose a method for defining groups.[1] Given that polarization is often understood in

[1] A notable exception is Draca and Schwarz (2023) who use survey data to identify different ideologies, and to

terms of "us" vs "them", conceptualizing the notion requires group structure, as also emphasized in the seminal contribution of Esteban and Ray (1994).

Our paper addresses both concerns. On the one hand, we look at a comprehensive list of values, and thus avoid making an *ad hoc* selection. On the other hand, we identify values-based clusters, to measure social polarization between groups. Our approach has two additional advantages. First, it provides a common metric to compare social and partisan polarization. If, as in much of the literature, social polarization is measured without defining groups, then it is difficult to compare it to partisan polarization. Since we measure social polarization between endogenously identified clusters, we can use the same metric as for partisan polarization between Democrats and Republicans. Second, our methodology allows us to identify multiple equilibria, which can shed light on the changing nature of cultural cleavages in the United States.

The endogenous clustering of individuals into groups based on their affinity in values connects with the literature on homophily. The idea is that "birds of a feather flock together": individuals tend to associate disproportionately with others who are similar to them. Most papers consider homophily in demographic or identity traits, such as age, gender, ethnicity, education, or occupation (McPherson et al., 2001; Jackson, 2010). Some also include homophily in political preferences (Verbrugge, 1977). We, instead, consider homophily in values as the basis for group formation.

This paper is also part of an extensive literature conceptualizing and measuring social heterogeneity. One strand of this literature focuses on measuring heterogeneity based on specific identity markers, such as ethnicity (Alesina et al., 2003). A limitation of this approach is that there is considerable heterogeneity in cultural values within ethnic groups, so partitions based on predefined identity traits may not be the most relevant dimensions of heterogeneity (Desmet et al., 2017; Desmet and Wacziarg, 2021). The novelty here is that we form groups based on homophily in values, without any regard to specific identity markers.

Our findings speak to the literature on the changing nature of political cleavages around the world. Salient recent examples include Gethin et al. (2022), Bonomi et al. (2021) and Enke et al. (2022). This literature pays a lot of attention to recent changes in voting behavior, party platforms and politically salient cleavages. In particular, they document a shift from socioeconomic issues dividing the electorate to cultural issues being the most relevant political cleavage. When considering the most efficient way of dividing the American public, we do not find a shift from a socioeconomic to a cultural divide. However, when we take into account the possibility of multiple equilibria, we find evidence that in the 1980s there existed an alternative partition that divided U.S. society by income class more than by moral values.

When endogenizing party platforms, the literature has considered different approaches. The Downsian model predicts that party positions converge to the views of the median voter. In

position individuals as probabilistic mixtures of these ideologies. They then measure polarization after forming groups based on people's dominant type share. In contrast to us, they find an increase in polarization in the U.S. between the early 1990s and the end of the 2000s.

contrast, others have argued that party platforms aggregate the preferences of their voters, so that party platform diverge (Baron, 1993; Ortuño-Ortín and Roemer, 2000; Gomberg et al., 2004). Our results are consistent with a shift from the Downsian model of political competition, where the mean positions of the voters of both parties are almost indistinguishable, to a model where party platforms increasingly diverge and now reflect the main values clusters in society.

2 Endogenous Partitions and Latent Polarization

In this section, we propose a micro-founded framework to measure polarization between latent values-based clusters. We start by providing a method to endogenously identify the latent values-based partitions of society, and then relate the antagonism between these endogenous groupings to the notion of latent polarization.

2.1 Partitions Based on Values

Consider a set S of J individuals in society. Each individual j is characterized by a values vector $v_j \in \mathbb{R}^Q$. There is homophily in values in the sense that individuals prefer to interact with others who have similar values. More specifically, an individual j experiences a level of disutility from interacting with individual k that is increasing in the distance between the vectors of values v_j and v_k:

$$u(v_j, v_k) = u(d(v_j, v_k)) \tag{1}$$

where $d(v_j, v_k)$ is a distance metric between v_j and v_k.

Values Identification Equilibrium (VIE). We assume that society is made up of a small number of non-overlapping groups, and that each individual can only affiliate with one group. For now, we set the number of groups to two. These groups are not exogenously defined, but are endogenously formed by individuals aiming to minimize their expected disutility from interacting or identifying with other members of their group.

Consider a partition A of the population set S into two groups, denoted A_1 and A_2, with $S = A_1 \cup A_2$ and $A_1 \cap A_2 = \emptyset$. Let $\mathcal{S}$ denote the set of all possible partitions of S into two groups, so that $A \in \mathcal{S}$. The expected disutility experienced by an individual with values v_j, when identifying with group A_i and interacting with each individual of that group with equal probability, is:

$$E(A_i, v_j) = \frac{1}{J_i} \sum_{k \in A_i} u(v_k, v_j) \tag{2}$$

where J_i is the number of people in group A_i. We refer to this expected disutility as the *antagonism* individual j experiences by identifying with group A_i. When deciding which group to identify with, individual j chooses group A_i over group A_{-i} if $E(A_i, v_j) \leq E(A_{-i}, v_j)$. We denote the group agent j belongs to by $A_{i(j)}$.

Definition of a Values Identification Equilibrium (VIE). *A partition $A \in \mathcal{S}$ is a Values Identification Equilibrium (VIE) if for each agent j we have $E(A_{i(j)}, v_j) \leq E(A_{-i(j)}, v_j)$.*

Thus, a VIE is a Nash equilibrium: taking as given the group identification of all other individuals, no agent wants to change her group identification. One can easily prove the existence of an equilibrium, because there is a finite number of individuals.[2] However, there may be multiple VIEs. We denote the set of possible VIEs by $\mathcal{V}$.

For any VIE $A \in \mathcal{V}$, average within-group antagonism in society is:

$$E(A) = \frac{1}{J} \sum_{j \in S} E(A_{i(j)}, v_j) \tag{3}$$

Within the set of possible VIEs, we refer to the one that minimizes within-group antagonism as the Global VIE, and denote that VIE by A^*. We denote the corresponding level of within-group antagonism by $E^* = E(A^*)$.

Definition of the Global VIE. *A^* is the Global Values Identification Equilibrium if for each $A \in \mathcal{V}$, $E(A) \geq E(A^*)$.*

2.2 Latent Values-Based Polarization

The partition that minimizes within-group value antagonism is also the partition that maximizes between-group differentiation. In fact, the well-known Φ_{ST} index of between-group differentiation can be written as:[3]

$$\Phi_{ST} = \frac{E(S) - E(A^*)}{E(S)} \tag{4}$$

where $E(S)$ is average overall antagonism in society when individuals are not partitioned into groups and everyone interacts with everyone else with equal probability.

Expression (4) defines between-group antagonism as the share of overall antagonism that does not occur within groups. It corresponds to our novel concept of *latent polarization.*

Definition of Latent Values Polarization. *Latent values polarization is equal to between-group differentiation Φ_{ST} when the VIE is A^*: the share of overall values antagonism that occurs between groups in the partition corresponding to the Global VIE.*

This novel definition of polarization captures the essence of how Esteban and Ray (1994) understand the concept. They view society as polarized if each cluster is similar in terms of the attributes of its members, but different clusters have members that are dissimilar. In our case, a society exhibits a high degree of latent polarization if within-group heterogeneity in values is low and between-group heterogeneity in values is high. As such, our definition generalizes the

[2] Gomberg et al. (2004) provide conditions for the existence of an equilibrium in the case of a continuum of agents.
[3] Appendix A.1 discusses the relationship between this index and the index of cultural differentiation in Desmet et al. (2017).

notion of polarization in Esteban and Ray (1994) to a setting where individuals' attributes are multi-dimensional.

We refer to our notion of polarization as *latent* because it is a measure of the underlying potential for polarization in society: since it is based on the partition that minimizes within-group heterogeneity and maximizes between-group differentiation, it captures the maximum scope for values-based polarization in society. We are not claiming that existing social groups (clubs, political parties, or other visible social groupings) necessarily reflect the underlying latent cultural groups. Indeed, society need not self-organize into these values-based clusters. Rather, in the empirical section of this paper, we ask whether and to what extent latent groups are reflected in actual groups, in particular those based on political party affiliation.

Since we are interested in measuring the maximum scope for polarization in society, our notion of latent polarization is based on the Global VIE. Of course, whenever there is more than one VIE, we could also define polarization for these other partitions. Comparing the degree of polarization in different VIEs may be of interest. For instance, if these VIEs display very different levels of latent polarization, we might overstate the extent of latent polarization when focusing only on the Global VIE.

2.3 Partitions into More Groups

We can easily generalize our approach to partitions of S into H groups A_h, with $h = 1, 2, ..., H$. In that case, a VIE would require that for each agent j, $E(A_{i(j)}, v_j) \leq E(A_{h(j)}, v_j)$ for any h. The Global VIE would then be the partition into H groups that minimizes within-group heterogeneity and maximizes between-group differentiation. We denote by A^*_H the Global VIE of a partition into H groups.

The corresponding degree of latent polarization for a partition into H groups is:

$$\Phi_{ST,H} = \frac{E(S) - E(A^*_H)}{E(S)} \tag{5}$$

When the number of groups increases, by construction the degree of within-group heterogeneity declines, and the degree of between-group differentiation goes up. Hence, latent polarization increases with the number of groups H. As a result, our measure of latent polarization should be interpreted as conditional on the number of groups. Therefore, when we compare the degree of latent polarization across time, it is important to keep the number of groups constant.

2.4 Comparison with Partisan Polarization

In the same way that we measure polarization between values-based clusters, we can measure partisan polarization between voters of political parties. Consider a partition $A^p \in \mathcal{S}$ based on partisan affiliation, with a first group A^p_1 containing Democratic voters, and a second group A^p_2 containing Republican voters. We keep the utility function of an individual unchanged. That is, an individual's utility continues to depend on the distance between his values and those of the individuals of his group. In general, the partisan-based partition A^p is not a VIE. The degree of values antagonism between partisan-based groups is then:

$$E(A^p) = \frac{1}{J} \sum_{j \in S} E(A^p_{i(j)}, v_j) \tag{6}$$

Thus, $E(A^p)$ measures within-group values antagonism in the case where individuals still only care about values, but where groups are defined by political party affiliation.

By analogy with latent polarization, we can define partisan polarization as the share of overall antagonism that is between partisan groups:[4]

$$\Phi^p_{ST} = \frac{E(S) - E(A^p)}{E(S)} \tag{7}$$

By definition, partisan polarization, Φ^p_{ST}, is (weakly) lower than latent polarization, Φ_{ST}. As such, group membership based on political parties implies more within-group antagonism but less polarization, compared to partitions strictly based on values.

A central goal of this paper is to analyze the relation between values-based clusters and partisan groups. Specifically, we are interested in understanding whether politics has become more or less reflective of underlying values-based clusters in society. Equations (4) and (7) enable a direct comparison between partisan polarization and latent polarization, because they provide a common polarization metric. Of course, comparing the level of partisan and latent polarization depends on the number of groups in each partition. In our baseline, we will focus on partitions into two groups, because the partisan divide is mainly between Democratic and Republican voters.

In this subsection, we considered a partition based on partisan affiliation. Needless to say, the same measure of polarization as in equation (7) can be applied to any partition. For example, we could split up people into groups, based on their gender, race, income, or education. In the empirical part of the paper, we compare latent and partisan polarization to other measures of polarization based on demographic or identity traits.

2.5 The Case of Squared Euclidean Distance

So far, we have not assumed a functional form for $u(d(v_j, v_k))$. In the rest of the paper, we take the disutility function (1) to be given by:

$$u(v_j, v_k) = \|v_j - v_k\|^2 \tag{8}$$

[4]This metric of between-group differentiation is closely related to the measure of cultural divides between identity groups, F_{ST}, used in Desmet and Wacziarg (2021).

where $\|v_j - v_k\|$ is the Euclidian distance between vectors v_j and v_k.[5] We use the squared Euclidian distance metric in our empirical application, but our framework can accommodate other metrics.

Using the squared Euclidean distance has several advantages. First, it allows us to use the standard k-means clustering method in order to create endogenous partitions. Second, adopting squared Euclidean distance implies an additive preference structure: the distance between two vectors can be computed by adding up distances in each of the Q dimensions. Third, in Appendix A.2, we show that the Global VIE A^* is the partition that minimizes within-group variance. This is equivalent to the partition that maximizes between-group variance. The between-group variance corresponding to partition A^* measures the degree of latent polarization in society.

Alternative interpretation of values identification. Consider two alternative ways to assign individuals to latent values-based groups. The first is the one we have assumed so far: each individual compares his values to those of all other individuals in his group and minimizes within-group antagonism. The second would be for each individual to join the group with a *mean* vector of values that is closest to his own. This type of group identification is closer to the one typically considered by economists (Akerlof and Kranton, 2000; Shayo, 2009; Bonomi et al., 2021), and perhaps a more realistic portrayal of how individuals reason about identifying with a group.

In what follows, we show that the two approaches deliver the same equilibrium partitions in large populations. Define an alternative definition of antagonism of individual j, based on his distance to the mean $\mu_{i(j)}$ of his group $A_{i(j)}$:

$$E'(A_{i(j)}, v_j) = \frac{1}{J_{i(j)}} \left\| v_j - \mu_{i(j)} \right\|^2 \tag{9}$$

In this case, we can view the mean value μ_i as the representative culture of group A_i, with the cost for an individual with values v_j to identify with group A_i given by the distance $\|v_j - \mu_i\|^2$. This yields an alternative definition of a VIE, which is the same as before except that antagonism is based on E'. We refer to this as a Mean Values Identification Equilibrium (MVIE):

Definition of a Mean Values Identification Equilibrium (MVIE). *A partition $A \in \mathcal{S}$ is a Mean Values Identification Equilibrium (MVIE) if for each agent j we have $E'(A_{i(j)}, v_j) \leq E'(A_{-i(j)}, v_j)$.*

In Appendix A.3, we prove the following proposition:

Proposition 1. *If the number of individuals is large enough and the distance metric is squared Euclidean, then any MVIE is a VIE.*

In other words, under the squared Euclidean distance assumption, if the number of individuals is large enough, then an equilibrium where each individual minimizes his distance to the group's

[5]This approach is related to that in Alesina et al. (2017), p. 183. They compute bilateral distances between all respondents of the European Values Survey using the squared Euclidian distance between vectors of individual answers. They then plot the densities of a monotonic function of these distances.

mean position is a VIE. As such, identification does not require an individual to know all bilateral distances to all other individuals, but only his distances to the mean positions of the groups.

3 Data and Algorithm

In this section, we describe the data and algorithm used to find endogenous values-based partitions.

3.1 Data

We use data from all seven waves of the World Values Survey for the United States. These waves correspond to 1981, 1990, 1995, 1999, 2006, 2011 and 2017. Vector v_j is given by agent j's answers to questions on values.

Sample of questions and respondents. We start with the integrated World Values Survey 1981-2022 and retain all questions that are about values and attitudes, disregarding those that concern respondents' demographic characteristics. These questions cover a wide and comprehensive set of issues that the WVS characterizes as related to life, work, family, politics, society, religion, morale, and identity. Among that set, we keep those with answers that can be ordered. These are either binary or ordered on a scale. We rescale answers so that they are always in the interval $[0, 1]$.

On average, we have 196 questions on values and attitudes per wave. To alleviate the issue of missing answers, we drop any question that is answered by less than 30 percent of the respondents. This eliminates an average of four questions, leaving us with 192 questions per wave. We also drop respondents who do not answer at least 70 percent of the questions. This only results in dropping an average of 0.7 percent of respondents. This leaves us with an average of 1,841 respondents per wave.

Appendix Table C1 reports the number of questions and respondents by wave. Over time, there is entry and exit of questions. As a result, the set of questions changes from wave to wave. Since the relevant issues also evolve over time, we view it as desirable to include for each wave the most comprehensive set of questions on values and attitudes. For example, if we were to drop questions that enter in later waves, we might ignore some of the more recent divisive issues. However, in the empirical section, we will explore the robustness of our findings to using a fixed set of questions across all seven waves.

Although we drop individuals and questions with a low response rate, we still have on average 3.1 percent of missing answers. To deal with this issue, we use a machine learning algorithm to impute their values. Let T be the set of respondents with no missing answers, and let $V_T = \{v_j : j \in V\}$ be a matrix whose columns correspond to the vectors of values of those respondents. Using V_T as the training sample, the machine learning algorithm yields a data matrix $V = \{v_1, v_2, ..., v_J\}$ with no missing values.[6]

[6] More specifically, we use the Mathematica (version 13.0.1) command "SynthesizeMissingValues" to replace miss-

Principal components analysis. Next, we reduce the dimensionality of the question space by using principal component analysis (PCA). There are multiple advantages to doing so. First, using PCA avoids the possible duplication of questions that capture similar values and are likely to be answered similarly by a given respondent. For example, there are separate questions on belief in Heaven and belief in Hell, with highly correlated answers. Second, to the extent that there is measurement error in the way individuals answer WVS questions, the use of principal components helps mitigate the problem. Third, by construction, PCA produces dimensions that are orthogonal to each other. In the case of using squared Euclidean as our distance metric, this allows an interpretation of the resulting measures of values antagonism as minimizing within-group variance in values (as captured by principal component positions - see Appendix A.2). Fourth, with fewer dimensions, finding a VIE is computationally less costly.[7]

For each wave, we compute the principal components of matrix $V = \{v_1, v_2, \ldots, v_J\}$. For each individual j we write the vector of her position on the different PC dimensions as $p_j = \{p_{j1}, p_{j2}, \ldots, p_{jQ}\}$, where Q is the number of questions.[8] We can use either the answers themselves or any number of principal components to find VIEs. In practice, we do the latter, and consider alternatively the first two, first three and first 75 principal components to create the endogenous partitions.[9] When using the first PC, the distance between individual j and individual k is given by:

$$u\left(p_{j1}, p_{k1}\right) = \left\|p_{j1}, p_{k1}\right\|^2 = \left(p_{j1} - p_{k1}\right)^2 \tag{10}$$

When using the first two PCs, the corresponding distance is:

$$u\left(\left\{p_{j1}, p_{j2}\right\}, \left\{p_{k1}, p_{k2}\right\}\right) = \left(p_{j1} - p_{k1}\right)^2 + \left(p_{j2} - p_{k2}\right)^2 \tag{11}$$

and so on for more dimensions.

3.2 Algorithm to Find Values-Based Partitions

Using bilateral distances between all individuals to find a VIE is computationally very demanding. Instead, we appeal to our equivalency result in Proposition 1, which allows us to rely on a simpler disutility function based on each individual's distance to the mean of each cluster, rather than the

ing values. The training sample used was formed by the answers given by the set individuals with no missing answers (T). We set the level of performance to "Quality" to maximize the synthesis quality. For each wave, Mathematica chooses the best machine learning algorithm from among "Multinormal", "Kernel Density Estimation", "Decision Tree", and "Gaussian Mixture". In wave 7, 58 percent of individuals were in the training sample. In rare cases, the algorithm can replace a missing value with a value that lies outside the $[0, 1]$ interval. In such cases, we assign a value of 0 or 1, depending on which is closest.

[7]We have verified that for wave 7, the results obtained *without* first reducing the question dimensionality using PCA are very similar to those obtained using PCA, in the sense that we obtain very similar clusters in both cases.

[8]In all cases, the number of questions is less than the number of individuals, so matrix $P = \{p_1, \ldots, p_N\}$ has Q columns.

[9]The share of the variance explained by the first two principal components ranges from 10.3% in wave 4 to 16.4% in wave 7, and increases over time. The share of the first 10 PCs ranges from 26.4% to 35.8% in waves 2 and 7. The case of 75 PCs is basically equivalent to considering all PCs.

sum of all within-group bilateral distances. With a distance metric that is squared Euclidian, in a VIE each individual is at a smaller distance from the mean answers of her own group than from the mean answers of any other group. To implement this, we use the k-means clustering algorithm in *Mathematica*. We use the sampling weights of the WVS to ensure that our underlying samples are representative.

Although our main focus is on partitions into two groups, we also consider partitions into three, four and five groups. As for the principal components, our main focus is on two principal components, but we also consider one, three and 75 principal components. For each of these 16 combinations, and for each wave, we need to find the Global VIE (A^*). To do so, we run the algorithm 1,000,000 times, using different random starting points. We then select the partition with the lowest antagonism, and denote this partition as A^*.

While in the later waves we find a unique VIE, in the earlier waves we detect multiple VIEs. Our main focus is on the Global VIE, because we are interested in getting a measure of the maximum scope for polarization of the U.S. public. However, the existence of other VIEs in early time periods may help explain the changing nature of partisan polarization, as we clarify below.

3.3 Validation Exercise

As a validation exercise, we ask if our methodology can identify the inhabitants of different countries from their positions in cultural space. We consider three countries on three different continents: the United States, China and Zimbabwe. We pool the respondents from Wave 7 of the WVS for these three countries. We then run principal components analysis on this joint sample.[10] Plotting individuals along the first two PCs makes the three countries appear distinctly (Figure 1, Panel A). We next run our algorithm on the pooled data (Figure 1, Panel B), allowing for three clusters. Our goal is to see whether our algorithm can recover the countries that constitute the pooled data. The results are telling: 96.5% of the individuals from China are classified in cluster 1 (in blue in the figure), 98.2% of the individuals from Zimbabwe are classified as belonging to cluster 3 (in green in the figure) and 79.5% of the individuals from the U.S. belong to cluster 2 (in orange in the figure). Most of the remaining US respondents (18.2%) are assigned to cluster 3 (the "Zimbabwe cluster").[11] Overall, our algorithm does very well at recovering the three underlying countries.

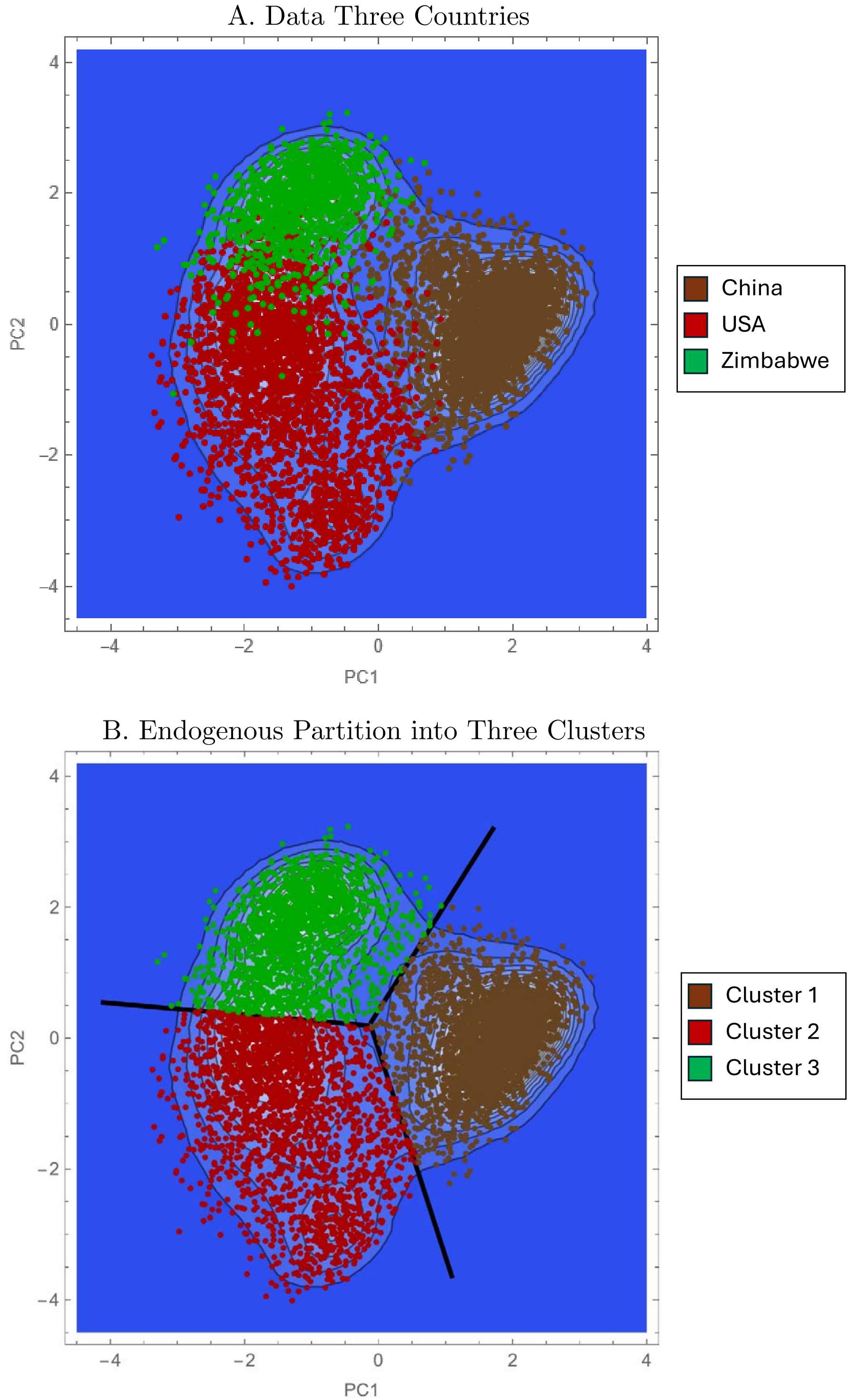
4 Latent Polarization in the U.S.

In this section, we document that the degree of latent polarization of the U.S. public has been high and stable for at least the last four decades.

[10]We find that the first PC explains 13% of the variance in answers, while the first 10 PCs explain 40%. These variance shares are not very different from those that we described above when looking at cultural variation within the United States.

[11]This 18.2% of the Americans sample of respondents consists of 246 Whites, 58 Blacks, and 74 Hispanics. As a percentage of the total of each group in the US they are: 14% of White respondents, 27.6% of Black respondents, and 16% of Hispanic respondents.

Figure 1: Values-Based Partition of China, US and Zimbabwe



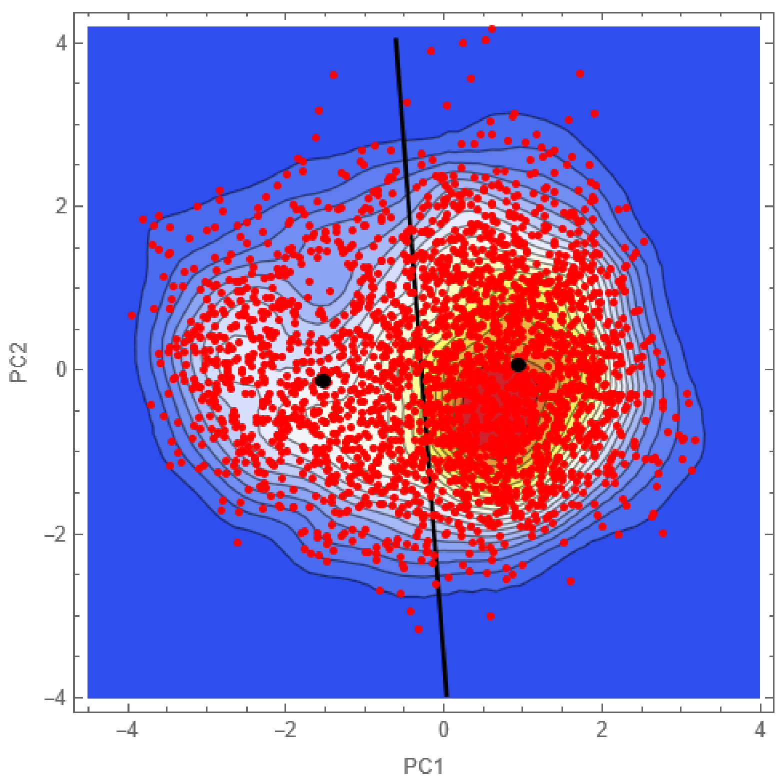
	China	USA	Zimbabwe
Cluster 1	96.5	2.2	0.0
Cluster 2	1.0	79.5	1.7
Cluster 3	2.4	18.2	98.2

4.1 The Degree of Latent Polarization

We start by focusing on the most recent wave of the World Values Survey, corresponding to 2017. To assess the extent of latent polarization, we compare it to polarization between partisan groups and between socio-demographic groups. Our baseline analysis measures latent polarization based on partitions into two groups, but we also examine polarization when allowing for more than two groups.[12] To visualize the clustering of individual respondents of the WVS into groups, we mainly focus on partitions based on the first two principal components.

Visual representation of two clusters. Starting with 198 questions for wave 7 of the WVS, we take the respondents' positions along the first two principal components. We then apply the k-means clustering algorithm 1,000,000 times, with the goal of partitioning the respondents into two groups. This yields a unique VIE, and we hence take it to be the Global VIE. Figure 2

Figure 2: Values-Based Partition into Two Clusters (2 PCs)



Note: Figure shows the partition of U.S. respondents to wave 7 into two clusters. The k-means clustering algorithm is applied to the positions of individuals along the first two principal components of their responses to 198 questions.

depicts this partition of the U.S. into two clusters. Each dot in the plot represents an individual's

[12] We use two-group partitions as our baseline because one of our main interests is to compare latent polarization to partisan polarization, and in the U.S. the latter involves two parties.

position along the two principal components. The contour lines show the density of individuals in this two-dimensional space. The black line separates the two clusters, whereas the two black dots correspond to the mean position of each cluster. The k-means clustering algorithm that generates this partition implies that each member of a cluster is closer to the mean of her cluster than to the mean of the other cluster.

The dividing line in Figure 2 is almost vertical, indicating that the two clusters are formed mostly along a single dimension. The preponderance of the first dimension is also obvious when comparing the means of the two clusters, which are mainly differentiated along the horizontal axis. The first principal component tends to load heavily on WVS questions related to religious and moral values. Specifically, the 5 questions with the largest loadings for the first principal component are: "Believe in: hell"; "Believe in: heaven"; "How important is God in your life"; "Important child qualities: religious faith"; and "Believe in: God". The second principal component mostly relates to social capital. The questions with the largest loadings for the second principal component include: "Most people can be trusted"; "Interest in politics"; "Political action: attending lawful/peaceful demonstrations"; and "How often do you attend religious services".

An important observation is that neither the first nor the second principal component attaches high loadings to questions related to economic issues, such as redistribution. This seems to indicate that the U.S. public is less divided on economic questions than on moral values. This observation comes with two caveats. First, while the first two principal components do not directly relate to economic questions, there could still be an important income gap between members of the clusters. However, as we show later, this is not the case: the mean income difference between clusters is minimal. Second, these findings might be driven by the structure of the World Values Survey, if there is a preponderance of questions on morals and religion in that survey. Later in this subsection, we argue that our findings are robust to this concern.

Degree of latent polarization. As can be seen in Table 1, the level of latent polarization between the two clusters is 41.2. Recall that latent polarization is measured as the share of overall antagonism that is between groups. It is high when members of each cluster have more similar positions along the first two principal components (within-group cohesion), but members of different clusters are quite different (between-group heterogeneity).

The degree of latent polarization in the U.S. is about average when compared to that in other countries. We applied our method to data from the latest wave of the integrated World Values Survey / European Values Survey across 81 countries, finding that the average degree of latent polarization stood at 41.7, with a standard deviation of 6.1. In countries such as Colombia, Mexico and Spain, latent polarization is higher, whereas in other countries, such as Japan and Jordan, it is lower.

Comparison with partisan polarization. While the degree of latent polarization in the U.S. is average compared to other countries, it is high compared to the degree of partisan polarization. If we split up the respondents of the WVS into groups defined by who they would plan vote for "if there were a national election tomorrow", the degree of partisan polarization based on the first two principal components is 14.7. Hence, latent polarization among the U.S. public is almost three times larger than polarization between voters of different political parties. At the same time, partisan polarization in the U.S. is almost one standard deviation higher than average partisan polarization across the 81 countries in our worldwide sample.

For our comparison of latent and partisan polarization to be valid, both measures must be based on partitions of society into the same number of groups. Because almost everyone votes for the Democratic or the Republican Party, we can directly compare our measure of partisan polarization to the one of latent polarization between two groups. It turns out that the small minority of respondents who vote for other parties, such as the Libertarian or the Green party, have no material impact. In fact, whether we exclude or include voters for other parties in the sample, the degree of partisan polarization is equal to 14.7.[13]

Table 1: Polarization for 2 PCs and 2 Clusters (WVS Wave 7)

A. Latent Polarization	41.2		
B. Polarization between Other Identity Groups (2 Groups)			
Partisan	14.7	Ethnicity	1.1
Religion	10.0	Education	6.2
Income	3.3	Social Class	2.9
Gender	1.1		

Notes: All polarization measures are between two groups: partisan (Democratic vs Republican), religion (religious denomination vs no religious denomination), income (above and below the median), gender (women and men), ethnicity (white and non-white), education (above and below the median), and social class (lower and middle-lower vs middle upper and upper). Appendix Table B1 reports polarization measures, using all the groups available in the WVS.

Comparison with demographic partitions. As a further way of assessing whether the degree of latent polarization in the U.S. is relatively high or relatively low, we compare it to the level of polarization between groups defined by particular demographic traits. We look at six types of demographic partitions. These are based on religion, income, ethnicity, education, social class, and gender. To facilitate comparison with latent and partisan polarization, we reduce each demographic

[13] In WVS wave 7, 24.5% of respondents would vote for neither the Republican Party nor the Democratic Party. In the baseline exercise, we exclude these respondents. When we include them, and treat their partisan alignments as separate groups (Libertarian, Green, Other and don't know/no answer), we find the same degree of partisan polarization as in the baseline.

trait to two groups.[14] More specifically, for religion we compare 'religious denomination' versus 'no religious denomination', for income we take 'above the median' and 'below the median', for ethnicity we consider 'white' and 'non-white', for education we analyze 'below' and 'above median', for social class we define 'lower' and 'middle lower' vs 'middle upper' and 'upper', and for gender we have 'men' and 'women'. For each one of these demographic traits, we consider how divided these groups are along the first two principal components. For example, in the case of gender, we compute the share of the overall heterogeneity in individuals' positions along the first two principal components that is between men and women.

As can be seen in Panel B of Table 1, in general polarization between groups based on demographic traits is low. For example, if partitions are formed based on gender, polarization in the U.S. would be a mere 1.1. We find similarly low levels of polarization if groups are formed based on ethnicity (1.1) or income (3.3). Polarization is somewhat larger when groups are defined by the religious and the non-religious (10.0), though this figure is still four times lower than latent polarization (41.2).

It is worth re-emphasizing that we are measuring polarization as the share of society-wide heterogeneity in values that is between groups. For example, polarization between income groups is not measured by how polarized people are in terms of their income, but rather by how polarized people of different income groups are in terms of their values. As explained before, the assumption is that individuals have a preference for interacting with like-minded individuals. The only difference between latent polarization, partisan polarization, or demographic polarization is the composition of the group members. Our results indicate that the degree of latent polarization of the U.S. public is much higher than both partisan and demographic polarization.

An alternative interpretation of our findings is that socially identifying with others based on demographic traits, such as ethnicity or gender, is 'inefficient'. For example, the fact that polarization between ethnic groups is 1.1 means that antagonism would only fall by 1.1 percent if (non-)whites were to interact exclusively with other (non-)whites, rather than with the society at large. Interacting with individuals based on whether they are religious or not is somewhat more efficient: heterogeneity in values would drop by 10.0 percent when people partition into a religious and a non-religious group. In contrast, identifying directly based on homophily in values leads to a reduction of 41.2 percent in heterogeneity. This is the sense in which we say that latent polarization is high.

Demographic characteristics of clusters. Given that groups based on demographic traits are substantially less efficient in reducing antagonism than groups based directly on homophily in values, we would expect demographic differences between the endogenous values-based clusters to be relatively small. This is in general what we see in Table 2, Panel A: when we endogenously

[14] In Appendix B.1, we conduct the same comparison with the original group definitions of the WVS, which in most cases involves more than two groups. Our conclusions are qualitatively unchanged, in the sense that polarization based on these groupings remains much smaller than latent polarization.

partition the U.S. into two clusters with the goal of minimizing within-group heterogeneity in values, the difference in, for example, the average years of schooling between the two groups is only 1.48 years. The corresponding differences in income is 0.11 (on a scale of 1 to 10) and the difference in age is only 2.01 years. In contrast, we find somewhat higher differences in the share of males (14.57 percent), the share of whites (13.59 percent), and the share belonging to the lower and middle-lower social class (9.83 percent). Overall, the fact that demographic differences between clusters tend to be relatively small confirms that values-based partitions are not too strongly aligned with demographic traits. Even for those demographic traits that exhibit larger differences, there is still substantial demographic heterogeneity within each of the two clusters.

Table 2: Differences between Cluster 1 and Cluster 2 (WVS Wave 7)

A. Difference in Demographic Traits between Clusters			
Male (percent)	14.57	Social Class (percent)	9.83
Age (years)	2.01	Education (years)	1.48
Income (scale 1-10)	0.11	White (percent)	13.59
B. Difference in Specific Questions between Clusters (0-1 Scale)			
Importance Family	0.05	Immigration Policy	0.11
Importance Religion	0.52	Fight for Country	0.15
Housewife Fulfilling	0.09	Believe in Heaven	0.64
If Don't Work, Become Lazy	0.14	Attend Religious Services	0.37
Generalized Trust	0.20	Homosexuality Justifiable	0.43
Confidence in Government	0.24	Abortion Justifiable	0.38
Membership Church or Religion	0.41	Election Officials Are Fair	0.09
Government vs Individual Responsibility	0.19	Democracy	0.13
Hard Work vs Luck	0.10	State Makes Income Equal	0.01

Values differences between clusters. Panel B of Table 2 reports the between-cluster differences in answers to some of the questions that receive the highest weights on the first two principal components, as well as to some of the questions related to economic issues.

Recall from Figure 2 that the endogenous division into two clusters is mainly along the first principal component. Hence, we would expect individuals of the different groups to be more divided on religious and moral questions than on other values. This is indeed the case: the difference in the mean positions between the two clusters is 0.64 for the question 'Believe in: heaven', 0.52 for the question 'Importance in life: religion', and 0.43 for the question 'Justifiable: homosexuality'. The differences in the mean positions on questions related to social capital that carry a high weight on the second principal component are on average a bit smaller, though still relatively large: 0.41 for the question 'Membership: church or religious organization', and 0.20 for the question 'Most people can be trusted'.

For questions related to economic values, the differences are markedly smaller: 0.11 for the question related to 'immigration policy', 0.10 for the question "Hard work brings success" and 0.00 for the question on whether it is an essential characteristic of democracy for the state to make people's income equal. In other words, differences between latent clusters mostly reflect cultural values and not the divisions over redistributive policies that traditionally define political cleavages (Gethin et al., 2022).

Number of dimensions. So far, our focus has been on the first two principal components. We could of course do the same analysis with more principal components. Within-group heterogeneity becomes progressively larger with more principal components, because more dimensions of heterogeneity make it more difficult to efficiently group individuals: two individuals can share an affinity along some dimensions but not others. Correspondingly, with more principal components, between-group differentiation becomes smaller and latent polarization decreases.

When increasing the number of principal components from two to 75, latent polarization drops from 41.2 to 8.3. This means that heterogeneity in values can drop by at most 8.3 percent when people partition into two clusters based on people's positions on 75 principal components. Importantly, increasing the number of principal components does not change our basic result. As reported in Table 3, when using 75 principal components, latent polarization (8.3) continues to be substantially larger than partisan polarization (4.3), and the difference is even higher when the comparison is with polarization between demographic groups (e.g., 1.1 in the case of ethnicity and 2.4 in the case of religion).

Table 3: Polarization for 75 PCs (WVS Wave 7)

A. Latent Polarization	8.3		
B. Polarization between Other Identity Groups (2 Groups)			
Partisan	4.3	Ethnicity	1.1
Religion	2.4	Education	1.7
Income	1.0	Social Class	0.9
Gender	1.2		

Cultural and economic divisions. Our analysis indicates that the two main dimensions that explain most of the heterogeneity across Americans have to do with religious/moral values and attitudes related to social capital. Divisions on economic issues seem less important.

One reason might be the composition of the World Values Survey, which focuses on values, norms, and attitudes. This is unlikely to be a major concern for at least two distinct reasons. First, the WVS does have many questions related to people's economic values and views, such as

"Democracy: People receive state aid for unemployment"; "Hard work brings success"; and "Income equality". Second, when applying the same analysis to other countries, the relevant dimensions change. For example, in the case of Nigeria, the first principal component mainly picks up political action, and the second principal component relates to confidence in institutions.

That said, if we were to select questions in the WVS so that only half were about cultural values (as opposed to economic views), we would expect economic issues to become more important. This is indeed what we find, but only to a limited extent. In an exercise where economics and culture each make up half of a subset of 78 WVS questions, the first principal component continues to be about morals and religion.[15] The second principal component now corresponds to a mix of social capital and economic questions.[16] We see the same relative importance of economics when taking the correlation between respondents' income and their positions along both principal components: an absence of correlation for the first principal component (-0.05) and a positive correlation for the second principal component (0.258). However, when considering the Global VIE that emerges from these alternative questions, our previous results are unchanged: society continues to be divided mainly along the first principal component (Figure 3). That is, the moral/religious dimension continues to be the main societal cleavage, even when we artificially increase the weight of economic questions in the survey.

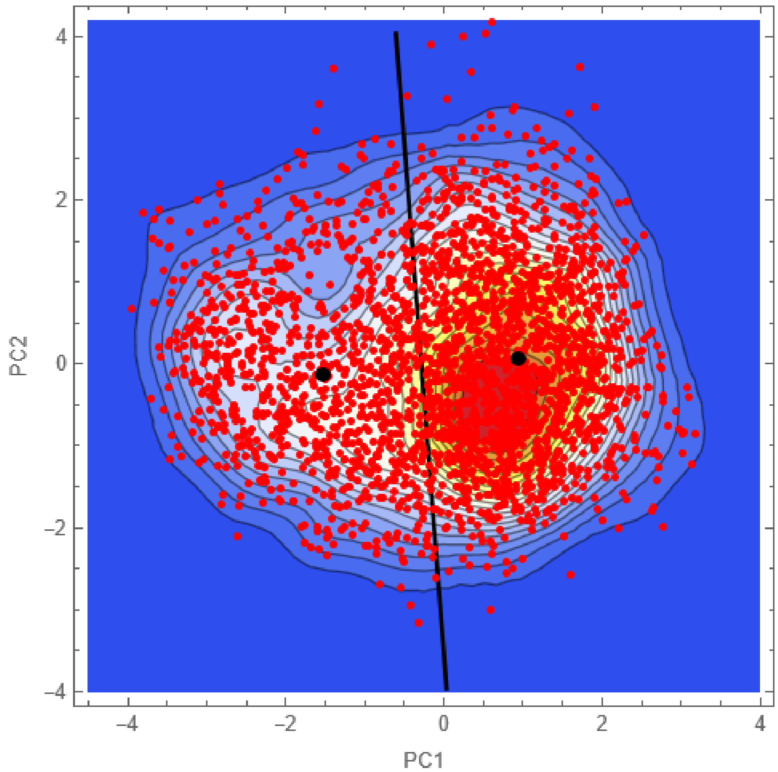
Number of clusters. As mentioned before, our main focus is on partitions into two clusters, because we want to compare the level of latent polarization to the level of partisan polarization in the U.S. context, where two political parties dominate. However, it is interesting to assess the characteristics of latent value-based groups when allowing for more than two clusters. Figure 4 depicts the endogenous partitions of individuals into three and four clusters, based on the first two principal components of answers. Recall from Figure 2 that the partition into two groups occurs mostly along the first principal component, with the group on the right scoring higher on religiosity (as captured by questions such as belief in heaven and the importance of God). When allowing for a third group, Panel A of Figure 4 shows the more religious group splitting into two groups, mostly along the second principal component, capturing questions on social capital (including questions on trust, political participation, and church attendance). When allowing for a fourth group, the two U.S. clusters on the right (corresponding to religious individuals, with either high or low levels of social capital) get subdivided into three clusters, while the group of more secular respondents on the left side of the figure remains largely unaffected.

This suggests that the main cleavage in the U.S. is between a minority of less religious liberals and a majority of more religious conservatives. The finding that the secular group does not further

[15] In this exercise, the 5 questions with the highest loadings for the first principal component are "Justifiable: Homosexuality", "Important in life: Religion", "Justifiable: Abortion", "Believe in: God", "Jobs Scarce: Employers should give priority to Americans", and "How often do you pray".

[16] The five questions with the highest loadings are "Most people can be trusted", "How often in a country's elections: Election officials are fair", "Democracy: The state makes people's incomes equal", "Worries: Not being able to give one's children a good education", and "Worries: Losing my job or not finding a job".

Figure 3: Values-Based Partition: Half of Questions Based on Economic Issues

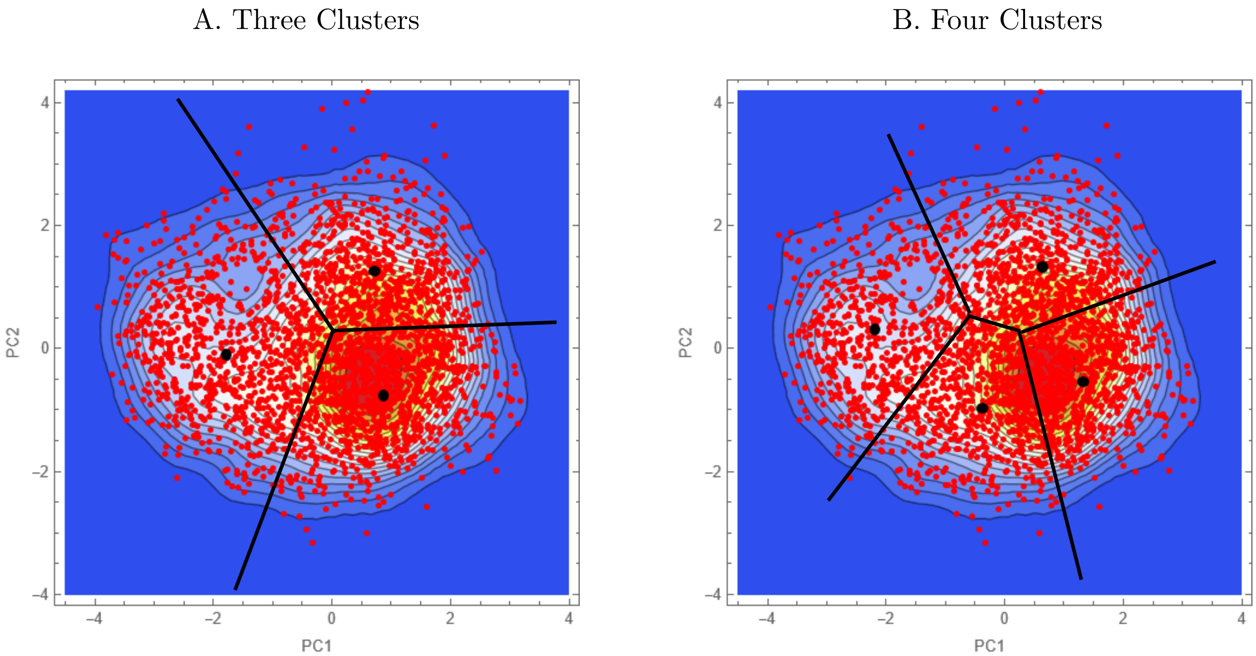


Note: Figure shows the partition of U.S. respondents to wave 7 into two clusters. The k-means clustering algorithm is applied to the positions of individuals along the first two principal components of their responses to 78 questions, of which half are about economics and half about culture.

split into smaller groups when we allow for more clusters reflects both their smaller size and their degree of distinctiveness.

When the number of clusters increases, within-group heterogeneity mechanically decreases, so that between-group differentiation increases (by construction within-heterogeneity is zero if each individual is its own cluster): the degree of latent polarization must be understood as conditional on the number of groups. When going from two to three clusters, latent polarization increases from 41.2 to 61.9. This number would further increase to 70.5 percent in the case of four clusters, and 75.2 percent in the case of five clusters (Figure 5). Hence, if latent polarization is already high when considering two clusters (compared to partisan and demographic polarization), it is even higher if we allow for more than two clusters. In some cases, the relevant comparison requires more than two groups. For example, the polarization between ethnic groups of 1.5 reported in Panel B of 1 is based on five different groups, and should hence be compared to the latent polarization between five clusters of 75.2.

Figure 4: Values-Based Partition into Three or Four Clusters (2 PCs)



Note: Figure shows the partition of U.S. respondents to wave 7 into three (panel A) or four clusters (panel B). The k-means clustering algorithm is applied to the positions of individuals along the first two principal components of their responses to around 200 questions.

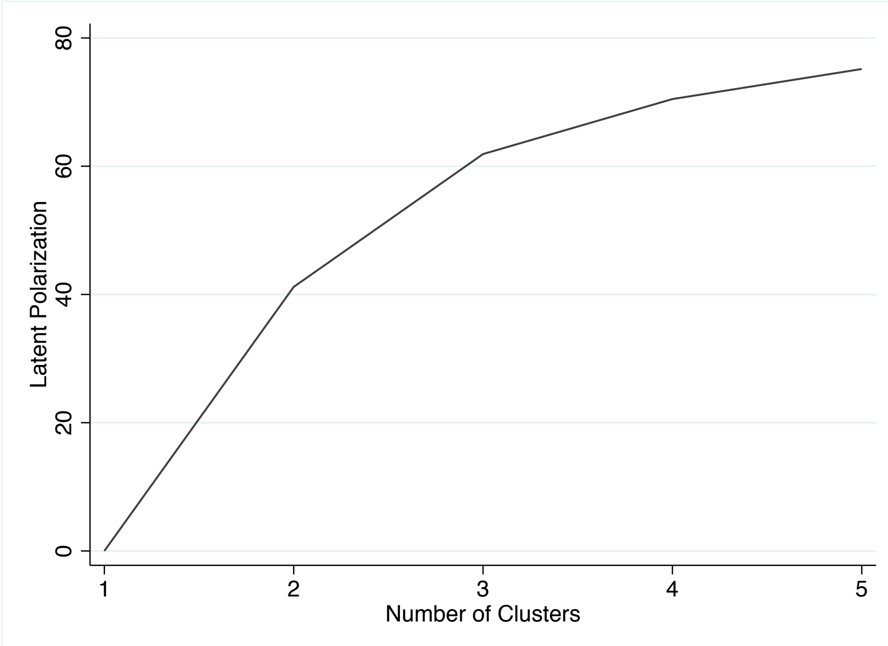
4.2 Latent Polarization Has Been Fairly Stable for Four Decades

In the previous subsection we concluded that in recent years latent polarization in the U.S. has been relatively high; we now turn our attention to exploring time variation in the degree of latent polarization across successive waves of the WVS.

For each of the seven waves, we determine the first two principal components of individuals' responses to all questions that relate to values and attitudes. Using the respondents' positions along the two PCs, we run the k-means clustering algorithm 1,000,000 times to partition the respondents into two stable groups. In contrast to wave 7, in some of the earlier waves of the WVS we find multiple VIEs – there is more than one way of partitioning society consistent with a Nash equilibrium defined by no agent having an incentive to switch groups. In most of our subsequent discussion, we focus on the Global VIE, corresponding to the VIE that minimizes within-group heterogeneity based on the 1,000,000 times that we run the clustering algorithm. However, the existence of other VIEs is informative, and we return to this question at the end of the subsection.

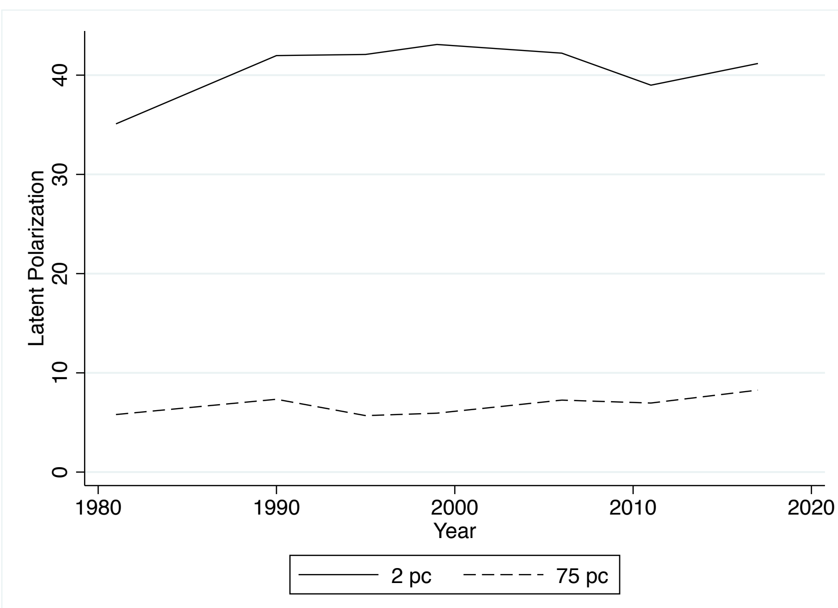
Stability of latent polarization. Figure 6 depicts the evolution of latent polarization in the U.S. between 1981 and 2017. Focusing on the case of two principal components, we see that latent polarization has been fairly stable. In fact, if we regress the degree of latent polarization on a time

Figure 5: Latent Polarization: Different Number of Clusters (2 PCs)



Note: Figure shows the degree of latent polarization when partitioning U.S. respondents to wave 7 into two, three, four, or five clusters. The partition is based on minimizing the degree of within-cluster heterogeneity of the positions of individuals along the first two principal components of their responses to around 200 questions.

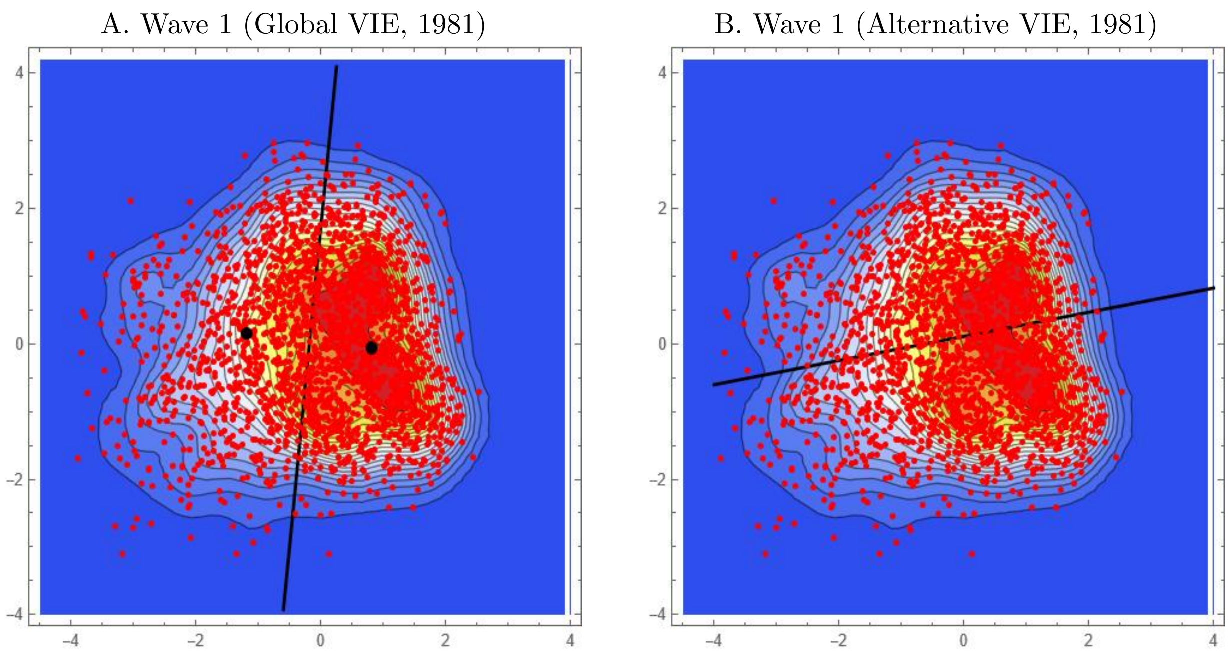
Figure 6: Latent Polarization: 1981 to 2017



trend that corresponds to the year of the different waves, the coefficient is statistically insignificant at the 10 percent level. When computing latent polarization based on 75 principal components, results are similar, with possibly a slight upward trend since the 2000s.

Stability of dividing cleavages. Our analysis is based on principal components that are computed separately wave by wave. The reason for doing so is to ensure that in each period we capture the most relevant dimensions of heterogeneity between people. However, when comparing the first two principal components between 1981 and 2018, the dimensions that best explain the heterogeneity across people do not change. In 2017, the first principal component captures religious/moral values, and the second principal component relates to social capital. For the 1981 wave, this is the case too.[17]

Figure 7: Values-Based Partitions into Two Clusters: Wave 1



In addition, comparing Panel A of Figure 7 and Figure 2 shows that the two clusters are formed mostly along the horizontal axis, both in 1981 and 2017. As such, for the last 40 years, the optimal endogenous partition has occurred along the religious/moral dimension. This suggests that the culture wars have been latent in U.S. society for many decades. Although it may not have expressed itself politically until more recently, the latent preconditions for this culture war already existed in the early 1980s.

Multiple equilibria in earlier waves. Authors such as Bonomi et al. (2021) have argued that political cleavages in the U.S. have shifted from the redistributive axis to a cultural axis. So far, our

[17] In 1981, the five questions that load most heavily on the first principal component are: "How often do you attend religious services"; "Important for successful marriage: Religious beliefs"; "Abortion when woman not married"; "Abortion if not wanting more children"; and "Member: Belong to none", whereas the five questions that load more heavily on the second principal component are: "Voluntary work: Unpaid work none"; "Political action: joining in boycotts"; "Political action: Signing a petition"; "Member: Belong to none"; "Important child qualities: Good manners".

results do not indicate the religious/moral cleavage being a recent phenomenon: the most efficient partition has been along cultural lines for at least four decades. However, in the earlier waves of the WVS there are multiple equilibria. When analyzing these alternative equilibria, a more complex picture emerges.

For wave 1 (1981), Figure 7 shows the existence of multiple equilibria. Panel A depicts the division into two clusters corresponding to the Global VIE, whereas Panel B depicts the division corresponding to one of the two other VIEs that we identify (these two other VIEs are very similar to each other). As can be seen, in the alternative VIE people are mostly divided along the second principal component. In that case, the values that divide respondents have to do with social capital, rather than with moral/religious issues.

Table 4: Multiple Equilibria

	Frequency	Latent Polarization	Difference between Clusters				
			Size	Men	Age	Income	Religious
Wave 1							
Global VIE	2.29	35.1	20.64	8.38	7.81	0.11	13.58
VIE	37.41	33.7	4.77	7.63	1.98	1.62	0.55
VIE	60.30	33.5	0.81	5.01	1.36	1.25	4.87
Wave 2							
Global VIE	94.01	42.0	29.03	3.70	7.26	0.35	38.68
VIE	5.99	26.7	14.25	3.87	4.92	0.96	0.12
Wave 3							
Global VIE	100.00	42.1	32.42	10.62	7.21	1.01	30.07
Wave 4							
Global VIE	99.97	43.1	20.42	12.79	4.09	0.03	31.53
VIE	0.03	25.1	15.10	2.18	3.14	1.50	5.00
Wave 5							
Global VIE	100.00	42.2	14.03	10.51	4.04	0.26	27.48
Wave 6							
Global VIE	100.00	39.0	23.52	2.80	4.46	0.38	40.03
Wave 7							
Global VIE	100.00	41.2	35.48	14.57	2.01	0.12	28.54

Table 4 reports some characteristics of the different equilibria for the different waves. In wave 1 (1981) and wave 2 (1990) there is more than one VIE. From wave 3 (1996) onwards, there is a unique equilibrium.[18] When focusing on the Global VIE of the different waves, religion is an important cleavage between clusters, whereas income is not. For example, in wave 1 the share of religious people is 13.58 percent higher in one cluster compared to the other, whereas income is only 0.11 points higher (on a scale of 1 to 10). The corresponding figures for wave 4 are 31.53 percent for religion and 0.03 points for income. In contrast, when focusing on the alternative VIEs in waves 1 and 2, the reverse is true: religious differences between clusters are relatively small,

[18] To be precise, in wave 4 (1999), there is also a second equilibrium, but it only occurs with a frequency of 0.03 percent (when running the algorithm 1,000,000 times).

whereas income differences are relatively large. For example, in wave 1, the religious difference is 0.55 and 4.87 percent in the two alternative equilibria (compared to 13.58 percent in the Global VIE), whereas the income difference between clusters is 1.98 and 1.36 points (compared to 0.11 in the Global VIE).

To further assess the alternative VIEs in waves 1 and 2, we are interested in how 'efficient' these equilibria are, as well as in the share of the 1,000,000 runs that generate these alternative VIEs. In wave 1, we see that the Global VIE corresponds to only 2.29 percent of the runs. In addition, latent polarization drops only slightly when going from the Global VIE (35.1) to the alternative VIEs (33.7 and 33.5). In that sense, the alternative equilibria are almost as 'efficient' as the Global VIE, and the algorithm generates them with a high probability. In contrast, in wave 2, the Global VIE is generated in 94.01 percent of the runs, and the difference in polarization is large, from 42.0 in the Global VIE to 26.7 in the alternative VIE. In that sense, the alternative equilibrium is much less 'efficient' than the Global VIE, and the algorithm generates it with a low probability.

From this, we conclude that in the earlier waves, there was an alternative equilibrium VIE that divided society more along people's income, and less along people's religiosity. In addition, in this alternative equilibrium the values that divided people had to do more with social capital and less with religious or moral issues. Moreover, in the earliest wave this alternative VIE is generated with a high probability. In the later waves, this alternative equilibrium vanished, and the only VIE becomes one where society is divided by moral/religious values.

These findings are related to our earlier discussion on the changing nature of political cleavages. Gethin et al. (2022), Bonomi et al. (2021) and Enke et al. (2022) document a shift from socioeconomic issues dividing the electorate to cultural issues being the most relevant political cleavage. While we find that the Global VIE has always divided the U.S. population along moral/religious lines, in the 1980s and early 1990s there was an alternative VIE where the two clusters showed more significant income differences. In sum, in later waves there appears a clear latent cleavage based on moral and religious values, whereas in earlier periods the nature of latent cleavages was more ambiguous, due to the existence of alternative VIEs.

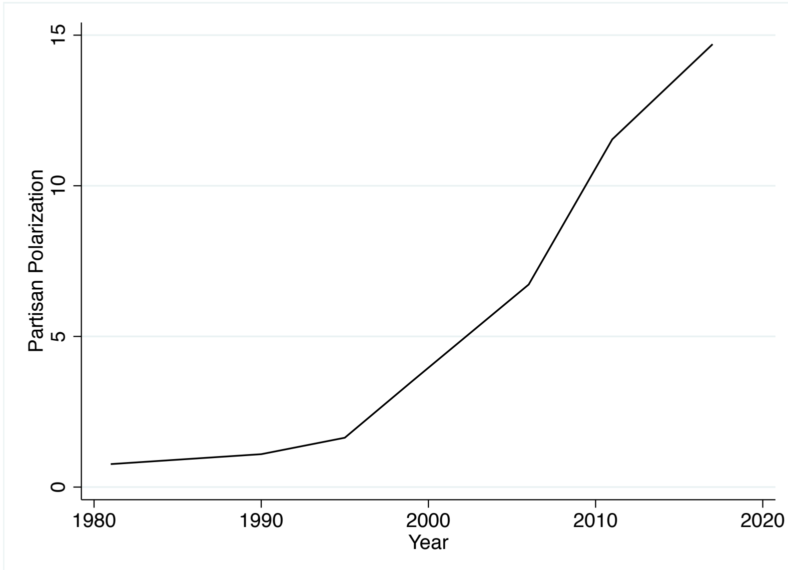
5 Partisan Polarization

In this section we analyze the evolution of partisan polarization over time, and ask how it relates to the degree of latent polarization of the U.S. public.

5.1 The Rise of Partisan Polarization

As can be seen in Figure 8, partisan polarization between voters of the Republican Party and voters of the Democratic Party has increased over time. Before 2000, partisan polarization was consistently below 5 percent. Around 2000, it started to increase, reaching 7 percent by the mid-2000s, and then further rose to around 15 percent in the last decade. This is consistent with recent work documenting a rising political divide in the U.S. (Desmet and Wacziarg, 2021; Boxell et al., 2022).

Figure 8: Partisan Polarization: 1981 to 2017



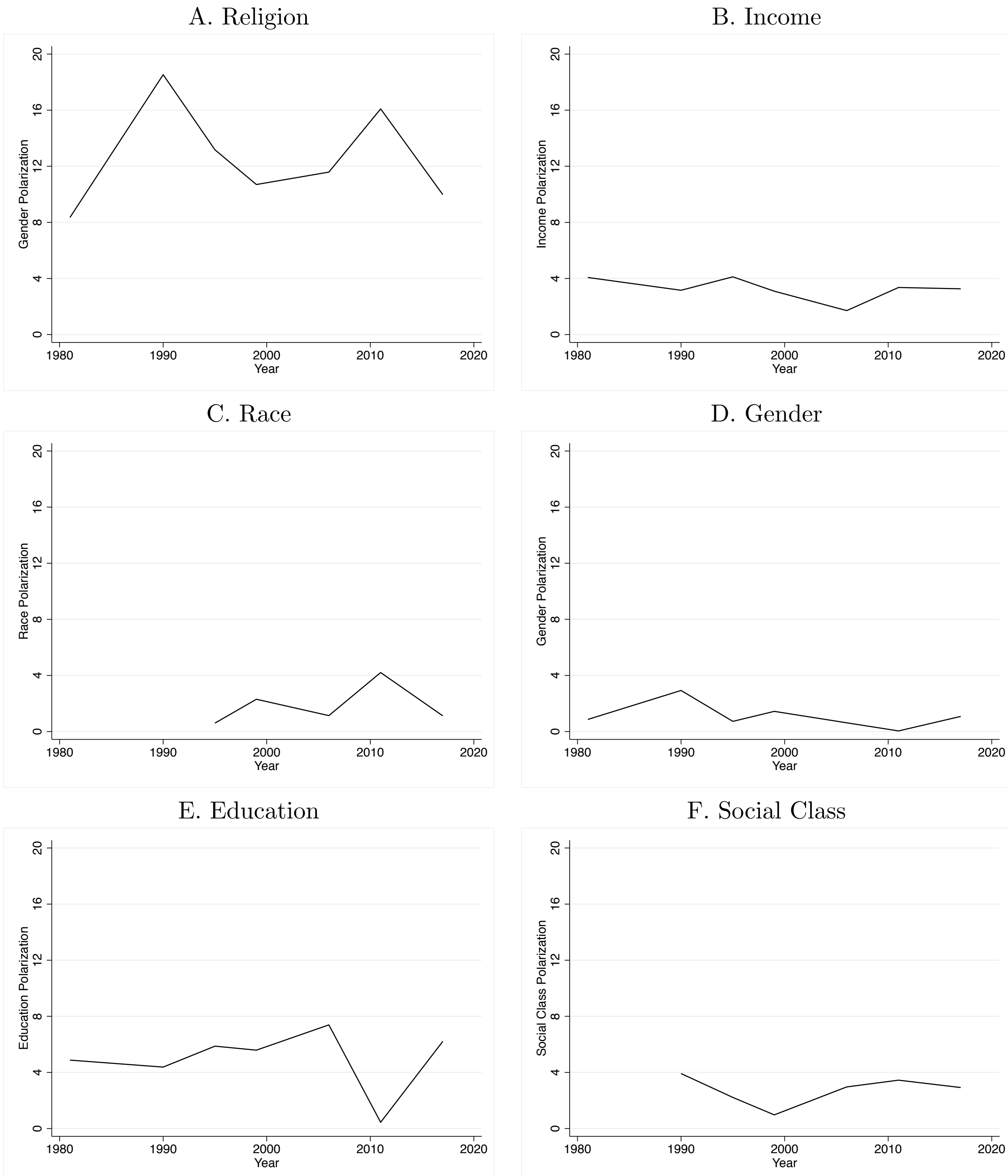
In contrast, when we look at polarization based on demographic or identity groups, such as gender, ethnicity, income, religion, social class, or education, there is no statistically significant trend (Figure 9). In this context of stable polarization between many identity traits, and of stable latent polarization, the rise in partisan polarization documented in Figure 8 stands out.

5.2 Increasing Partisan Sorting

In what precedes, we showed that latent polarization has been relatively stable over time. We also just showed that partisan polarization has increased. How can these two facts be reconciled? The answer could be increasing alignment between values and voting behavior, with voters increasingly sorting into different parts of the cultural space. Figure 10 depicts the positions of all respondents to the WVS along the two main principal components. Red dots refer to Republican voters, and blue dots to Democratic voters. The black dots refer to the mean positions of both types of voters, and the black lines separate individuals into two clusters based on minimizing the distance to these mean positions. Two results stand out.

First, the dividing line into two voting blocks changes from being more horizontal in the early 1980s to increasingly vertical over time. This reflects the relevant political cleavage shifting from the social capital dimension (principal component 2) to the moral/religious dimension (principal component 1). This change is also consistent with the political divide being more related to income in the earlier waves, and more related to religion in the later waves. In fact, in wave 1, when the political divide is more along principal component 2, the correlation of that principal component

Figure 9: Polarization between Demographic or Identity Groups: 1981 to 2017



with income is 0.36, whereas it is only 0.06 with religion. In contrast, in wave 7, when the political divide is along principal component 1, the correlation of that principal component with income is 0.06, whereas it is 0.35 with religion.

Second, in the earlier waves the red and blue dots largely overlap, implying that the distributions of Democratic and Republican voters were similar. Gradually, the blue and red dots overlap less, and by 2017 there is a much greater separation between Republicans, who are overwhelmingly conservative along the religious/moral dimension, and Democrats, who are overwhelmingly progressive along that same dimension.

5.3 Rising Partisan Polarization Reflects Latent Polarization

Our findings suggest that the growing political divide in the U.S. is not due to people's underlying values becoming more polarized, as the degree of latent polarization has remained stable over the last four decades. Instead, we argue that the rising political divide reflects an increased alignment between people's political affiliation and their values. We focus on two pieces of evidence. First, we show that political differences between endogenous clusters have increased over time. Second, we document that the mean positions of the voters of the different parties have been converging to the mean positions of the endogenous values-based clusters. As such, political parties are increasingly giving voice to latent values clusters in the U.S.

Partisan and ideological differences between endogenous clusters. Figure 11 helps us evaluate changes over time in the alignment between values clusters on the one hand, and partisan and ideological characteristics on the other hand. It displays differences in the share of Democrats and differences in ideology (on a left-right scale of 1-10) between the two endogenous clusters, for all waves of the WVS, from 1981 to 2017.

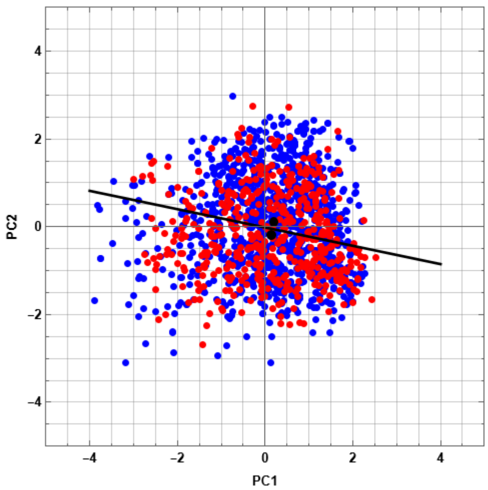
We uncover a notable pattern. The political division between clusters increased sharply over time: the difference in the share of Democrats between the two clusters, which stood at 4.6 percentage points in 1995, grew to 29.0 percentage points by 2017, with a more pronounced increase in the later waves. Thus, the values clusters have become much more politically patterned. That is, partitions derived from values are more predictive of political positions today than they were in the past: there is a growing alignment between political preferences and values clusters.

The difference in political orientation between clusters, on a left-right scale from 1 to 10, has also increased steadily, from 0.63 points in 1981 to 2.15 points in 2017, with the biggest increase occurring in the last decade of the sample. This is consistent with our finding that differences in partisan affiliation between clusters has also increased sharply: as partisan preferences have become more closely aligned with the underlying values clusters in society, the same has occurred with people's political orientations on the left-right scale.[19]

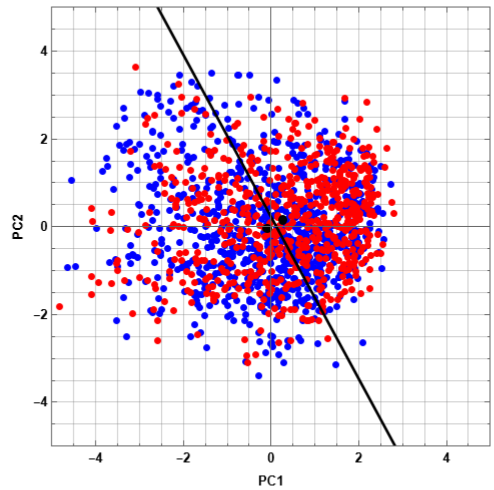
[19] Abramowitz and Saunders (2008) and Fiorina and Abrams (2008) focus on ideology to measure social polarization across time, using a 1-7 scale from the ANES that captures a spectrum from very liberal to very conservative. This could be interpreted as a summary measure of values-based differences between individuals, but our measure of latent polarization relies on a much more comprehensive set of values and our methodology lets the data speak as to which values matter more in defining cultural clusters.

Figure 10: Positions of Voters: 1981 to 2017

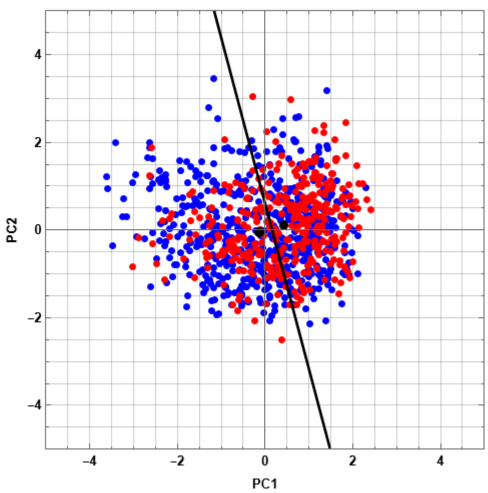
A. Wave 1 (1981)



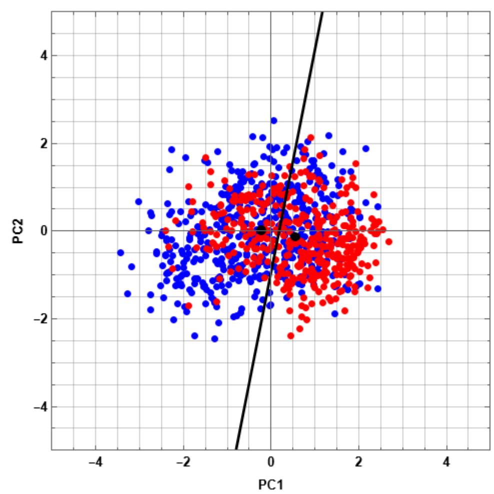
B. Wave 2 (1990)



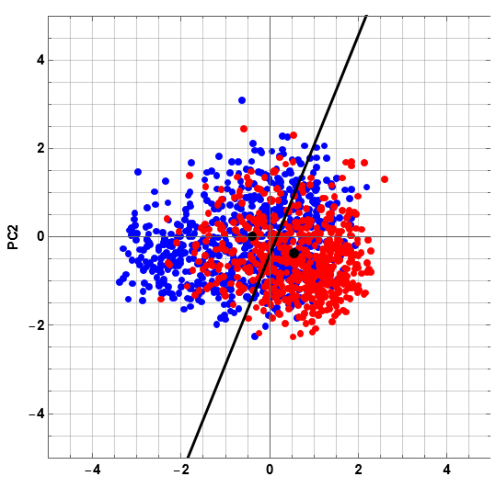
C. Wave 4 (1999)



D. Wave 5 (2006)



E. Wave 6 (2011)



F. Wave 7 (2017)

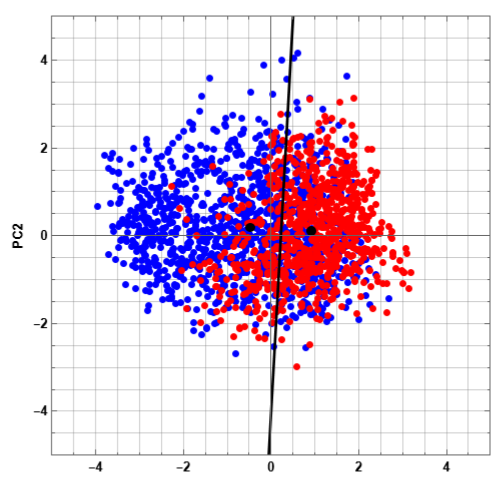
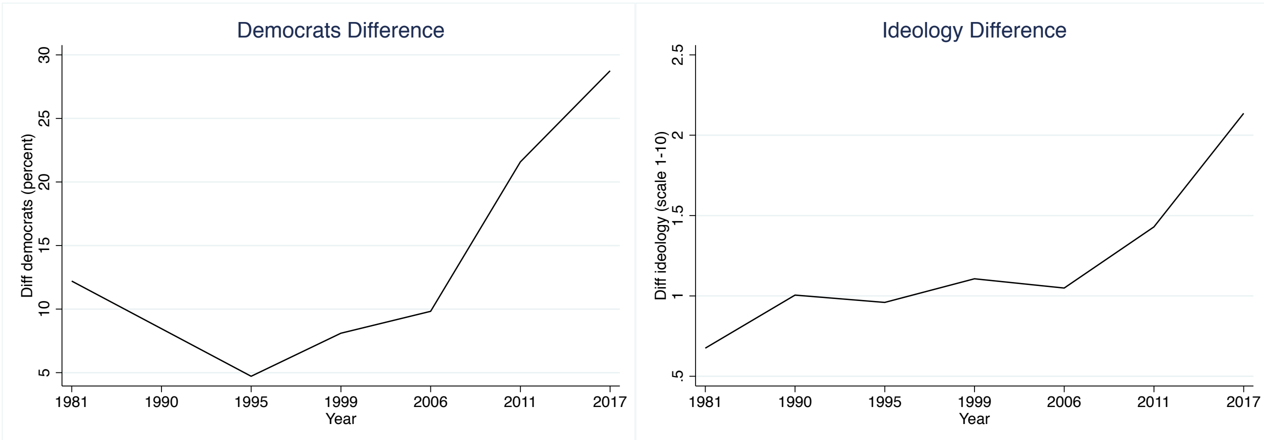


Figure 11: Partisan and Ideological Differences between Clusters: 1981 to 2018



Mean positions of voters are converging with mean positions of clusters. Figure 12 compares the mean positions of the two endogenous clusters in the U.S. along the first two principal components to the mean positions of the individuals that vote for different political parties. It displays these positions for six of the seven waves of the WVS.[20]

In 1981 (wave 1), the mean positions of Democrats and Republicans almost coincide, and they are relatively far removed from the mean positions of clusters 1 and 2. Voters of the two main political parties in the U.S. did not differ much in terms of values on average, and their values were not well aligned with those of the endogenous clusters. Starting in 1990 (wave 2), the mean positions of the two partisan groups start to gradually diverge, with Republican voters moving closer to the mean position of cluster 1 and Democratic voters moving closer to the mean position of cluster 2.

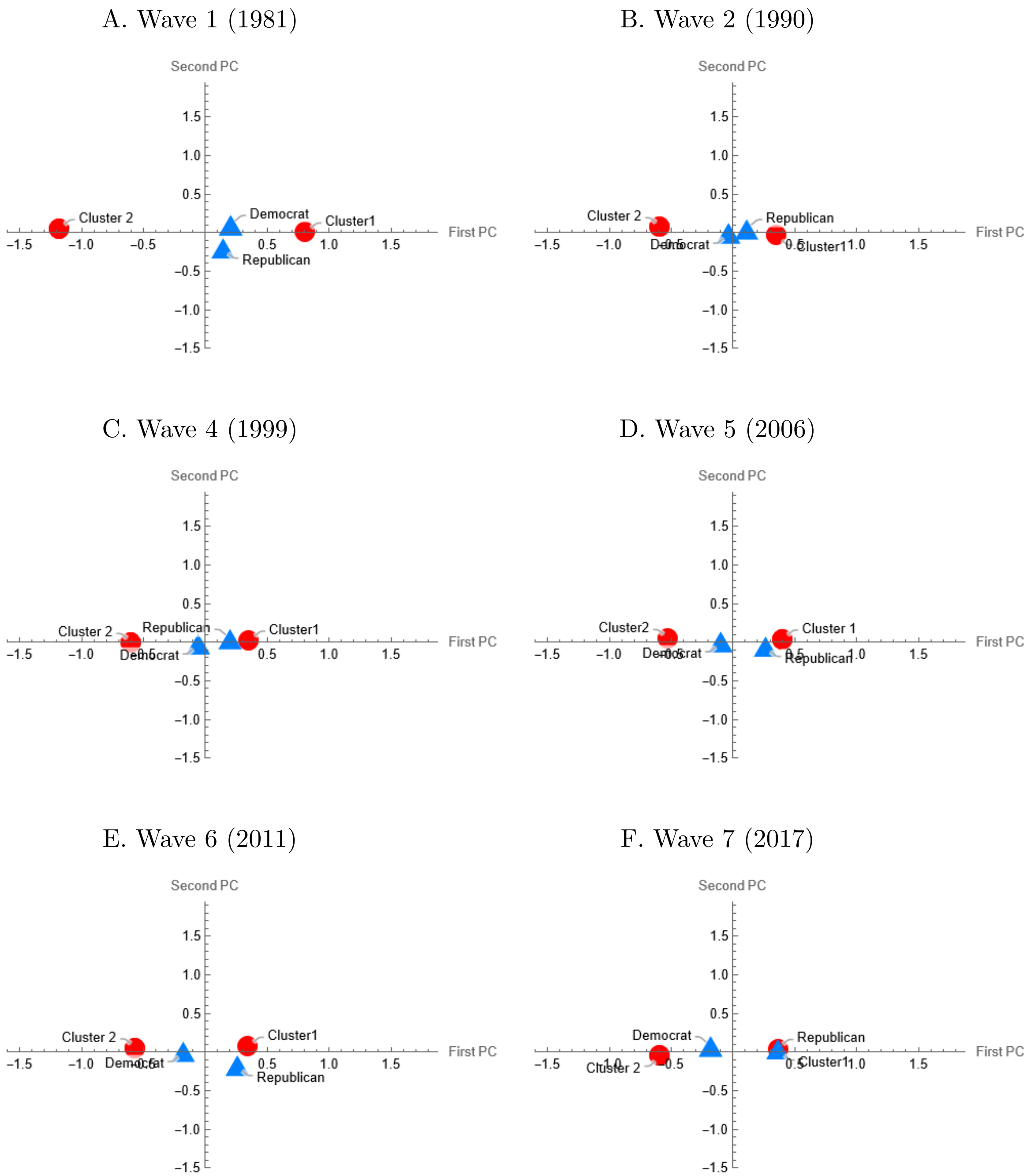
By 2017 (wave 7), the average respondent who aligned with the Republican Party has fully converged to the mean position of cluster 1 - the majority cluster that makes up around two-thirds of the U.S. population. The Democrats have moved closer to cluster 2, but continue to be more centrist than the mean position of the second cluster. This is consistent with the vote share of the two parties being close to 50-50, as cluster 2 only makes up only about one-third of the population: if the mean values of supporters of the Democratic Party had fully moved to those of cluster 2, the Democratic Party would only command a vote share of roughly one-third.[21]

Overall, in the U.S. there has been both a notable divergence between voters of the two main political parties on values, and a notable alignment of partisan positions with the positions of values-based clusters. This is consistent with rising political polarization, in spite of relatively stable latent polarization. This result is reminiscent of Fiorina and Abrams (2008) who argue that

[20]We omit wave 3 (1995) for readability. This wave display patterns that are in between the adjacent waves.

[21]When measuring how the distance between the mean positions of either the clusters or the voters has changed over time, we have not controlled for changes in the average distance between individuals. In Appendix B.2, we use a normalized distance measure that amounts to setting the average distance between two individuals equal to one.

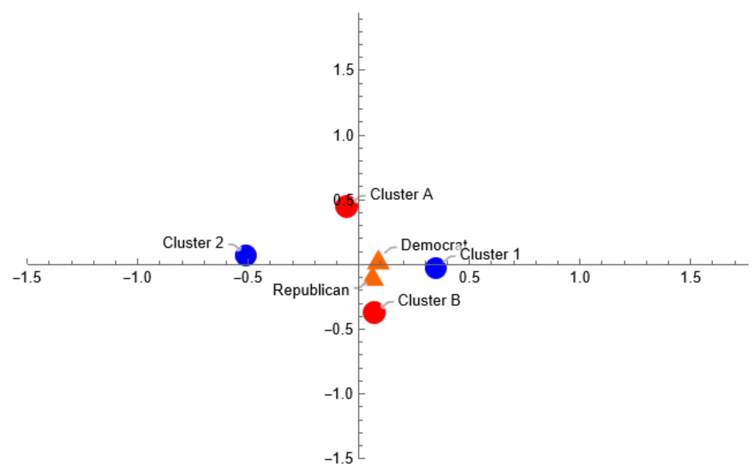
Figure 12: Mean Positions of Voters and Clusters: 1981 to 2017



citizens' positions on public policy issues have not changed much, but that there is increased sorting of voters with certain policy views into certain partisan identifications.

Our findings also speak to the question of how latent values-based clusters can find an actual expression in society. Our premise is that individuals have a natural tendency to associate with like-minded individuals. However, in reality there may not be an easily available technology to express values-based clusters in visible social divisions. In that sense, our findings show that partisan identification is becoming an increasingly efficient way for society to reflect latent values-based clusters. The convergence of partisan positions with the underlying values clusters shows precisely that.

Figure 13: Mean Positions of Voters and Clusters in 1981: Multiple VIEs



Alternative VIEs in early waves. The existence of multiple equilibria in the early waves may provide clues about the reasons behind the growing alignment of partisan positions with latent clusters. Figure 13 depicts the mean positions of the clusters in the Global VIE and in the alternative VIE, together with the mean positions of the Democratic and Republican voters, for wave 1. Given the existence of multiple VIEs, it is *ex ante* not obvious whether we should expect voters to align with clusters along the Global VIE rather than with the alternative equilibrium. What we observe, instead, is mean partisan positions located somewhere in the middle, and very close to each other. In later waves, when the alternative equilibrium disappears, the Global VIE becomes the only equilibrium, and the actual alignment of partisans with underlying values-based clusters becomes more unambiguous. One interpretation of our findings is that the mean positions of voters align with the clusters from the Global VIE in recent periods, due to the lack of an alternative cultural partition.

6 Conclusion

In this paper, we proposed a novel methodology to endogenously partition society into groups based on homophily in values, without considering predefined identity traits. These partitions minimize within-group antagonism, and hence maximize between-group antagonism. As such, the difference in values between these endogenous clusters provides a measure of the maximum attainable polarization in society, a concept we refer to as latent polarization.

We found that values-based partitions reduce antagonism by an order of magnitude more than partitions based on exogenous identity traits, such as gender or ethnicity. If individuals have a preference for associating with other like-minded people, then using identity traits to determine group membership is costly. This is consistent with our past work in Desmet, Ortuño-Ortín and Wacziarg (2017) and Desmet and Wacziarg (2021), showing that identity traits are not very predictive of people's values. The novelty here is that, in spite of high overall heterogeneity in values, there are ways to cluster people into relatively homogeneous groups. This suggests that political parties and other social organizations can create more cohesive coalitions when focusing directly on people's values, or viewpoints.

We also found that over the last four decades, values-based polarization in the U.S. has been rather stable. In addition, the main values dimension along which people disagree the most has not changed since the early 1980s. The culture war that has come to the fore in recent years has been latent for a long time. However, the differences in values between voters of the two main political parties have increased: partisan polarization between Democrats and Republicans was relatively low and stable until the early 2000s, but has since then tripled in magnitude. During this process, the mean positions of the Democrats and the Republicans have become increasingly aligned with the endogenous values-based clusters that we identified. As such, the realignment of political parties in the U.S. and the concurrent rise in political polarization do not stem from an increasingly divided and polarized society, but rather from political affiliations becoming more reflective of latent values-based clusters.

References

Abramowitz, A. I. and K. L. Saunders (2008). Is Polarization a Myth? *Journal of Politics 70*(2), 542–555.

Akerlof, G. and R. E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth 8*(2), 155–194.

Alesina, A., G. Tabellini, and F. Trebbi (2017). Is Europe an Optimal Political Area? *Brookings Papers on Economic Activity 48*(1 (Spring), 169–234.

Ansolabehere, S., J. Rodden, and J. M. Snyder (2006). Purple America. *Journal of Economic Perspectives 20*(2), 97–118.

Baldassarri, D. and A. Gelman (2008). Partisans without Constraint: Political Polarization and Trends in American Public Opinion. *American Journal of Sociology 114*(2), 408–446.

Baron, D. P. (1993). Government Formation and Endogenous Parties. *American Political Science Review 87*(1), 34–47.

Bonomi, G., N. Gennaioli, and G. Tabellini (2021). Identity, Beliefs, and Political Conflict. *Quarterly Journal of Economics 136*(4), 2371–2411.

Boxell, L., M. Gentzkow, and J. M. Shapiro (2022, 01). Cross-Country Trends in Affective Polarization. *The Review of Economics and Statistics*, 1–60.

DellaPosta, D. (2020). Pluralistic Collapse: The "Oil Spill" Model of Mass Opinion Polarization. *American Sociological Review 85*(3), 507–536.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2017). Culture, Ethnicity, and Diversity. *American Economic Review 107*(9), 2479–2513.

Desmet, K. and R. Wacziarg (2021). The Cultural Divide. *The Economic Journal 131*(637), 2058–2088.

Draca, M. and C. Schwarz (2023). How Polarised Are Citizens? Measuring Ideology from the Ground Up. *The Economic Journal forthcoming*.

Enke, B., M. Polborn, and A. A. Wu (2022). Morals as Luxury Goods and Political Polarization. Technical report.

Esteban, J. and D. Ray (1994, July). On the Measurement of Polarization. *Econometrica 62*(4), 819–851.

Fiorina, M. P. and S. J. Abrams (2008). Political Polarization in the American Public. *Annual Review of Political Science 11*(1), 563–588.

Fiorina, M. P., S. J. Abrams, and J. C. Pope (2005). *Culture War? The Myth of a Polarized America*. New York: Pearson Longman.

Gethin, A., C. Martínez-Toledano, and T. Piketty (2022). Brahmin Left Versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948-2020. *Quarterly Journal of Economics 137*(1), 1–48.

Gomberg, A., F. Marhuenda, and I. Ortuño-Ortín (2004). A Model of Endogenous Political Party Platforms. *Economic Theory 24*(2), 373–394.

Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science 22*(1), 129–146.

Jackson, M. (2010). Inequality's Economic and Social Roots: The Role of Social Networks and Homophily. Working paper, Stanford University.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology 27*(1), 415–444.

Ortuño-Ortín, I. and J. E. Roemer (2000). Endogenous Party Formation And The Effect Of Income Distribution On Policy. Working Paper 2000-06, Instituto Valenciano de Investigaciones Economicas.

Shayo, M. (2009). A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review 103*(2), 147–174.

Verbrugge, L. (1977). The Structure of Adult Friendship Choices. *Social Forces 56*(2), 576–597.

A Antagonism and Existing Measures of Heterogeneity

A.1 Antagonism, Cultural Fractionalization and Cultural Differentiation

The values of individuals are in a Q-dimensional space. Thus, an individual j is characterized by the vector of values $v_j = \left\{v_j^1, v_j^2, ..., v_j^Q\right\}$. Antagonism within a given group A_i is defined as the average pairwise distance between individuals in that group:

$$E(A_i) = \frac{1}{J_i}\sum_{j=1}^{J_i} E(A_i, v_j) = \frac{1}{J_i}\sum_{j=1}^{J_i} \frac{\sum_{k=1}^{J_i} d(v_j, v_k)}{J_i} = \frac{1}{J_i^2}\sum_{j=1}^{J_i}\sum_{k=1}^{J_i} d(v_j, v_k) \tag{12}$$

We now compare within-group antagonism (12) with the CF_D index of cultural fractionalization that incorporates distances, as defined in Appendix A.4.1 of Desmet et al. (2017). Start with the one-dimensional case (i.e., $Q = 1$). Applied to our setup, in that case Desmet et al. (2017) define within-group cultural fractionalization that takes account of distances as:

$$CF_D^{A_i} = \sum_{j=1}^{J_i}\sum_{k=1}^{J_i} \frac{1}{J_i}\frac{1}{J_i} d_{jk} = \frac{1}{J_i^2}\sum_{j=1}^{J_i}\sum_{k=1}^{J_i} d_{jk} \tag{13}$$

where d_{jk} is the distance between individuals in one-dimensional space. $CF_D^{A_i}$ is nothing else than Greenberg's B index: the expected distance between the answers given by two randomly picked individuals. In the one-dimensional case, $E(A_i) = CF_D^{A_i}$.

When considering the Q-dimensional case, Desmet et al. (2017) define $CF_D^{A_i}$ as the average cultural fractionalization over the Q dimensions:

$$CF_D^{A_i} = \frac{1}{Q}\frac{1}{J_i^2}\sum_{q=1}^{Q}\sum_{j=1}^{J_i}\sum_{k=1}^{J_i} d_{jk}^q \tag{14}$$

In the case of the squared Euclidean distance, we can compare (12) and (14). With this distance metric, $d(v_j, v_k) = \|v_j - v_k\|^2$, so that (12) can be written as:

$$E(A_i) = \frac{1}{J_i^2}\sum_{j=1}^{J_i}\sum_{k=1}^{J_i} d(v_j, v_k) = \frac{1}{J_i^2}\sum_{j=1}^{J_i}\sum_{k=1}^{J_i}\sum_{q=1}^{Q}\left(v_j^q - v_k^q\right)^2 \tag{15}$$

The index $CF_D^{A_i}$ is now:

$$CF_D^{A_i} = \frac{1}{Q}\frac{1}{J_i^2}\sum_{q=1}^{Q}\sum_{j=1}^{J_i}\sum_{k=1}^{J_i}\left(v_j^q - v_k^q\right)^2 = \frac{1}{Q}E(A_i) \tag{16}$$

Next, consider I groups, $A_1, A_2, ..., A_I$, with the number of individuals in group i given by J_i. The set of all individuals is $S = A_1 \cup A_2 \ldots \cup A_I$. In this case, average within-group antagonism can be written as:

$$E(A^I) = \sum_{i=1}^{I} \frac{J_i}{J} E\left(A_i\right) \tag{17}$$

which can be related to the within-group cultural fractionalization expression in Desmet et al. (2017) is:

$$CF_D^W = \sum_{i=1}^{I} \frac{J_i}{J} CF_D^{A_i} = \frac{E\left(A^I\right)}{Q} \tag{18}$$

Finally, Φ_{ST} as defined in the Appendix in Desmet et al. (2017) coincides with our measure:

$$\Phi_{ST} = \frac{CF_D^S - CF_D^W}{CF_D^S} = \frac{\frac{E(S)}{Q} - \frac{E\left(A^I\right)}{Q}}{\frac{E(S)}{Q}} = \frac{E(S) - E\left(A^I\right)}{E(S)} \tag{19}$$

Given the number of groups I, it is immediate to see that the groups A_1^*, $A_2^*, \ldots, A_I^*$ that minimize within-group antagonism CF_D^W are the ones that maximize Φ_{ST} (i.e. between-group antagonism) – which also happens to be our measure of latent polarization.

A.2 Between-Cluster Variance in the Squared Euclidian Case

We next show the well-known fact that in the case of Squared Euclidean distances:

$$E(A_i) = \frac{1}{J_i^2} \sum_{j=1}^{J_i} \sum_{k=1}^{J_i} \|v_j - v_k\|^2 = \frac{2}{J_i} \sum_{j=1}^{J_i} \|v_j - \mu_i\|^2 \tag{20}$$

where μ_i denotes the vector of mean values of group A_i:

$$\mu_i = \frac{\sum_{j=1}^{J_i} v_j}{J_i} \tag{21}$$

The proof is straightforward:

$$\begin{aligned} E(A_i) &= \frac{1}{J_i^2} \sum_{k=1}^{J_i} \sum_{j=1}^{J_i} \|v_j - v_k\|^2 = \frac{1}{J_i^2} \sum_{k=1}^{J_i} \sum_{j=1}^{J_i} \sum_{q=1}^{Q} \left(v_j^q - v_k^q\right)^2 = \frac{1}{J_i^2} \sum_{k=1}^{J_i} \sum_{j=1}^{J_i} \sum_{q=1}^{Q} \left(v_j^q - \mu_i^q - v_k^q + \mu_i^q\right)^2 \\ &= \frac{1}{J_i^2} \sum_{k=1}^{J_i} \sum_{j=1}^{J_i} \sum_{q=1}^{Q} \left(\left(v_j^q - \mu_i^q\right)^2 + \left(v_k^q - \mu_i^q\right)^2 - 2\left(v_j^q - \mu_i^q\right)\left(v_k^q - \mu_i^q\right) \right) \\ &= \frac{2}{J_i} \sum_{k=1}^{J_i} \sum_{q=1}^{Q} \left(v_k^q - \mu_i^q\right)^2 - 0 = \frac{2}{J_i} \sum_{k=1}^{J_i} \|v_k - \mu_i\|^2 \end{aligned} \tag{22}$$

We can also relate the antagonism in A_i with the variance of the values $v_j = \left\{v_j^1, v_j^2, ..., v_j^Q\right\}$. In the one-dimensional case, the variance of values in group A_i is:

$$Var(A_i) = \frac{1}{J_i}\sum_{j=1}^{J_i}(v_j - \mu_i)^2 \tag{23}$$

Thus, $E(A_i) = 2Var(A_i)$. With I groups, $A_1, A_2, \ldots, A_I$ we have:

$$E(A^I) = \sum_{i=1}^{I}\frac{J_i}{J}E(A_i) = 2\sum_{i=1}^{I}\frac{J_i}{J}Var(A_i) \tag{24}$$

Thus, average within-group values antagonism is closely related to the average within-group variance when the distance metric is squared Euclidian. Another way to state this result is that the groups A_1^*, $A_2^*, \ldots, A_I^*$ that minimize average within-group antagonism are the ones that minimize average within-group variance. Given that the total variance is constant, this is equivalent to the partition that maximizes average between-group variance.

In the multidimensional case ($Q > 1$), the interpretation is not exactly the same. We have:

$$E(A_i) = \frac{2}{J_i}\sum_{q=1}^{Q}\sum_{j=1}^{J_i}\left(v_j^q - \mu_i^q\right)^2 = 2\sum_{q=1}^{Q}Var(v_i^q) \tag{25}$$

where $Var(v_i^q)$ stands for the (sample) variance in dimension q. Obviously, $\sum_{q=1}^{Q}Var(v_i^q)$ is not the variance of the random vector v_i, since it ignores covariances. However, if we apply our method using principal components of the value questions, the Q dimensions are by construction uncorrelated, so that $Cov\left(v^q, v^{q'}\right) = 0$. In this case , the results obtained in the one-dimensional case apply for $Q > 1$ as well, and minimizing within-group antagonism is equivalent to maximizing between-group variance.

A.3 Partitions under a Distance to the Mean Criterion

Here, we show that (i) an MVIE is not, in general, a VIE and (ii) when group sizes are sufficiently large, then an MVIE is also a VIE. Point (ii) consists of the Proof of Proposition 1.

If a partition $A = (A_1, A_2)$ is an MVIE, it solves the problem:

$$\underset{A}{\operatorname{argmin}}\sum_{i=1}^{2}\sum_{j\in A_i}\|v_j - \mu_i\|^2$$

As we showed in Appendix A.2, this is equivalent to solving the problem:

$$\underset{A}{\operatorname{argmin}}\sum_{i=1}^{2}\frac{1}{J_i}\sum_{j\in A_i}\sum_{k\in A_i}\|v_j - v_k\|^2$$

We first show that partition A need not be a VIE. Consider an alternative partition $A' = \{A'_1, A'_2\}$ that differs from A in that we move one individual c from A_1 to A_2. Since partition A is an MVIE, we have

$$\sum_{i=1}^{2} \frac{1}{J_i} \sum_{j \in A_i} \sum_{k \in A_i} \|v_k - v_j\|^2 \leq \frac{1}{J_1 - 1} \sum_{j \in A'_1} \sum_{k \in A'_1} \|v_k - v_j\|^2 + \frac{1}{J_2 + 1} \sum_{j \in A'_2} \sum_{k \in A'_2} \|v_k - v_j\|^2$$

which can be rewritten as

$$\begin{aligned} &\frac{2}{J_2 + 1} \sum_{j \in A'_2} \|v_c - v_j\|^2 - \frac{2}{J_1} \sum_{j \in A_1} \|v_c - v_j\|^2 \\ &\quad \geq \left(\frac{1}{J_1} - \frac{1}{J_1 - 1}\right) \sum_{j \in A_1 \backslash \{c\}} \sum_{k \in A_1 \backslash \{c\}} \|v_k - v_j\|^2 + \left(\frac{1}{J_2} - \frac{1}{J_2 + 1}\right) \sum_{j \in A_2} \sum_{k \in A_2} \|v_k - v_j\|^2 \end{aligned} \tag{26}$$

The left-hand side of inequality (26) is the change in antagonism experienced by agent v_c, i.e., $E(A'_2, v_c) - E(A_1, v_c)$. If the partition A is a VIE, then $E(A'_2, v_c) - E(A_1, v_c) \geq 0$. However, the right-hand side of equation (26) need not be positive. Hence, $E(A'_2, v_c)$ is not necessarily larger than $E(A_1, v_c)$, so that the partition A need not be a VIE.

Proof of Proposition 1.

Proposition 1 states that if group sizes are large enough, any MVIE is also a VIE. Note that by increasing the size of J_1 and J_2 we can make the right-hand side of (26) as close as we want to zero. Hence, we can ensure that the left-hand side of (26) is positive. In other words, starting from an MVIE, any agent switching to the other group would experience an increase in antagonism. This implies that an MVIE is also a VIE. □

B Additional Results

B.1 Socio-demographic partitions into more than two groups

In this subsection, we discuss polarization based on socio-demographic partitions into more than two groups. As a reference point, Panel A of Table B1 reports latent polarization between two clusters, for the U.S. and the average of 81 countries. Panel B reports polarization based on three types of demographic partitions (as well as on partisan partitions). In contrast to Table 1, for each of these partitions, we consider all the groups that appear in the WVS. For example, in the case of the U.S. ethnicity we do not simply distinguish between whites and non-whites but between five groups: 'white', 'Black', 'Hispanic', 'other', and 'two or more'. For the other demographic traits, income consists of ten deciles, and religion involves ten groups ('no denomination', 'Roman Catholic', 'Protestant', 'Orthodox', 'Jew', 'Muslim', 'Hindu', 'Buddhist', 'Other Christian', 'Other'). For partisan affiliation, the options are 'Republican Party', 'Democratic Party', 'Libertarian', and 'Green Party'. We then consider how divided these groups are along the first two principal components.

Table B1: Latent Polarization and Polarization between Identity Groups (WVS Wave 7)

	U.S.	81 Countries	
		Mean	s.d.
A. Two Clusters, Two PCs			
Latent Polarization	41.2	41.7	6.1
B. Polarization between Other Identity Groups (All Groups)			
Religion	13.4	13.4	10.5
Income	4.8	4.4	3.2
Ethnicity	1.5	3.4	4.7
Partisan	14.7	9.0	6.6

Notes: All polarization measures are between all groups in the WVS: for the case of the U.S., religion ('no denomination', 'Roman Catholic', 'Protestant', 'Orthodox', 'Jew', 'Muslim', 'Hindu', 'Buddhist', 'Other Christian', 'Other'), income (deciles), ethnicity ('white', 'Black', 'Hispanic', 'other', 'two or more'), and partisan ('Republican Party', 'Democratic Party', 'Libertarian', 'Green Party').

In general, polarization between demographic or partisan groups is low. For example, polarization is only 1.5 if partitions are based on ethnicity, and 4.8 if based on income deciles. Polarization is somewhat larger when groups are defined by different religions (13.4) or different political parties (14.7), though these figures are still much lower than latent polarization (41.2). In the rest of the world, polarization levels between groups based on income, ethnicity, religion, and partisan groups are similarly low. For example, when considering 81 countries, the average degree of polarization between income groups is 4.4, compared to 4.8 in the U.S.

Because our measure of polarization is increasing in the number of groups, the levels are slightly higher than in 1, where each demographic traits is reduced to only two groups. For example, polarization between groups based on income deciles is 4.4, compared to 3.3 when the only two income groups are 'above' and 'below the median'. Similarly, polarization between religious groups is 13.4 when we distinguish all groups, compared to 10.0 when the only two groups are 'religious denomination' and 'no religious denomination'. However, the qualitative result is unchanged: polarization between groups defined by demographic or partisan groups is low, compared to latent polarization.

B.2 Partisan Alignment with Values Based Clusters: Normalized Measure

When measuring how the distance between the mean positions of either the clusters or the voters has changed over time, we have not controlled for changes in the average distance between individuals. To see why this might be an issue, consider the following hypothetical example. Suppose the average distance between Republicans and Democrats has been increasing at the same rate as the average distance between individuals. Based on our baseline measure we would say that the political divide is growing, but it is debatable whether we would actually want to conclude that Republican and

Democratic voters are drifting further apart.

To address this possible concern, we compute a normalized distance measure between groups that makes the average distance between individuals equal to one. More specifically, denote the positions of an individual j along principal components 1 and 2 by v_j^1 and v_j^2. In the squared Euclidean case, denote the average non-normalized distance between any two individuals by $\bar{d} = \frac{1}{J}\sum_{j=1}^{J}\sum_{q=1}^{2}(v_j^q - \mu^q)^2$, where $\mu^q = \frac{1}{J}\sum_{j=1}^{J} v_j^q$. Now divide the position of an individual j along principal component q by $\sqrt{d}$, and denote this normalized position by $v_j^{q\prime} = v_j^q/\sqrt{d}$. When using these normalized positions, the average normalized distance between two individuals is one. Figure B1 shows the results when using this normalized measure. The results are unchanged: the positions of the two main political parties are diverging and becoming increasingly aligned with the positions of the endogenous values-based clusters.

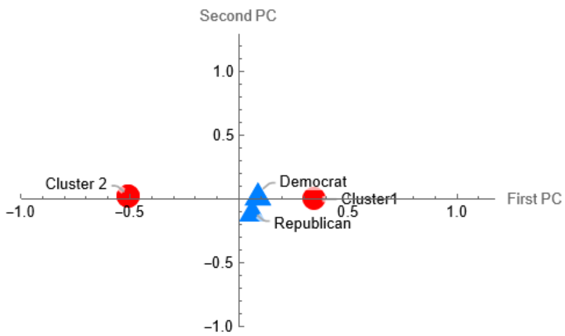
C Additional Tables

Table C1: Questions and Respondents

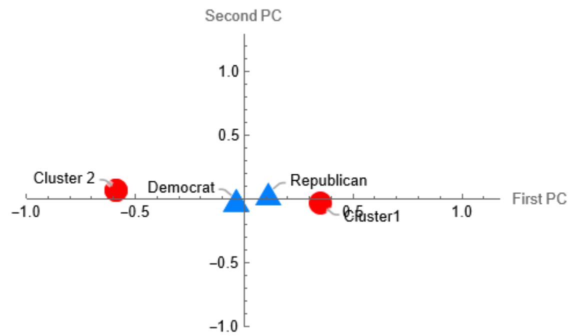
	# Questions		# Respondents		% Imputed
	Original	Used	Original	Used	
WVS 1	176	173	2325	2322	4.3
WVS 2	292	276	1839	1836	5.5
WVS 3	161	161	1542	1536	3.1
WVS 4	202	202	1200	1200	1.4
WVS 5	173	170	1249	1217	3.3
WVS 6	170	166	2232	2198	3.0
WVS 7	198	198	2596	2578	1.1

Figure B1: Mean Positions of Voters and Clusters: Normalized Distances

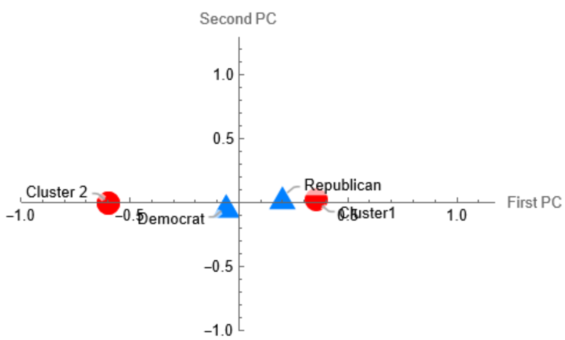
A. Wave 1 (1981)



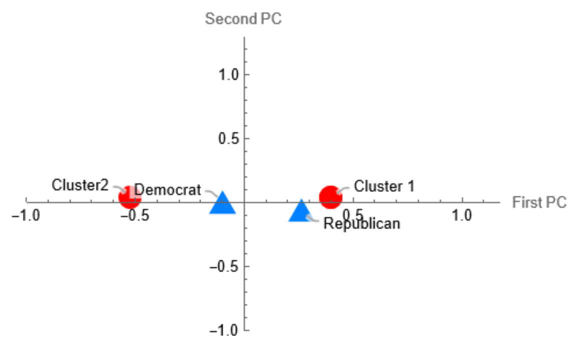
B. Wave 2 (1990)



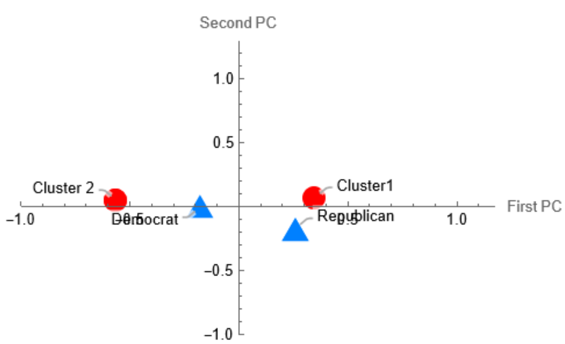
C. Wave 4 (1999)



D. Wave 5 (2006)



E. Wave 6 (2011)



F. Wave 7 (2017)