Contrastive Multiple Correspondence Analysis (cMCA): Applying the Contrastive Learning Method to Identify Political Subgroups

Takanori Fujiwara[*†] and Tzu-Ping Liu[*‡]

University of California, Davis

July 20, 2020

Abstract

Ideal point estimation and dimensionality reduction have long been utilized to simplify and cluster complex, high-dimensional political data (e.g., roll-call votes, surveys, and text) for use in (preliminary) analysis and visualization. These methods often work by finding the directions or principal components (PCs) on which either the data varies the most or respondents make the fewest decision errors. However, these PCs, which usually reflect the left-right political spectrum (Coombs 1964), are sometimes uninformative in explaining significant differences in the distribution of the data (e.g., how to categorize a set of highly-moderate voters). To tackle this prevalent issue, we adopt an emerging analysis approach, called *contrastive learning*. Contrastive learning—e.g., contrastive principal component analysis (cPCA) (Abid et al. 2018)—works by first splitting the data by predefined groups, and then deriving PCs on which the target group varies the most but the background group varies the least. As a result, cPCA can often find 'hidden' patterns, such as subgroups within the target group, which PCA cannot reveal when some variables are the dominant source of variations across the groups. We contribute to the field of contrastive learning by extending it to multiple correspondence analysis (MCA) in order to enable an analysis of data often encountered by social scientists—namely binary, ordinal, and nominal variables. We demonstrate the utility of contrastive MCA (cMCA) by analyzing three different surveys: The 2015 Cooperative Congressional Election Study, 2012 UTokyo-Asahi Elite Survey, and 2018 European Social Survey. Our results suggest that, first, for the cases when ordinary MCA depicts differences between groups, cMCA can further identify characteristics that divide the target group; second, for the cases when MCA does not show clear differences, cMCA can successfully identify meaningful directions and subgroups, which traditional methods overlook.

---

[*] Equal contribution and listed alphabetically.

[†] Takanori Fujiwara is Ph.D. Candidate, Department of Computer Science, One Shields Avenue, Davis, CA 95616 (tfujiwara@ucdavis.edu).

[‡] Tzu-Ping Liu is Ph.D. Candidate, Department of Political Science, One Shields Avenue, Davis, CA 95616 (tpliu@ucdavis.edu).

Introduction

Ideal point estimation is a prolific and popular topic for research throughout political science. Scholars have developed and utilized a diverse set of methods to identify political actors' "positions" on the left-right ideological space. This has been accomplished by utilizing one of two general ideas: Either "optimal scoring" (Jacoby 1999), these methods include principal component analysis (PCA), multiple correspondence analysis (MCA), and factor analysis (e.g., Brady 1990; Davis, Dowley, and Silver 1999). Or "spatial voting," these methods include Aldrich-McKelvey scaling, the NOMINATE class of models, two-variable item response theory (IRT), and optimal classification (OC) (e.g., Aldrich and McKelvey 1977; Clinton, Jackman, and Rivers 2004; Hix, Noury, and Roland 2006; Martin and Quinn 2002; Poole 2000; Poole and Rosenthal 1991). While these methods have traditionally been applied only to roll-call votes, they have been increasingly applied to range of new data including surveys and text (Hare, Liu, and Lupton 2018; Imai, Lo, and Olmsted 2016; Monroe, Colaresi, and Quinn 2008; Poole 1998; Wilkerson and Casas 2017).

In general, these methods are used to uncover the similarities (differences) between observations in a given dataset. These similarities are then used to estimate a small number of (underlying) dimensions that reflect the similarities as a set of spatial distances between all of the estimated points. These lower-dimensional representations can both help us understand the general structure of our data, such as its distribution, and identify underlying patterns, like clusters and outliers (Fujiwara, Kwon, and Ma 2020; Wenskovitch et al. 2018). More precisely, these methods are often explicitly used to (1) identify clusters; (2) derive influential factors (dimensions); and (3) compare the derived clusters and/or factors with predefined classes or groups (e.g., Armstrong et al. 2014; Brady 1990; Davis, Dowley, and Silver 1999; Grimmer and King 2011; Hare et al. 2015; McCarty, Poole, and Rosenthal 2001; Rosenthal and Voeten 2004). Consequently, ideal point estimation can be considered both a dimensionality reduction and clustering technique.

In addition to deriving lower-dimensional spaces, correctly interpreting these recovered dimensions is just as, if not more, important. Since the work of Coombs (1964), the latent left-right dimension of ideology is often recovered as a principal component/direction (PC) in political data. Indeed, almost all of the methods discussed previously recover ideology as their first PC. Nevertheless, these methods can face situations when the left-right scale is not informative, i.e., when respondents are highly moderate and clustered in the middle of the scale. Because of this clustering, it is difficult to delineate between voters of some given group or category, such as partisans. While this does not happen particularly often, cases of strong moderation are found throughout political science (see the UTAS and ESS analyses below) and these traditional methods will often fail to meaningfully distinguish between groups (parties).

Although the derived results may not be informative when respondents are highly moderate, it does not *necessarily* mean that supporters from different parties are politically similar. Rather, supporters from different parties may be divided by certain issues or (synthetic) directions that are not captured by the left-right scale. This is because these methods identify dimensions on which the data as a whole varies maximally or respondents make the minimum amount of decision errors.[1] Thus, to deal with this common issue of uninformative PC(s), for instance, instead of analyzing data as a whole, Abid et al. (2018) derive a contrastive learning procedure of PCA, named contrastive PCA (cPCA), which first splits data into different groups, usually by predefined classes (e.g., party ID), and then compares the data structure of the target group and the background group to find PC(s) on which the target group varies maximally and the background group varies minimally.

We contribute to the field of contrastive learning by extending this idea from PCA to multiple correspondence analysis (MCA) so as to enable the derivation of contrastive PC(s) using different types of features/data that social scientists encounter frequently: binary, ordinal, and nominal data. In applying contrastive MCA (cMCA), we find that even when

1. The idea of maximal variance and the idea of minimal decision errors are equivalent. This is because while the projected data points can vary the most on a PC, it indicates that middle point of that PC provides the maximum distance between point clouds on both sides.

voters are highly moderate and homogeneous across party lines, we can still explore how specific issue(s) effectively identify subgroups within the parties themselves. Similar to cPCA, cMCA first splits the data by predefined groups, such as partisanship, treatment/control groups, etc; second, as is the case with MCA, cMCA applies a one-hot encoder to convert a categorical dataset into a binary format matrix, called a disjunctive matrix; finally, cMCA takes one of the groups as the background/pivot and then compares this group with the target group to derive PCs on which the positions of members' ideal points from target groups vary the most but those from the background group vary the least.

To demonstrate the utility of cMCA, we analyze three surveys, the 2015 Cooperative Congressional Election Study (CCES 2015), the 2012 Asahi-Todai Elite Survey (UTAS 2012), and the 2018 European Social Survey (ESS 2018, the U.K. module only). The results show that first, even under conditions where voters are highly polarized and traditional PC(s) can divide voters along party lines, as demonstrated by the results from the 2015 CCES, cMCA enables researchers to better capture the unique distributions among Democratic and Republican supporters, respectively. Second, when voters are highly moderate and traditional PC(s) are uninformative, as shown by the results from the 2012 UTAS and 2018 ESS, cMCA detects subgroups along with contrastive directions, such as the anti-Korean and Chinese mood among supporters of the Liberal Democratic Party (LDP) in Japan or the pro- and anti-EU attitudes among supporters from the U.K's Conservative and Labour Party, separately.

This work makes an important contribution to dimensional analysis by enabling researchers to identify dimensions that effectively classify and divide observations (voters) even when the left-right scale cannot effectively discriminate between classes/groups (partisanship). This method, as evidenced by its application to voter surveys, meaningfully identifies dimensions that would be, and are, overlooked by existing methods. Simply put, cMCA's results provide important information for researchers to better understand and make meaningful distinctions between underlying groups in their data. Moreover, compared with other ideal point estimation methods such as the ordered optimal classification (Hare, Liu, and Lupton 2018), the basic-space procedure (Poole 1998), and the Bayesian mixed factor analysis model (Martin, Quinn, and Park 2011; Quinn 2004), cMCA is exceedingly fast and efficient in computation since cMCA utilizes MCA as the base of the algorithm. Finally, by revealing the overlooked influential factors, cMCA provides a springboard for further analyses of high-dimensional data, such as feature selection. Overall, we believe that this method provides an important contribution to data visualization and unsupervised learning. The paper proceeds with a description of cMCA, its application to our three voter surveys, and then concludes with general thoughts about and rules of thumb for using cMCA.

Contrastive Learning and Contrastive MCA

Contrastive learning (CL) is an emerging machine learning approach, which analyzes high-dimensional data to capture "patterns that are specific to, or enriched in, one dataset relative to another" (Abid et al. 2018). Unlike ordinary approaches, such as PCA and MCA, that usually aim to capture the characteristics of one entire dataset, CL compares subsets of that dataset relative to one another (target versus background group). The logic behind CL is to explore unique components or directions that contain more salient latent patterns in one dataset (the target group) than the other (background group). So far, this approach has been applied to several machine learning methods, including PCA (Abid et al. 2018), latent Dirichlet allocation (Zou et al. 2013), hidden Markov models (Zou et al. 2013), regressions (Ge and Zou 2016), and variational autoencoders (Abid and Zou 2019; Severson, Ghosh, and Ng 2019). We utilize this contrastive learning approach by applying it to MCA, an enhanced version of PCA for nominal and ordinal data analysis (Greenacre 2017; Le Roux and Rouanet 2010).

Multiple Correspondence Analysis (MCA)

When applying PCA to non-continuous data, it handles each level or category in the data as a real numerical value. As a result, PCA unnecessarily ranks the categories and assumes that the intervals are all equal and represent the differences between categories. To overcome these issues, MCA first converts an input dataset $\mathbf{X} \in \mathbb{R}^{p\times d}$ (p: the number of data points, d: the number of variables) of non-continuous data into what is called a disjunctive matrix $\mathbf{G} \in \mathbb{R}^{p\times K}$ (K: the total number of categories) by applying one-hot encoding to each of d categorical dimensions (Harris and Harris 2010). For illustration, assume that $\mathbf{X}$ consists of two columns/variables: `color` and `shape`, and each variable has three (`red`, `green`, `blue`) and two (`circle`, `rectangle`) levels, respectively. In this case, the disjunctive matrix $\mathbf{G}$ will contain five categories: `red`, `green`, `blue`, `circle`, and `rectangle`. For instance, a piece of blue rectangle paper will be recorded as $[0, 0, 1, 0, 1]$ in $\mathbf{G}$.

We could further derive a probability/correspondence matrix $\mathbf{Z}$ through dividing each value of $\mathbf{G}$ by the grand total of $\mathbf{G}$ (i.e., $\mathbf{Z} = N^{-1}\mathbf{G}$ where N is the grand total of $\mathbf{G}$). This probability matrix can now be treated as a typical dataset for use in analyses. Similar to PCA, we first normalize $\mathbf{Z}$ and then obtain what is called a Burt matrix, $\mathbf{B}$, with $\mathbf{B} \stackrel{\text{def}}{=} \mathbf{Z}^\top\mathbf{Z}$. This Burt matrix $\mathbf{B}$ under MCA corresponds to a covariance matrix under PCA. Thus, as in PCA, to derive principal directions, MCA performs eigenvalue decomposition (EVD) to $\mathbf{B}$ to preserve the variance of $\mathbf{G}$.[2]

Extending MCA to cMCA

To apply CL to MCA, we first split the original dataset $\mathbf{X} \in \mathbb{R}^{p\times d}$ into a target dataset $\mathbf{X}_T \in \mathbb{R}^{n\times d}$ and a background dataset $\mathbf{X}_B \in \mathbb{R}^{m\times d}$ (where $p = n+m$) by a certain predefined boundary, such as individuals' partisanship or whether they were assigned to treatment or control groups. We then further derive two Burt matrices from the target and background datasets, $\mathbf{B}_T$ and $\mathbf{B}_B$, respectively. Importantly, because the CL procedure utilizes matrix

2. Similar to PCA, one could apply singular value decomposition (SVD) to $\mathbf{Z}$ to obtain the same principal directions.

differences, the Burt matrices must have the same dimensions, i.e., $\mathbf{B}_T \in \mathbb{R}^{K\times K}$ and $\mathbf{B}_B \in \mathbb{R}^{K\times K}$.

Let $\mathbf{u}$ be any unit vector with K dimensions. Similar to cPCA (Abid et al. 2018), we can derive the variances of the target and background datasets along with $\mathbf{u}$ by using $\sigma_T^2(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{u}^\top \mathbf{B}_T \mathbf{u}$ and $\sigma_B^2(\mathbf{u}) \stackrel{\text{def}}{=} \mathbf{u}^\top \mathbf{B}_B \mathbf{u}$. To obtain the optimal direction $\mathbf{u}^*$ on which $\mathbf{Z}_T$ has the highest variance and $\mathbf{Z}_B$ has the lowest, one needs to solve:

$$\mathbf{u}^* = \arg\max_{\mathbf{u}} \ \sigma_T^2(\mathbf{u}) - \alpha\sigma_B^2(\mathbf{u}) = \arg\max_{\mathbf{u}} \ \mathbf{u}^\top(\mathbf{B}_T - \alpha\mathbf{B}_B)\mathbf{u} \tag{1}$$

where α ($0 \leq \alpha \leq \infty$) is a hyperparameter of cMCA, named a contrast parameter. From Eq. 1, we can see that $\mathbf{u}^*$ corresponds to the first eigenvector of the matrix $(\mathbf{B}_T - \alpha\mathbf{B}_B)$. Note that all of the eigenvectors of $(\mathbf{B}_T - \alpha\mathbf{B}_B)$ can be derived through performing EVD over $(\mathbf{B}_T - \alpha\mathbf{B}_B)$. We call these derived eigenvectors as contrastive principal components (cPCs). The contrast parameter α controls the trade-off between having high target variance versus low background variance. When $\alpha = 0$, the resulting cPCs only maximize the variance of a target dataset producing results that are equivalent to applying standard MCA to only the target dataset. As α increases, cPCs place greater emphasis on directions that reduce the variance of a background dataset. In addition to the manual selection of α, we can also utilize (semi-)automatic selection methods developed for other CL methods (Abid et al. 2018; Fujiwara et al. 2020).

Data-Point Coordinates, Category Coordinates, and Category Loadings

As in ordinary MCA, in cMCA, we provide three essential tools for helping researchers relate data points and categories/variables in the contrastive PC space: (1) data-point coordinates (also known as coordinates of rows (Greenacre 2017), or clouds of individuals (Le Roux and Rouanet 2010)), (2) category coordinates (also known as coordinates of columns (Greenacre 2017), or clouds of categories (Le Roux and Rouanet 2010)), and (3) category loadings. These three tools provide different auxiliary information.

Data-point coordinates provide a lower-dimensional representation of the data, which is standard in most dimensional reduction techniques. Furthermore, category coordinates and loadings provide essential information on how to interpret these data-point coordinates. Similar to data-point coordinates, category coordinates present the position of each category/level in a lower-dimensional space. By comparing category coordinates with data-point coordinates, which both lie on the same contrastive latent space, we can better understand the associations and relationships between the data points and categories. When a data point has a close proximity to a category's location, it indicates that this point and the corresponding category are highly associated with each other (Pagès 2014). On the other hand, category loadings indicate how strongly each category relates to each derived contrastive PC.[3]

Similar to MCA, for a target dataset $\mathbf{X}_T$, cMCA's data-point coordinates $\mathbf{Y}_T^{\text{row}} \in \mathbb{R}^{n\times K'}$ ($K' < K$) can be derived as:

$$\mathbf{Y}_T^{\text{row}} = \mathbf{Z}_T\mathbf{W}_T \tag{2}$$

where $\mathbf{W}_T \in \mathbb{R}^{K\times K'}$ is the top-K' eigenvectors obtained by EVD. Similarly, we can obtain data-point coordinates, $\mathbf{Y}_B^{\text{row}}$, of a background dataset $\mathbf{X}_B$ onto the same low-dimensional space of $\mathbf{Y}_T^{\text{row}}$ through:

$$\mathbf{Y}_B^{\text{row}} = \mathbf{Z}_B\mathbf{W}_T \tag{3}$$

As discussed in Fujiwara et al. (2020), visualizing low-dimensional representations of both target and background datasets can help researchers determine whether or not a target dataset has certain specific patterns relative to a background dataset. When the scatteredness/shape of plotted data points of $\mathbf{Y}_T^{\text{row}}$ is larger than $\mathbf{Y}_B^{\text{row}}$, we can conclude that there exists a set of unique patterns within $\mathbf{Y}_T^{\text{row}}$.

In MCA, the common way of deriving a dataset $\mathbf{X}$'s category coordinates, $\mathbf{Y}^{\text{col}}$, is through $\mathbf{Y}^{\text{col}} = \mathbf{D}\mathbf{W}$, where $\mathbf{D}$ is a diagonal matrix with the sum of each column of $\mathbf{Z}$ on the

3. We provide a detailed discussion of the category selection procedure in Appendix A.

diagonal and $\mathbf{W}$ is a matrix whose column vectors are the top-K' eigenvectors obtained with EVD on $\mathbf{B}$. However, because cMCA applies EVD to the matrix, $(\mathbf{B}_T - \alpha\mathbf{B}_B)$, the result is necessarily affected by $\mathbf{X}_B$ as well. As a consequence, we cannot simply compute category coordinates using the methods described above. To overcome this issue, instead, we utilize MCA's *translation formula* derived from data-point coordinates to calculate category coordinates (Greenacre 2017). More precisely, the category coordinates of the target dataset, $\mathbf{Y}_T^{\text{col}} \in \mathbb{R}^{K \times K'}$, can be derived through:

$$\mathbf{Y}_T^{\text{col}} = \mathbf{D}_T^{-1}\mathbf{Z}_T^{\top}\mathbf{Y}_T^{\text{row}}\text{diag}(\boldsymbol{\lambda})^{-1/2} \tag{4}$$

where $\mathbf{D}_T \in \mathbb{R}^{K \times K}$ is a diagonal matrix with the sum of each column of $\mathbf{Z}_T$ on the diagonal and $\boldsymbol{\lambda} \in \mathbb{R}^{K'}$ is a vector of the top-K' eigenvalues derived from Eq. 1.

Finally, since MCA's derivation procedure is highly similar to that of PCA, we utilize the concept of principal component loadings in PCA to calculate category loadings in cMCA. More precisely, category loadings, $\mathbf{L}_T \in \mathbb{R}^{K \times K'}$, under cMCA can be derived through:

$$\mathbf{L}_T = \mathbf{W}_T \, \text{diag}(\boldsymbol{\lambda})^{1/2} \tag{5}$$

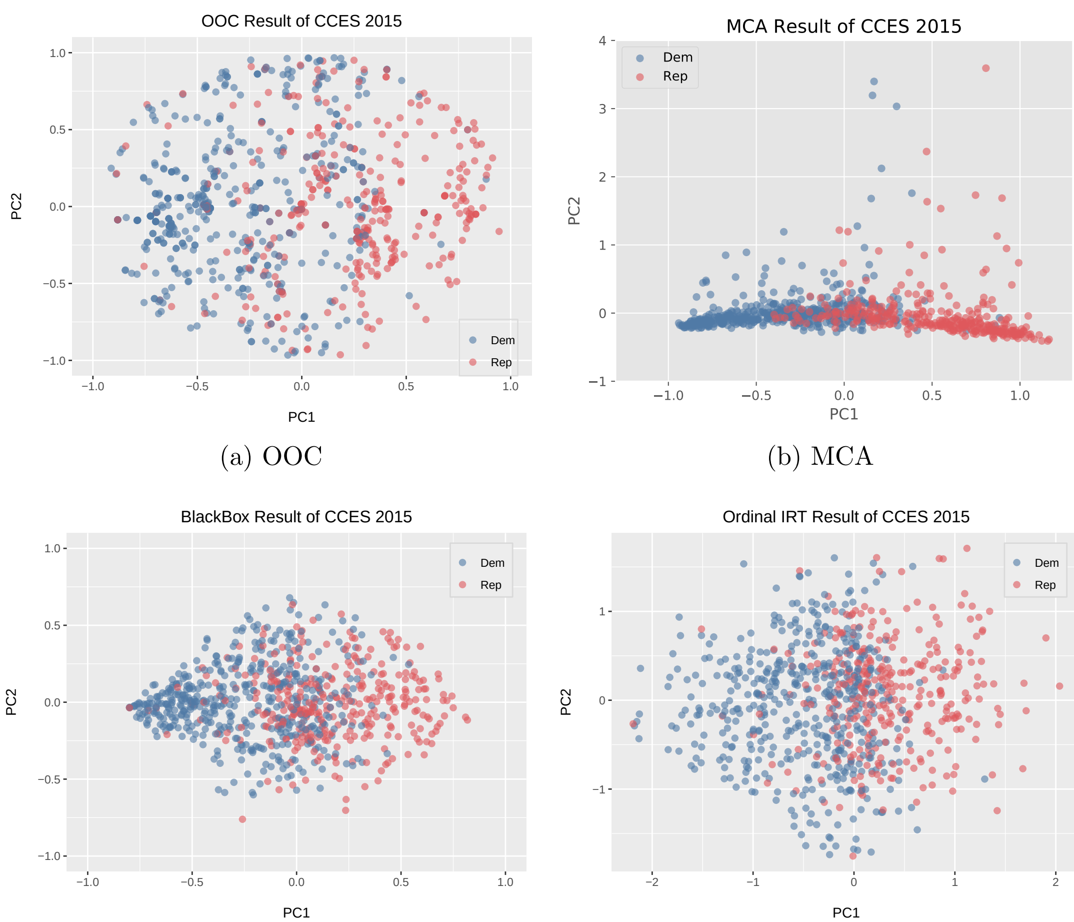
Application

To demonstrate the utility of cMCA, we analyze three surveys—the CCES 2015, the UTAS 2012, and the ESS 2018 (the U.K. module).[4] Given that the home countries of these surveys, the U.S., Japan, and the U.K., have varying levels of political polarization, these surveys provide a wide-range of political situations for testing. Furthermore, given that the general political systems and cultures of each country are unique, we also test the consistency of the derived cMCA results with the observed, qualitative political realities in each country.

4. Please refer to online Appendix B for the variable (re)coding scheme used with each survey.

Case One: CCES 2015—Democrats and Republicans Are Divided by Different Issues

The first survey we examine is the 2015 Cooperative Congressional Election Study (Ansolabehere and Schaffner 2017). We first present the estimated results of all respondents through the ordered optimal classification (OOC) (Hare, Liu, and Lupton 2018), MCA, the Basic-Space (BS) procedure (Poole 1998), and the Bayesian mixed factor analysis model (BMFA) (Martin, Quinn, and Park 2011; Quinn 2004) separately in Fig. 1.[5]



(a) OOC (b) MCA

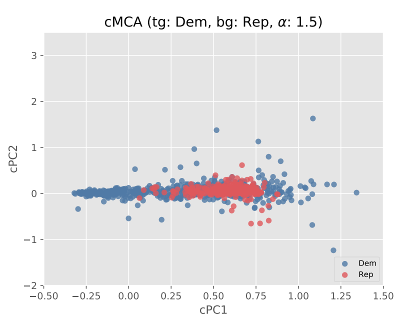
(c) BS (d) BMFA

Figure 1: OOC, MCA, BS, and BMFA Results of CCES 2015

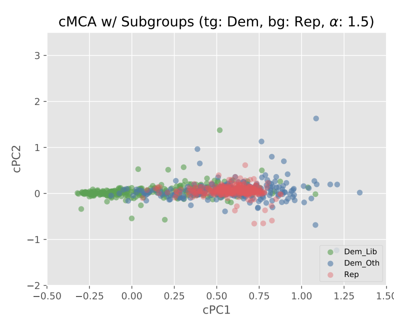
Although the four approaches utilize different methods for dimensional reduction and

5. Given that the purpose is only to demonstrate how well each derived result is able to cluster data, none of the sub-figures through this paper adjust the structure of the latent space to align the first PC with the left-right scale.

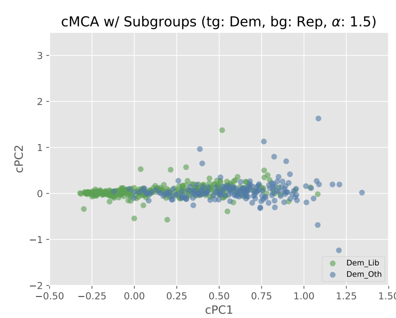
point estimation, the different results (see Fig. 1) together rather demonstrate one similar pattern among American voters. That is, the derived first PC/left-right scale clearly splits American voters along with partisanship (`Dem`: Democrat or `Rep`: Republican), with some exceptions. Given that the positions are estimated through voters' self-reported values, these results demonstrate that the predispositions which the U.S. voters heavily rely on are highly associated with their vote choices and party preferences (Bolsen, Druckman, and Cook 2014; Zaller et al. 1992). In other words, all sub-figures of Fig. 1 demonstrate the existence of ideological polarization among the U.S. public identified previously in the literature (Fiorina and Abrams 2008; Iyengar and Westwood 2015; Lelkes 2016; Smidt 2017). We further apply cMCA to this case to examine its utility in situations that regular methods produce meaningful distinctions between subgroups.



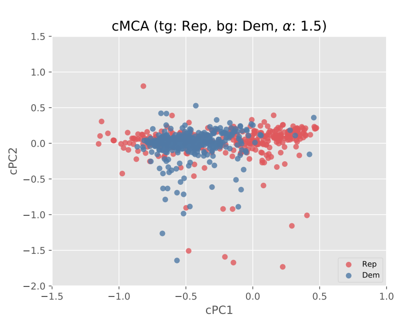
(a) Target: Dem, Background: Rep (Original cMCA Results)



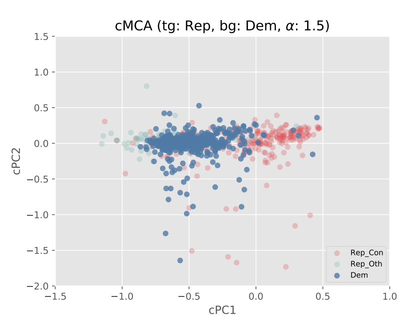
(b) Target: Dem, Background: Rep (with Background)



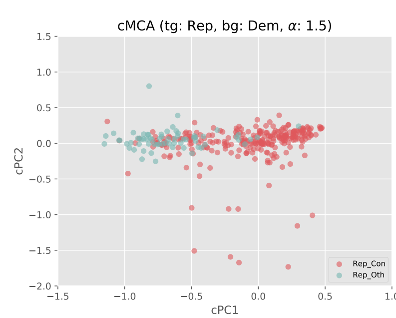
(c) Target: Dem, Background: Rep (without Background)



(d) Target: Rep, Background: Dem (Original cMCA Results)



(e) Target: Rep, Background: Dem (with Background)



(f) Target: Rep, Background: Dem (without Background)

Figure 2: cMCA Results of CCES 2015 (Dem versus Rep)

We begin by assigning Democrats to the target group and Republicans to the background group and present the cMCA results (i.e., data-point coordinates) on the top panel of Fig. 2. As Fig. 2a shows, when α is set to 1.5, the Democrats (in `blue` color) can be divided by the Republicans (in `red` color) along the first contrastive PC (cPC1). This indicates that cMCA uncovers the direction (i.e., cPC1) along which the Democrats have much higher variations than the Republicans. Utilizing the auxiliary information provided by category coordinates and loadings in Appendix A, we decide to further categorize Democrats based on three variables with the highest loadings, which are `military-diplomacy priority`, `military strength`, and `unequal right not a big problem`.[6] More precisely, the Democrats who hold liberal opinions over at least one of these issues are colored in `green` (`Dem_Lib`) and the rest are colored in `blue` (`Dem_Oth`) (see Fig. 2b and Fig. 2c). By and large, what we conclude through Fig. 2c is that compared with the rest of the selected issues, these three variables together, `military-diplomacy priority`, `military strength`, and `unequal rights not a big problem`, are the issues which divide the Democrats into two sub-groups: the subgroup of extremely liberal Democrats that differ from the rest of the party's supporters.

On the other hand, cMCA also discovers a specific pattern within Republicans when assigning Democrats to the background group. As presented in Fig. 2d, when α is set to 1.5, we find that Republicans (in `red` color) are split by Democrats (in `blue` color) into two subgroups along cPC1. According to the category coordinates and factor loadings, the variables that divide Republicans are not the same as those that divide Democrats. Instead, we find that the fundamental variables, `social ideology`, `economic ideology`, and `national security ideology`, are the most influential.[7] More precisely, as Fig. 2e illustrates, the method identifies a subgroup of extremely conservative Republicans (`Rep_Con` in `red` color) that differ from the rest of the party's supporters (`Rep_Oth` in `teal` color).

By and large, Fig. 2 demonstrates that even when the PCs derived from the traditional

6. All three variables are three-point scales where 1 refers to ideologically conservative, 2 refers to ideologically moderate, and 3 refers to ideologically liberal. Please refer to online Appendix B for more details.

7. Again, all three variables are three-point scales where 1 refers to ideologically conservative, 2 refers to ideologically moderate, and 3 refers to ideologically liberal. Please refer to online Appendix B for more details.

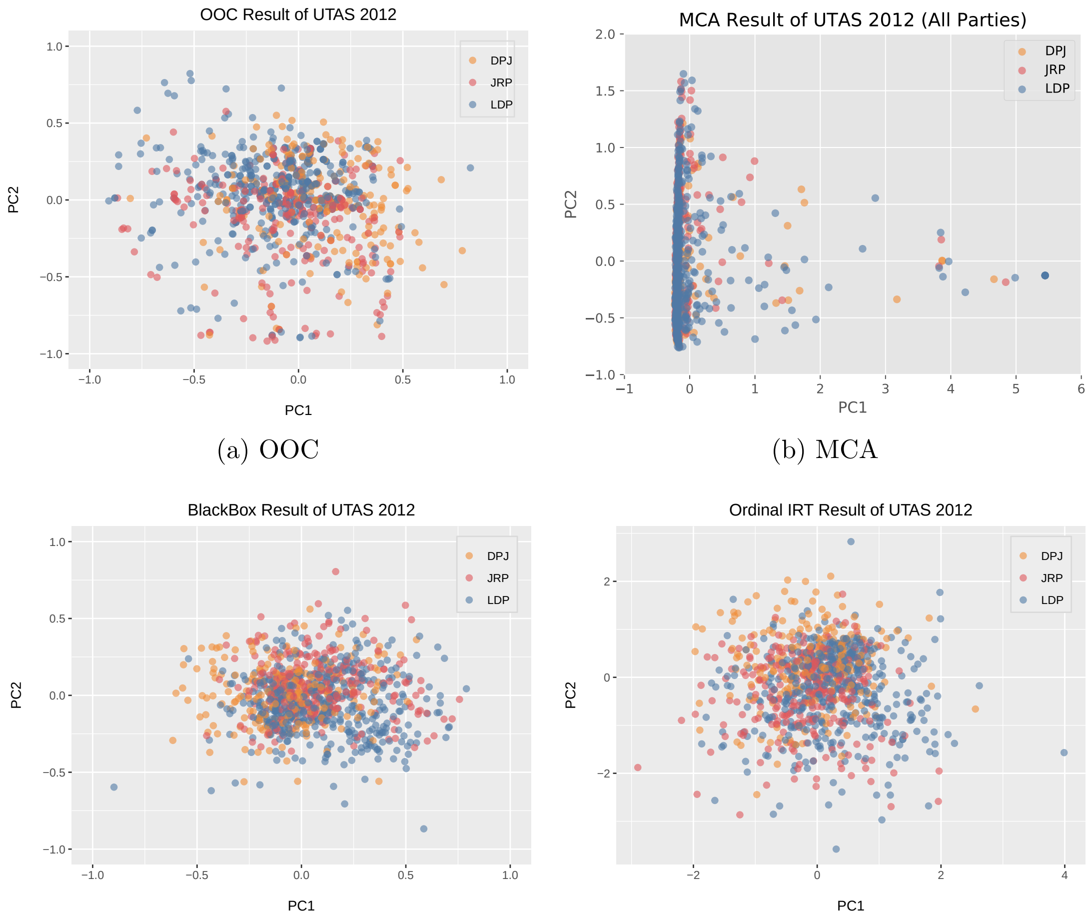
methods are informative, cMCA can help identify the more influential variables and categories within each of predefined groups relative to the other. The cMCA results provide deeper contextual explanation regarding the hidden patterns within each group and how those groups are comprised internally.

Case Two: UTAS 2012—the Hidden Xenophobic Wing of the LDP

Next, we turn our attention to the 2012 UTokyo-Asahi Survey, conducted in Japan during their House of Representatives election in 2012.[8] For each election, UTAS surveys both candidates and voters, and thus contains three sets of survey questions: candidate-specific, voter-specific, and shared questions. We focus on the voter-specific and shared questions only, and again, with OOC, MCA, BS, and BMFA, we present the positions of all respondents who support either the Liberal Democratic Party (LDP), the Democratic Party of Japan (DPJ), or the Japan Renewal Party (JRP) in Fig. 3.

Compared with the results demonstrated in Fig. 1 where the derived PC space and the left-right scale can easily divide American voters by partisanship, the UTAS results instead highlight how Japanese voters are highly mixed and overlapping along these dimensions. Unlike the CCES results, not one of the sub-figures of Fig. 3 reveal any obvious patterns that represent clear and distinct party lines. Importantly, this implies that the level of polarization in Japanese politics is relatively low, at least compared to the United States. These results are consistent with previous research finding a downward trend in the ideological distinctions between Japanese parties, elites, and voters (Endo and Jou 2014; Jou and Endo 2016; Miwa 2015). In short, while simple, left-right ideology may still be the heuristic that Japanese voters rely on for determining their vote-choice, the political culture of Japan has become decidedly more centrist. This is a perfect example illustrating when standard dimensional analyses can lead to uninformative PCs. To further illustrate cMCA's contribution when

8. UTAS is conducted by Masaki Taniguchi of the Graduate Schools for Law and Politics, the University of Tokyo and the Asahi News. The original data can be retrieved from http://www.masaki.j.u-tokyo.ac.jp/utas/utasindex.html.



(a) OOC

(b) MCA

(c) BS

(d) BMFA

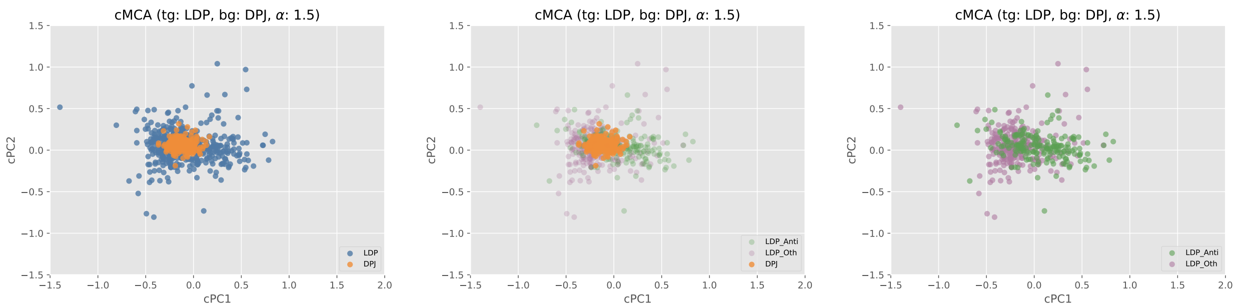
Figure 3: OOC, MCA, BS, and BMFA Results of UTAS 2012

standard methods provide lackluster results, we apply cMCA to UTAS 2012.

As Fig. 4 shows, when we assign LDP supporters to the target group and DPJ supporters to the background group, cMCA identifies two subgroups, with minor overlap, within LDP supporters. When α is set to 1.5, cMCA detects that the LDP supporters (in `blue` color) can be divided by the DPJ supporters (in `orange` color) along cPC1, as demonstrated in Fig. 4a. More precisely, based on the auxiliary information provided in Fig. 10a and Fig. 10b, we find that LDP supporters who are located on the right of Fig. 4a are strongly associated with two categories: `extremely feeling distant to South Korea` or `extremely feeling distant to China`.[9] We further divide the LDP supporters into two

9. The original variables for these two categories are `Feeling Close to South Korea` and `Feeling`

subgroups and present them in Fig. 4b and Fig. 4c: those who feel extremely distant to South Korea or China are colored in `green` (`LDP_Anti`) and the rest is colored in `purple` (`LDP_Oth`). We consider these feelings of extreme distance as anti-South Korea and anti-China attitudes, respectively. Importantly, this xenophobia identified within the LDP is consistent with the reality in current Japanese politics.



(a) Target: LDP, Background: DPJ (Original cMCA Results)
(b) Target: LDP, Background: DPJ (with Background)
(c) Target: LDP, Background: DPJ (without Background)

Figure 4: cMCA Results of UTAS 2012 (LDP versus DPJ)

Indeed, scholars have consistently identified a wave of xenophobic activism in Japan against Koreans, Chinese, and other minorities since the late 2000s—the "Action Conservative Movement" (ACM) (Smith 2018). These new right-wing groups are considered different than the traditional right-wing politicians and parties of post-war Japan. While these traditional parties, which are the predecessors of the modern LDP and the Shinzo Abe administration, focused on anti-communism and the emperor-centered view of history and the state of nature, these new parties are much more preoccupied with nativistic and xenophobic issues and policies (Gill 2018; Smith 2018; Yamaguchi 2018). Interestingly, although these new right-wing groups in the LDP have been identified and discussed by the news media and scholarly research, they are rarely identified through surveys and traditional methods of ideal point estimation. The use of cMCA on this data perfectly illustrates the utility of this method in capturing influential features or categories, that are substantively important, and

`Close to China` separately. Both variables are five-point Likert scales where 1 refers to feeling extremely not close to South Korea/China and 5 refers to feeling extremely close to South Korea/China. Please refer to online Appendix B for more details.

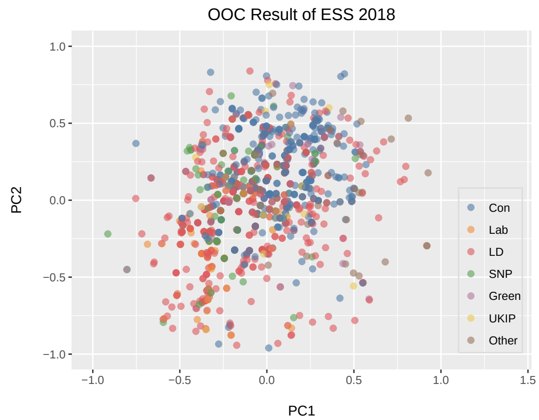
which are often left unidentified by traditional methods.

Case Three: ESS 2018 (the U.K. Module)—the Brexit Cleavage among the Labor and Conservative Parties

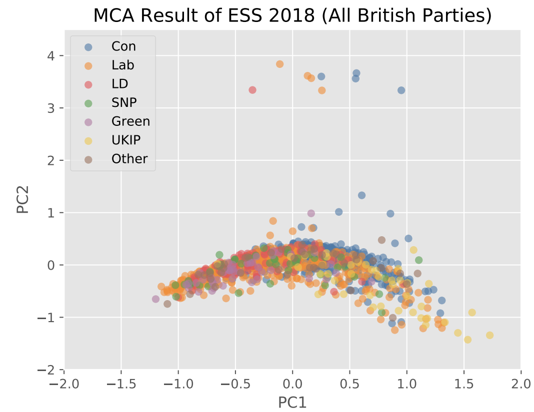
Finally, we focus on the British module of the 2018 European Social Survey. Fig. 5 shows the OOC, MCA, BS, and BMFA results for supporters of all major parties in the U.K.: the Conservative (Con), Labour (Lab), Liberal Democratic (LD), Scottish National (SNP), Green, and UK Independence Party (UKIP) (and others). According to Fig. 5, similarly to the case of UTAS 2012, all of the estimated results indicate that there are no clear party lines under the ordinary PC space. This suggests that the status of political polarization among American voters does not clearly exist among the mass public in the U.K. This conclusion is indeed consistent with the current academic understanding that the U.K. has undergone a period of political depolarization since the second wave of ideological convergence between the elites of the two major parties, the Conservative and Labour parties (Adams, Green, and Milazzo 2012a, 2012b; Leach 2015).

Furthermore, this survey was conducted from August 10th, 2018 through February 22nd, 2019, which was between the two snap elections regarding the UK's departure from the European Union (EU), commonly known as Brexit (Cutts et al. 2020; Heath and Goodwin 2017). Theoretically, given how salient and polarized the issue of Brexit is, one should expect that such high salience should be reflected in voters' responses and be revealed on the derived PC space. Similar to the results of UTAS 2012, in which traditional methods did not reveal the xenophobic subgroup of the LDP, the ESS results from traditional methods also fail to separate voters by preferences related to Brexit, which is often considered the most salient and important issue in contemporary British politics.

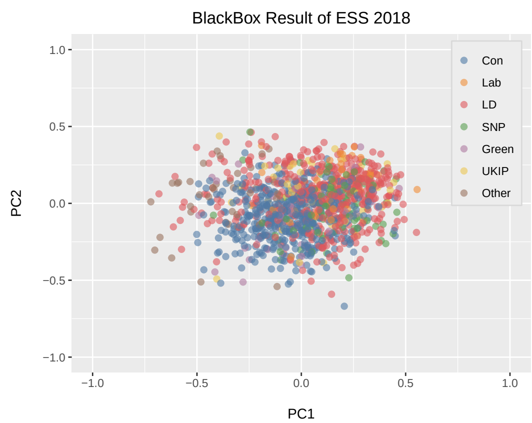
We apply cMCA to ESS 2018 and present the results in Fig. 6 and Fig. 7. As shown in Fig. 6, the first pair we examine includes the two major parties, Conservative and the Labour. However, the results are relatively uninteresting. Fig. 6a and Fig. 6b together show



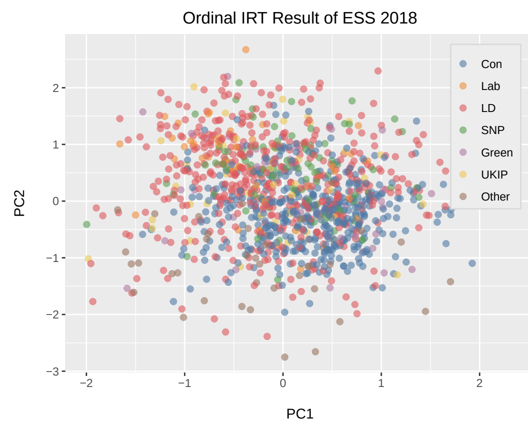
(a) OOC



(b) MCA



(c) BS

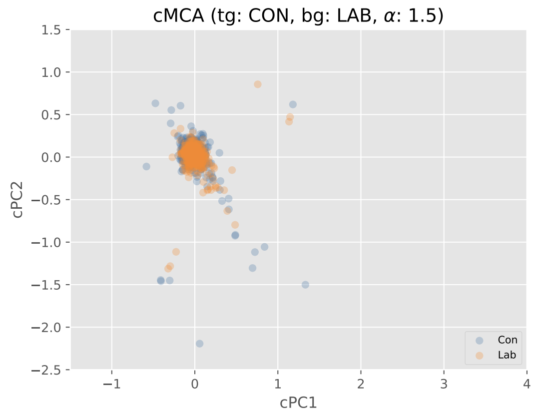


(d) BMFA

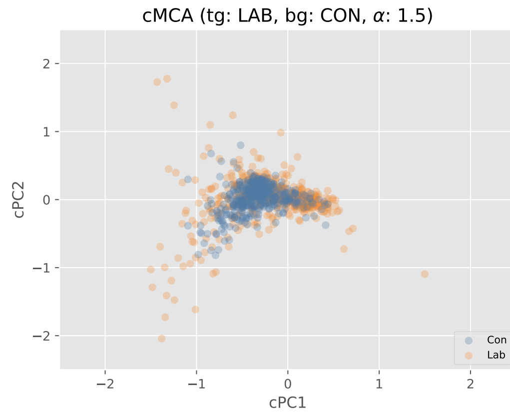
Figure 5: OOC, MCA, BS, and BMFA Results of ESS 2018

that the two main parties fail to divide each other in a meaningful way. In both figures, given that the clusters of Conservative and the Labour supporters almost perfectly stack onto one another, regardless of which one is the target or background group, we conclude that there are no hidden patterns discernible by cMCA between these two main parties.

Nevertheless, the sub-figures of Fig. 7 tell a very different story. Instead of comparing the two main parties themselves, we separate the parties and compare them as target groups with UKIP assigned to the background group. Through Fig. 7a, we observe that when UKIP supporters (the `black` points in Fig. 7a) are the anchor, it splits Conservatives (the `pink` points in Fig. 7a) into two adjacent sub-groups along with the second contrastive PC (cPC2) when α is set equal to 100. Based on auxiliary information provided by sub-figures



(a) Target: Con, Background: Lab (Original cMCA Results)



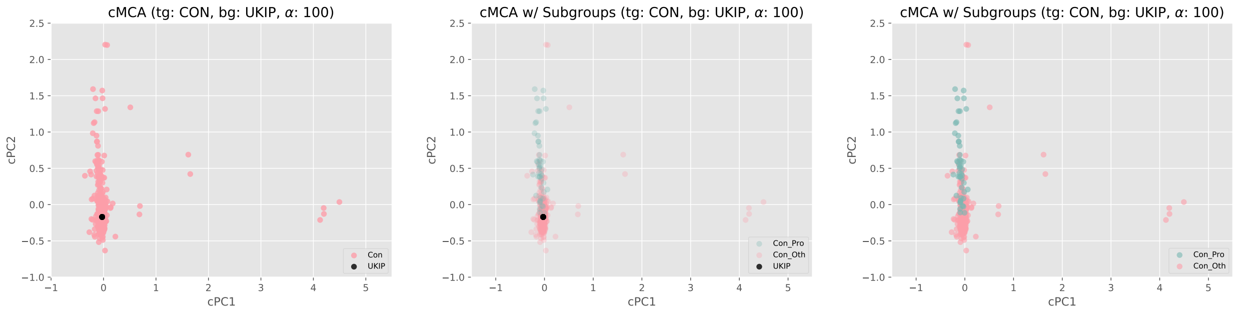
(b) Target: Lab, Background: Con (Original cMCA Results)

Figure 6: cMCA Results of ESS 2018 (Con versus Lab)

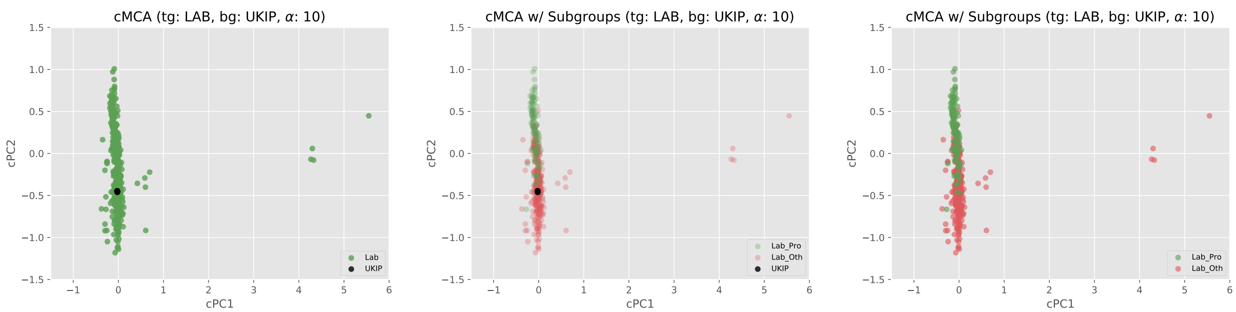
on the bottom panel of Fig. 11 in Appendix A, we find that factors related to attitudes regarding the European Union (EU) and immigration polarize Conservatives. More precisely, we find that Conservative supporters who are on the top of Fig. 7a are associated with extremely open views over at least one of the three variables: `Immigration to Living Quality`, `Immigration to Culture`, and `Emotionally Attached to Europe`.[10] As shown in Fig. 7b and Fig. 7c, we further color this extremely EU-supportive sub-group (`Con_Pro`) in `teal` and color the rest of the Conservative supporters (`Con_Oth`) in `pink`.

We also find a sub-group within the Labour party by using UKIP supporters as the background group. Similar to the previous results, we find that when α is set equal to 100, Labour supporters (the `green` points in Fig. 7d) are split by UKIP supporters (the `black` points in Fig. 7d) along cPC2. The results in the bottom panel of Fig. 12, indicate that some Labour supporters are pro-immigration and pro-EU. We further divide the Labour supporters into two sub-groups with different colors and present them in Fig. 7e and Fig. 7f: those who hold extremely open opinions over at least one of the three variables, i.e., `Immigration to Living Quality`, `Immigration to Culture`, or `Emotionally Attached to Europe`, are in `green` (`Lab_Pro`) and the rest colored `red` (`Lab_Oth`).

10. All three variables are five-point scales where 1 refers to the feeling of extreme anti-EU or anti-immigrant and 5 to the feeling of extreme pro-EU or pro-immigrant. Please refer to online Appendix B for more details.



(a) Target: Con, Background: UKIP (Original cMCA Results) (b) Target: Con, Background: UKIP (with Background) (c) Target: Con, Background: UKIP (without Background)



(d) Target: Lab, Background: UKIP (Original cMCA Results) (e) Target: Lab, Background: UKIP (with Background) (f) Target: Lab, Background: UKIP (without Background)

Figure 7: cMCA Results of ESS 2018 (Con versus UKIP, Lab versus UKIP)

Overall, given that recent research has linked negative immigration attitudes to anti-EU sentiments (Colantone and Stanig 2016), we conclude that both Conservative and Labour party supporters are internally divided by Brexit attitudes. These results demonstrate, similar to traditional methods, that cMCA can identify divisions that generalize *across* pre-defined groups. As we may expect, although the two parties are divided by Brexit attitudes, the size of the pro-EU sub-group is significantly larger among Labour supporters, as compared to Conservative ones.

The ESS 2018 analysis emphasizes an important component of cMCA that was not present in the first two cases. That is, selecting the background group can be highly important for deriving meaningful results. While this is intuitive, given the primary idea of contrastive learning methods, it is important to keep in mind when utilizing cMCA. If the selected target and background groups are highly similar, cMCA may not be able to capture any informative patterns, as illustrated by the results in Fig. 6a and Fig. 6b. Nevertheless, this uninformative situation can still be informative from a different perspective—researchers can apply the contrastive learning method to any two groups to examine the level of their similarity. In this perspective, the results of Fig. 6a and Fig. 6b in truth validate the previous findings that there has been a wave of ideological convergence between the two major parties in British politics.

Discussion

Ideal point estimation and dimensional reduction techniques have been widely used for both pattern recognition and latent-space derivation. Nevertheless, due to the structure of different datasets and the methodology behind these ordinary PC methods, the derived results are often uninformative. In this article, we contribute to this literature by applying the contrastive learning to multiple correspondence analysis enabling researchers to derive contrasted dimensions using categorical data.

Through the analytical results from three surveys, CCES 2015, UTAS 2012, and ESS 2018, we illustrate the two main advantages of cMCA, relative to traditional PC methods. First, even when traditional PC methods find meaningful low-dimensional spaces and divisions, cMCA can provide complementary information regarding what factors/issues are salient *within* each group. Secondly, cMCA can identify substantively important dimensions and divisions among (sub-)groups in situations where traditional PC results are uninformative. Note that while cMCA explores principal directions for individual groups mainly—it can still derive general PCs that can be applied to all groups. As demonstrated in Fig. 7, when Conservative and Labour supporters are distributed almost identically, cMCA identifies the Brexit division that exists in both parties.

References

Abid, Abubakar, Martin J. Zhang, Vivek K. Bagaria, and James Zou. 2018. "Exploring Patterns Enriched in a Dataset with Contrastive Principal Component Analysis." *Nature Communications* 9 (1): 1–7.

Abid, Abubakar, and James Zou. 2019. "Contrastive Variational Autoencoder Enhances Salient Features." *arXiv:1902.04601.*

Adams, James, Jane Green, and Caitlin Milazzo. 2012a. "Has the British Public Depolarized along with Political Elites? An American Perspective on British Public Opinion." *Comparative Political Studies* 45 (4): 507–530.

———. 2012b. "Who Moves? Elite and Mass-Level Depolarization in Britain, 1987–2001." *Electoral Studies* 31 (4): 643–655.

Aldrich, John H, and Richard D McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections." *American Political Science Review* 71 (1): 111–130.

Ansolabehere, Stephen, and Brian Schaffner. 2017. *CCES, Common Content, 2015.* Version V2. doi:10.7910/DVN/SWMWX8. https://doi.org/10.7910/DVN/SWMWX8.

Armstrong, David, Ryan Bakker, Royce Carroll, Christopher Hare, Keith Poole, and Howard Rosenthal. 2014. *Analyzing Spatial Models of Choice and Judgment with R.* FL: CRC Press.

Bolsen, Toby, James N Druckman, and Fay Lomax Cook. 2014. "The Influence of Partisan Motivated Reasoning on Public Opinion." *Political Behavior* 36 (2): 235–262.

Brady, Henry. 1990. "Dimensional Analysis of Ranking Data." *American Journal of Political Science:* 1017–1048.

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355–370.

Colantone, Italo, and Piero Stanig. 2016. "Global Competition and Brexit." *American Political Science Review* 112 (2): 201–218.

Coombs, Clyde H. 1964. *A Theory of Data.* New York: Wiley.

Cutts, David, Matthew Goodwin, Oliver Heath, and Paula Surridge. 2020. "Brexit, the 2019 General Election and the Realignment of British Politics." *The Political Quarterly* 91 (1): 7–23.

Davis, Darren, Kathleen Dowley, and Brian Silver. 1999. "Postmaterialism in World Societies: Is It Really a Value Dimension?" *American Journal of Political Science* 43 (3): 935–962.

Endo, Masahisa, and Willy Jou. 2014. "How Does Age Affect Perceptions of Parties' Ideological Locations?" *Japanese Journal of Electoral Studies* 30 (1): 96–112.

Fiorina, Morris, and Samuel Abrams. 2008. "Political Polarization in the American Public." *Annual Review of Political Science* 11:563–588.

Fujiwara, Takanori, Oh-Hyun Kwon, and Kwan-Liu Ma. 2020. "Supporting Analysis of Dimensionality Reduction Results with Contrastive Learning." *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 45–55.

Fujiwara, Takanori, Jian Zhao, Francine Chen, Yaoliang Yu, and Kwan-Liu Ma. 2020. "Interpretable Contrastive Learning for Networks." *arXiv:2005.12419.*

Ge, Rong, and James Zou. 2016. "Rich Component Analysis." In *Proceedings of the International Conference on Machine Learning,* 48:1502–1510.

Gill, Tom. 2018. "The Nativist Backlash: Exploring the Roots of the Action Conservative Movement." *Social Science Japan Journal* 21 (2): 175–192.

Greenacre, Michael. 2017. *Correspondence Analysis in Practice.* New York: Chapman & Hall/CRC.

Grimmer, Justin, and Gary King. 2011. "General Purpose Computer-Assisted Clustering and Conceptualization." In *Proceedings of the National Academy of Sciences,* 108:2643–2650. 7. National Acad Sciences.

Hare, Christopher, David A Armstrong, Ryan Bakker, Royce Carroll, and Keith T Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions." *American Journal of Political Science* 59 (3): 759–774.

Hare, Christopher, Tzu-Ping Liu, and Robert N Lupton. 2018. "What Ordered Optimal Classification Reveals about Ideological Structure, Cleavages, and Polarization in the American Mass Public." *Public Choice* 176 (1-2): 57–78.

Harris, David, and Sarah Harris. 2010. *Digital Design and Computer Architecture.* CA: Morgan Kaufmann Publishers.

Heath, Oliver, and Matthew Goodwin. 2017. "The 2017 General Election, Brexit and the Return to Two-Party Politics: An Aggregate-Level Analysis of the Result." *The Political Quarterly* 88 (3): 345–358.

Hix, Simon, Abdul Noury, and Gerard Roland. 2006. "Dimensions of Politics in the European Parliament." *American Journal of Political Science* 50 (2): 494–520.

Imai, Kosuke, James Lo, and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110 (4): 631–656.

Iyengar, Shanto, and Sean J Westwood. 2015. "Fear and Loathing Across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707.

Jacoby, William G. 1999. “Levels of Measurement and Political Research: An Optimistic View.” *American Journal of Political Science* 43 (1): 271–301.

Jou, Willy, and Masahisa Endo. 2016. “Ideological Understanding and Voting in Japan: A Longitudinal Analysis.” *Asian Politics & Policy* 8 (3): 456–473.

Le Roux, Brigitte, and Henry Rouanet. 2010. *Quantitative Applications in the Social Sciences: Multiple Correspondence Analysis.* CA: SAGE Publications, Inc.

Leach, Robert. 2015. *Political Ideology in Britain.* London: Macmillan International Higher Education.

Lelkes, Yphtach. 2016. “Mass Polarization: Manifestations and Measurements.” *Public Opinion Quarterly* 80 (S1): 392–410.

Martin, Andrew, and Kevin Quinn. 2002. “Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999.” *Political Analysis* 10 (2): 134–153.

Martin, Andrew, Kevin Quinn, and Jong Hee Park. 2011. “MCMCpack: Markov Chain Monte Carlo in R.” *Journal of Statistical Software* 42 (9): 1–21.

McCarty, Nolan, Keith Poole, and Howard Rosenthal. 2001. “The Hunt for Party Discipline in Congress.” *American Political Science Review* 95 (3): 673–687.

Miwa, Hirofumi. 2015. “Voters’ Left–Right Perception of Parties in Contemporary Japan: Removing the Noise of Misunderstanding.” *Japanese Journal of Political Science* 16 (1): 114–137.

Monroe, Burt, Michael Colaresi, and Kevin Quinn. 2008. “Fightin’ Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict.” *Political Analysis* 16 (4): 372–403.

Pagès, Jérôme. 2014. *Multiple Factor Analysis by Example Using R.* FL: CRC Press.

Poole, Keith. 1998. "Recovering a Basic Space from a Set of Issue Scales." *American Journal of Political Science* 42 (3): 954–993.

———. 2000. "Nonparametric Unfolding of Binary Choice Data." *Political Analysis* 8 (3): 211–237.

Poole, Keith, and Howard Rosenthal. 1991. "Patterns of Congressional Voting." *American Journal of Political Science* 35 (1): 228–278.

Quinn, Kevin. 2004. "Bayesian Factor Analysis for Mixed Ordinal and Continuous Responses." *Political Analysis* 12 (4): 338–353.

Rosenthal, Howard, and Erik Voeten. 2004. "Analyzing Roll Calls with Perfect Spatial Voting: France 1946–1958." *American Journal of Political Science* 48 (3): 620–632.

Severson, Kristen A, Soumya Ghosh, and Kenney Ng. 2019. "Unsupervised Learning with Contrastive Latent Variable Models." In *Proceedings of the AAAI Conference on Artificial Intelligence,* 33:4862–4869.

Smidt, Corwin. 2017. "Polarization and the Decline of the American Floating Voter." *American Journal of Political Science* 61 (2): 365–381.

Smith, Nathaniel M. 2018. "Fights on the Right: Social Citizenship, Ethnicity, and Postwar Cohorts of the Japanese Activist Right." *Social Science Japan Journal* 21 (2): 235–257.

Wenskovitch, John, Ian Crandell, Naren Ramakrishnan, Leanna House, and Chris North. 2018. "Towards a Systematic Combination of Dimension Reduction and Clustering in visual Analytics." *IEEE Transactions on Visualization and Computer Graphics* 24 (1): 131–141.

Wilkerson, John, and Andreu Casas. 2017. "Large-Scale Computerized Text Analysis in Political Science: Opportunities and Challenges." *Annual Review of Political Science* 20:529–544.

Yamaguchi, Tomomi. 2018. "Revisionism, Ultranationalism, Sexism: Relations between the Far Right and the Establishment over the 'Comfort Women' Issue." *Social Science Japan Journal* 21 (2): 193–212.

Zaller, John R, et al. 1992. *The Nature and Origins of Mass Opinion.* Cambridge: Cambridge University Press.

Zou, James Y, Daniel J Hsu, David C Parkes, and Ryan P Adams. 2013. "Contrastive Learning Using Spectral Methods." In *Proceedings of the International Conference on Neural Information Processing Systems,* 26:2238–2246.

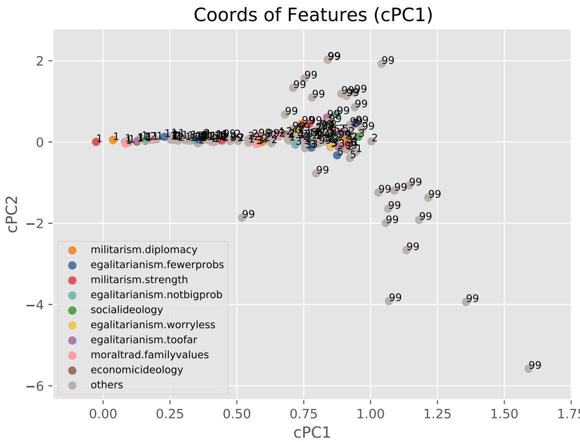
Appendix A

CCES 2015: Category Coordinates Plots and Category Loadings Plots

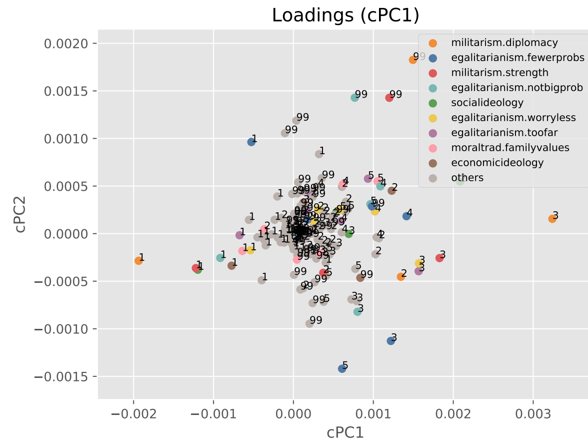
To decide which categories or variables heavily affect division within the original group in the latent space, one needs two pieces of auxiliary information: the category coordinates plot and the category loadings plot. The first plot demonstrates the position of each category in the derived latent space. In addition to estimating respondents' positions in the latent space like PCA, MCA could estimate positions of each category of each variable. Similar to PCA, the location of a certain category derived by MCA indicates that respondents whose positions are around that area should be associated with that category (Pagès 2014).

Information conceived in the second plot, the category loadings plot, is how much each category contributes to the dispersion of data on each principal dimension under the latent space. In such a plot, the presented position reflects the magnitude of contribution. For instance, when a given category locates on the very left of the plot, it indicates that respondents whose positions are on the left of the plot are highly likely to be associated with that specific category. In addition, we further construct variable loadings through normalizing category loadings first and summing all the absolute values of category loadings by PCs and variables. Based on this formula, variables that have higher loadings are those whose categories are dispersed away from each other. In other words, variables with high loadings indicate they are likely the ones that divide data into different clusters.

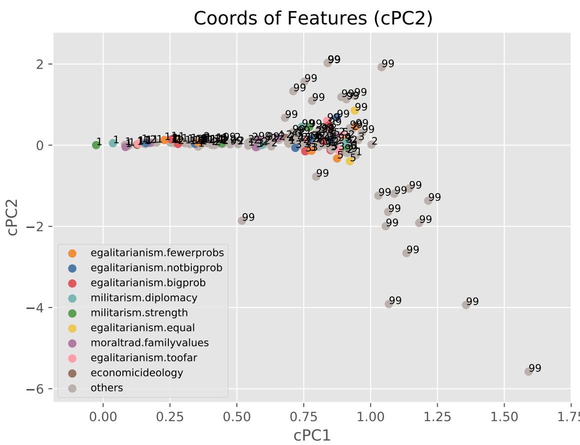
In this appendix, we demonstrate how to practically interpret both category coordinates and category loadings plots and further retrieve auxiliary information to re-cluster data. We first discuss results of CCES 2015 in Fig. 8. First of all, we demonstrate category coordinates in Fig. 8a and Fig. 8c for the combination of the Democrats as the target group and the Republicans as the background group. As demonstrated, both figures present the same category coordinates. However, because the CCES 2015 data contains many different categorical variables, we color-code only the top nine variables in Fig. 8a and Fig. 8c based on



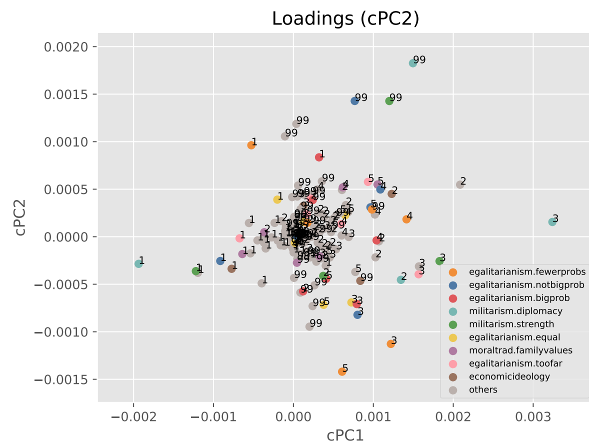
(a) Category coordinates colored by each variable's total of absolute category loadings along cPC1.



(b) Category loadings colored by each variable's total of absolute category loadings along cPC1.



(c) Category coordinates colored by each variable's total of absolute category loadings along cPC2.



(d) Category loadings colored by each variable's total of absolute category loadings along cPC2.

Figure 8: CCES2015 (Target: Dem, Background: Rep): Category Coordinates Plots and Category Loadings Plots

each variable's sum total of normalized absolute category loadings along the first contrastive PC (cPC1) and the second contrastive PC (cPC2), respectively. The order of the colors from top to bottom in the figure legend correspond to the ranks in which the variables were loaded.[11] We apply this color-coding scheme to all of category coordinates and category loadings plots. As presented in Fig. 2c, given that the target group is almost only divided along with the first cPC, we will focus on Fig. 8a to derive influential categories and variables.

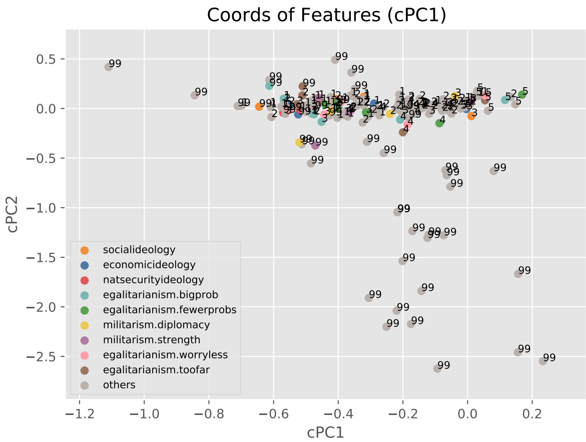
11. For example, as shown in Fig. 8a and Fig. 8c, while `militarism.diplomacy` has the highest total absolute category loadings along cPC1, `egalitarianism.fewerprobs` has the highest rank along cPC2.

Fig. 8a demonstrates that when respondents hold extremely liberal opinions over the first three questionnaires with the highest loadings, they are more likely to be clustered on the left of Fig. 2c. Furthermore, we demonstrate loadings of each category in Fig. 8b and Fig. 8d (again, the latter one is dropped from the analysis). As discussed, Fig. 8a shows that variables with the highest loadings are the ones whose categories are dispersed more. For instance, compared with other variables, the categories of the variable `egalitarianism.fewerprobs` are separated away from each other, especially the category of 1 and the categories of 3, 4, and 5—this indicates that the Democratic supporters who hold extremely liberal opinion over this issue are distinct from the rest of the Democratic supporters. Overall, based on category coordinates and loadings, we decide to take three variables: `militarism.diplomacy`, `egalitarianism.fewerprobs`, and `militarism.strength` to further cluster the Democratic supporters into two subgroups—those who hold an extremely liberal opinion on either of the three variables and the rest, and present the results in Fig. 2c.

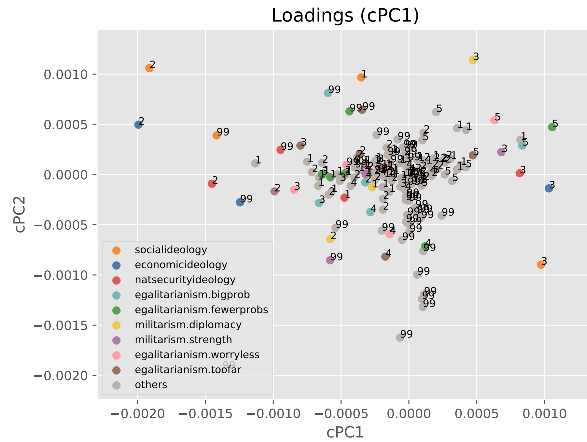
We further demonstrate how to interpret figures in Fig. 9 to cluster Republican supporters. As Fig. 2e presents, similarly, given that data is mainly dispersed along with cPC1, we only focus on Fig. 9a and Fig. 9b. Based on derived information from the two subfigures, one can find that respondents who are clustered to the right end of the Fig. 2e are associated with (extremely) conservative opinions over questionnaires/variables. We further cluster Republican supporters by their opinions on three main variables—`socialideology`, `economicideology`, and `natsecurityideology`, i.e. the three variables with the largest loadings. More specifically, we cluster Republican supporters into two subgroups, those who responded the category of 3 on one of the three variables and the rest. These results are presented in Fig. 2f.

UTAS 2012: Category Coordinates Plots and Category Loadings Plots

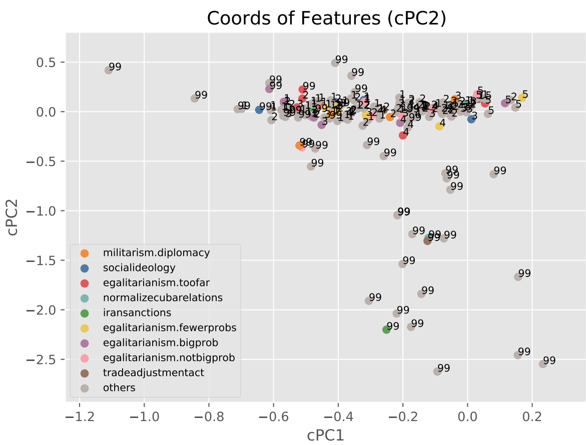
We next demonstrate the logic we adopt to cluster sub-groups within the LDP in Fig. 4c. Similarly to the CCES 2015 case, we only focus on auxiliary information from Fig. 10a and



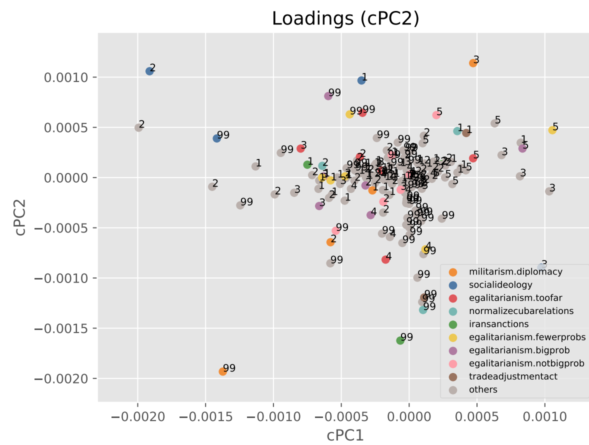
(a) Category coordinates colored by each variable's total of absolute category loadings along cPC1.



(b) Category loadings colored by each variable's total of absolute category loadings along cPC1.



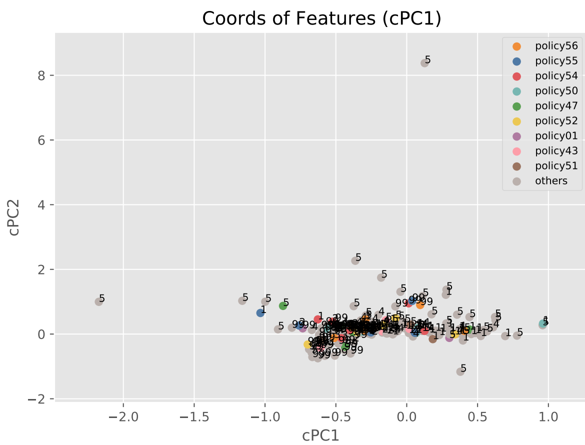
(c) Category coordinates colored by each variable's total of absolute category loadings along cPC2.



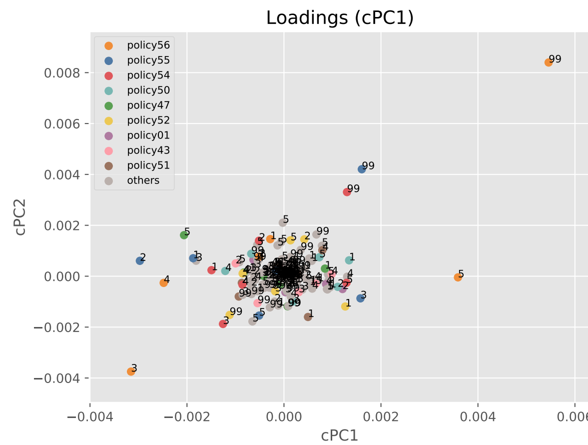
(d) Category loadings colored by each variable's total of absolute category loadings along cPC2.

Figure 9: CCES2015 (Target: Rep, Background: Dem): Category Coordinates Plots and Category Loadings Plot

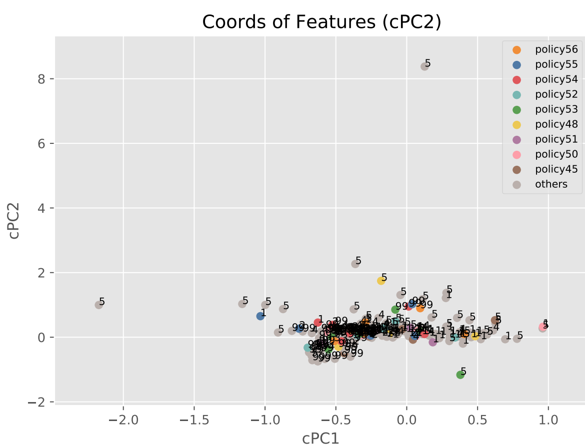
Fig. 10b given that the sub-groups of LDP are separated mainly along with cPC1. First of all, we find that the top three questionnaires with the highest loadings are `policy56`, `policy55`, and `policy54` which refer to the feeling thermometer to South Korea, Russia, and China individually. Furthermore, we find that LDP supporters who are clustered on the left end of Fig. 4c are associated with extreme unfamiliarity with South Korea and China (the category of 5) and moderate familiarity with Russia (the category of 3). Based on the political reality of Japan and category similarity, we only take `policy56` and `policy54` to further cluster the LDP supporters. Those who are extremely unfamiliar either with South Korea or with China are separated from the rest of the LDP supporters.



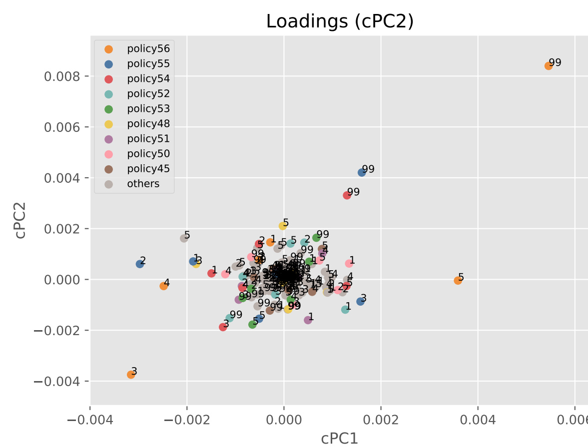
(a) Category coordinates colored by each variable's total of absolute category loadings along cPC1.



(b) Category loadings colored by each variable's total of absolute category loadings along cPC1.



(c) Category coordinates colored by each variable's total of absolute category loadings along cPC2.

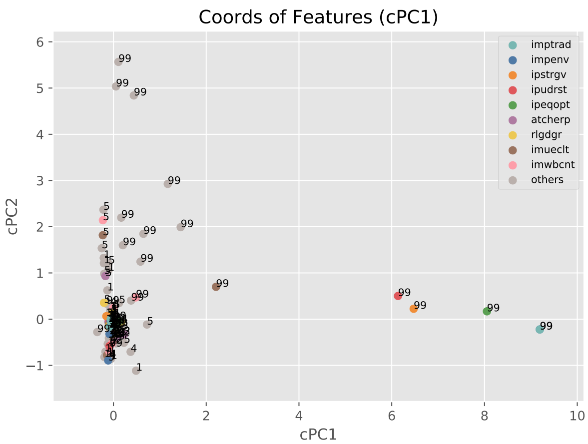


(d) Category loadings colored by each variable's total of absolute category loadings along cPC2.

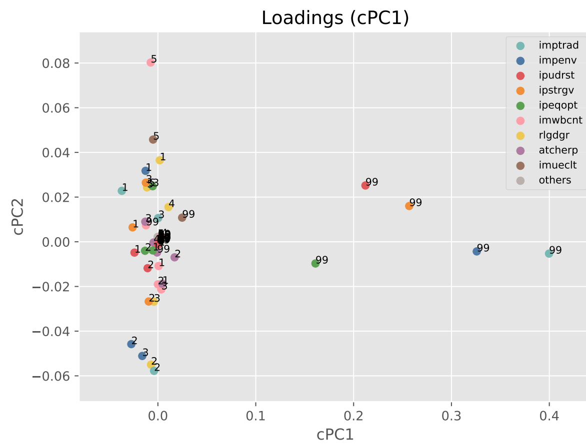
Figure 10: UTAS 2012 (Target: LDP, Background: DPJ): Category Coordinates Plots and Category Loadings Plot

ESS 2018: Category Coordinates Plots and Category Loadings Plots

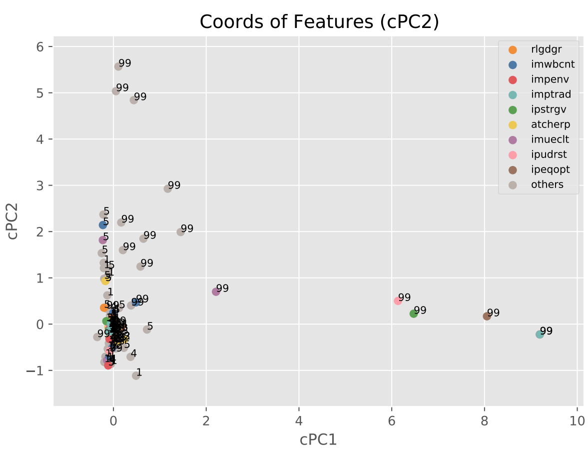
We finally demonstrate how to interpret category coordinates and category loadings to derive sub-groups in Fig. 7. Given that sub-groups of the Conservative supporters are dispersed along with cPC2, we concentrate on the result of Fig. 11c and Fig. 11d. According to the two sub-figures, we find that the Conservative supporters can be divided by their responses toward two immigrant-related questionnaires, `imwbcnt` (whether the mother country's cultural life is undermined or enriched by immigrants) and `imueclt` (whether the mother country becomes a worse or a better place to live with immigrants). More precisely, we further cluster the Conservative supporters by those who hold extremely liberal opinions on either of the two questionnaires (the category of 5) and the rest, such as presented in Fig. 7c.



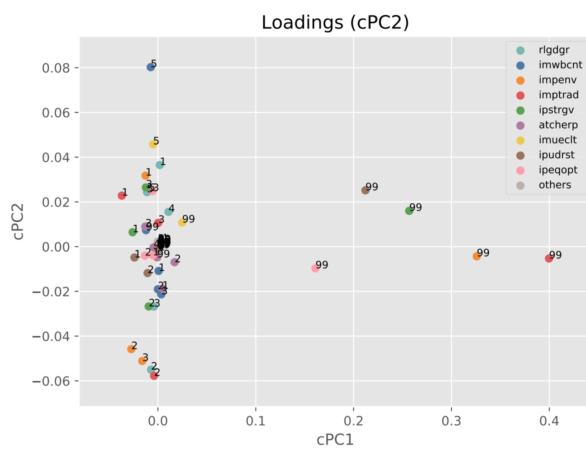
(a) Category coordinates colored by each variable's total of absolute category loadings along cPC1.



(b) Category loadings colored by each variable's total of absolute category loadings along cPC1.



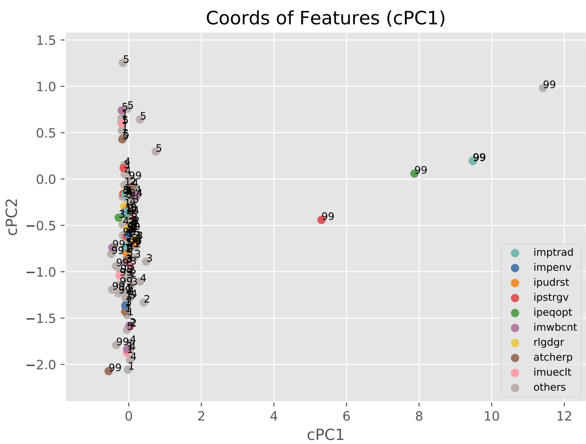
(c) Category coordinates colored by each variable's total of absolute category loadings along cPC2.



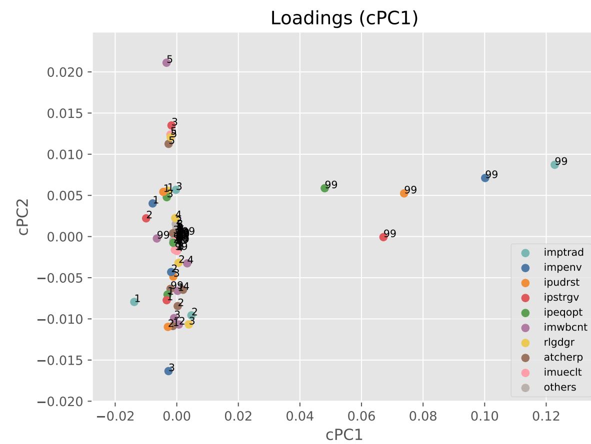
(d) Category loadings colored by each variable's total of absolute category loadings along cPC2.

Figure 11: ESS 2018 (Target: CON, Background: UKIP): Category Coordinates Plots and Category Loadings Plot

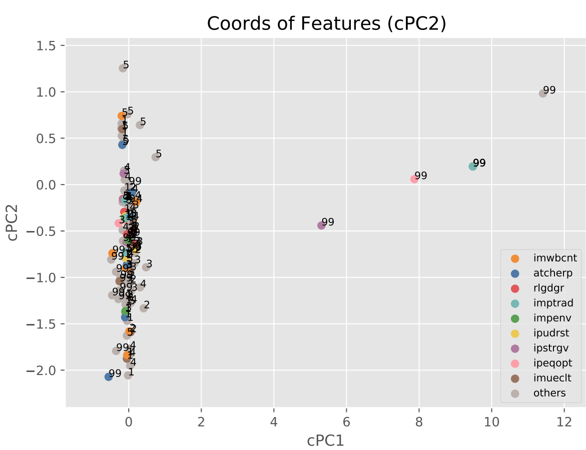
Furthermore, in Fig. 12, we also observe that the Labour supporters could be split into two sub-groups by the UKIP supporters. Through Fig. 12c and Fig. 12d (we only focus on coordinates and loadings along with cPC2 for this sub-case, too), we find that the two immigrant-related questionnaires above still decide how the Labour supporters could be divided. Similarly to the Conservative-UKIP case, the auxiliary information indicates that we can further cluster the Labour supporters in two sub-groups—those who hold extremely liberal responses on either `imwbcnt` or `imueclt` and the rest.



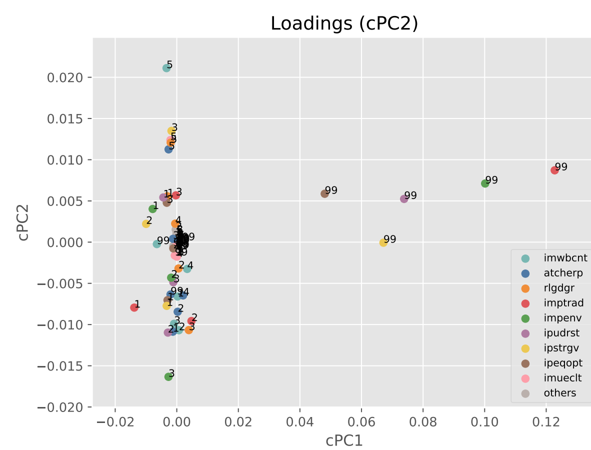
(a) Category coordinates colored by each variable's total of absolute category loadings along cPC1.



(b) Category loadings colored by each variable's total of absolute category loadings along cPC1.



(c) Category coordinates colored by each variable's total of absolute category loadings along cPC2.



(d) Category loadings colored by each variable's total of absolute category loadings along cPC2.

Figure 12: ESS 2018 (Target: LAB, Background: UKIP): Category Coordinates Plots and Category Loadings Plot