Check for updates

Bootstrap Confidence Regions for Multidimensional Scaling Solutions

William G. Jacoby Michigan State University
David A. Armstrong II University of Wisconsin–Milwaukee

Multidimensional scaling (or MDS) is a methodology for producing geometric models of proximities data. Multidimensional scaling has a long history in political science research. However, most applications of MDS are purely descriptive, with no attempt to assess stability or sampling variability in the scaling solution. In this article, we develop a bootstrap resampling strategy for constructing confidence regions in multidimensional scaling solutions. The methodology is illustrated by performing an inferential multidimensional scaling analysis on data from the 2004 American National Election Study (ANES). The bootstrap procedure is very simple, and it is adaptable to a wide variety of MDS models. Our approach enhances the utility of multidimensional scaling as a tool for testing substantive theories while still retaining the flexibility in assumptions, model details, and estimation procedures that make MDS so useful for exploring structure in data.

The term "multidimensional scaling" (usually abbreviated "MDS") refers to a general methodology for producing a geometric model of proximities data.[1] Multidimensional scaling has a long history in political science. However, almost all applications have been descriptive in nature; there have been few attempts to assess the stability of MDS models or to generate inferences about a population structure from a scaling solution that is based upon observed sample data.

In this article, we develop a bootstrap resampling strategy for constructing confidence regions in multidimensional scaling solutions. Our procedure is very simple, it is adaptable to a wide variety of MDS models, and it overcomes some of the limitations associated with earlier attempts to assess statistical stability in multidimensional scaling solutions. This is important because the bootstrap resampling procedure we lay out below helps transform multidimensional scaling from a mere tool for exploring data into a powerful strategy for testing theories.

Background

As motivation for the methodological discussion to follow, we begin with a substantive example in which MDS could be used to address theoretically relevant issues. There is widespread agreement that American politics has polarized along ideological lines during the first decade of the twenty-first century (e.g., Mann and Ornstein 2012; McCarty, Poole, and Rosenthal 2006). But, scholars disagree over the extent to which this polarization has been manifested in public opinion (e.g., Abramowitz and Saunders 2008; Fiorina, Abrams, and Pope 2008).

Most of the previous research on this topic has focused (reasonably enough) on citizens' own issue attitudes. We take a slightly different approach and ask whether people actually perceive the ideological gulf that almost all commentators agree exists at the elite level. From one perspective, the answer to this question might seem obvious: after all, political antagonists' claims and

William G. Jacoby is Professor in the Department of Political Science, Michigan State University, 368 Farm Lane, S303, South Kedzie Hall, East Lansing, MI 48824 (jacoby@msu.edu). David A. Armstrong II is Assistant Professor in the Department of Political Science, University of Wisconsin–Milwaukee, PO Box 413, NWQ B, Room 5450, 2025 E. Newport Avenue, Milwaukee, WI 53201 (armstrod@uwm.edu).

Earlier versions of this article were presented at the November 2010 meeting of the University of North Carolina Political Methods Group and the 2011 annual meetings of the Midwest Political Science Association. The late George Rabinowitz provided a great deal of assistance during the early stages of this project; we are very grateful for his encouragement and suggestions. We also thank Ryan Bakker and Saundra K. Schneider for their many helpful comments on various drafts of the manuscript. All replication materials, including data files, SAS commands, R scripts, and the R package mentioned in the text are available either directly from the authors or online at http://dvn.iq.harvard.edu/dvn/dv/ajps.

[1]Rabinowitz (1975) provides an excellent introduction to nonmetric multidimensional scaling from a political science perspective. More recent comprehensive treatments of MDS are provided by Cox and Cox (2001) and Borg and Groenen (2005).

American Journal of Political Science, Vol. 58, No. 1, January 2014, Pp. 264–278

©2013, Midwest Political Science Association DOI: 10.1111/ajps.12056

accusations are voiced loudly and frequently in the media; it is difficult to imagine that anybody could fail to recognize the differences between them. But, many political scientists assert otherwise. A very long line of research has demonstrated repeatedly that Americans are "innocent of ideology" (e.g., Converse 1964; Jacoby 2002). So, it really is not clear whether the general public does perceive a significant distinction between liberal and conservative political actors, and the question remains: do citizens think about the political world in ideological terms?

We will address the preceding question by analyzing citizens' evaluations of 13 prominent political figures from the period of the 2004 presidential election: George W. Bush, John Kerry, Ralph Nader, Richard Cheney, John Edwards, Laura Bush, Hillary Clinton, Bill Clinton, Colin Powell, John Ashcroft, John McCain, the Democratic party, and the Republican party. Specifically, we will create an empirical estimate of the cognitive structure that the mass public brings to bear when evaluating these political figures. Our approach does not specify in advance the judgmental standards that people use to differentiate political figures. Instead, we will try to infer the public's evaluative criteria from the contents of a model that is, itself, constructed from the data. If the empirical results are consistent with an ideological ordering of the 13 figures, then it would be strong evidence that the mass public really does employ the liberal-conservative continuum as an organizational principle for differentiating political stimuli. Multidimensional scaling is a tool that is ideally suited for generating precisely the kind of "mental map" that we are seeking in this research context.

Objectives, Utility, and Applications of MDS

A typical multidimensional scaling analysis begins with data on the pairwise dissimilarities (or similarities) among k stimulus objects. In the substantive example we are using, the "stimulus objects" are the political figures, so $k = 13$. MDS represents each stimulus as a point within m-dimensional space. In the example, m would be the number of distinct evaluative criteria that citizens use to differentiate the political figures. Typically, the dimensionality of the scaling solution (and, hence, the value of m) is not known a priori; instead, it is determined over the course of the analysis. The dissimilarities between the stimuli are represented by interpoint distances: specifically, greater dissimilarity between two objects is shown by larger distance between the two points representing those objects, and vice versa.[2] Analytically, the immediate objective of an MDS analysis is to find the most parsimonious—that is, the lowest-dimensioned—scaling solution that produces sufficiently close consistency between the input dissimilarities and the scaled interpoint distances.

The substantive objective of most multidimensional scaling analyses is to determine whether there is any interesting structure contained in the point configuration generated by the MDS routine. In other words, the analyst examines the placement of the k points in m-space to determine whether the relative positions of the points provide any insights about why pairs of stimuli are similar to, or different from, each other. So, in our substantive example, we would examine the scaled locations of the points representing the various political figures in order to determine whether the relative placements from the scaling solution provide insights about the criteria that citizens used to evaluate the figures. While m could be any value from one through $k - 1$, it is generally desirable to obtain an MDS solution in a space with few dimensions relative to the number of stimulus objects.[3]

Within the scaled point configuration, clusters of points may correspond to groups of stimuli that are distinct from each other (i.e., across clusters) in terms of their substantive characteristics. For example, the points representing the political figures might fall into distinct groups based upon partisanship. And, directions within the space may correspond to properties of the objects that vary in a more continuous manner. In this case, we want to see whether the scaled points are arrayed in a way that is consistent with ideological differences across the political figures. It is important to emphasize that any substantively meaningful clusters or directional arrays of points in the scaling solution emerge from the data and

[2]The only conceptual difference between dissimilarities data and similarities data is that, with the former, larger values correspond to more dissimilar pairs of stimulus objects. With similarities data, larger values indicate more similar object pairs. Most MDS software assumes that the input data are dissimilarities. This is not problematic at all because similarities can be reflected (i.e., multiplied by -1 or subtracted from a constant) to transform them into dissimilarities.

[3]While precise guidelines for determining dimensionality depend upon the details of the analysis (e.g., metric versus nonmetric MDS, a single dissimilarities matrix versus dissimilarities from multiple data sources, etc.), it will always be the case that increasing the number of dimensions improves the fit of the scaling solution to the data. In principle, the researcher in an exploratory analysis would try solutions in several dimensionalities, from one to some larger number, and then select the lowest-dimensioned solution that provides an adequate fit. In reality, two- and occasionally three-dimensional solutions are the norm, because they can be represented easily on a printed page or other two-dimensional display medium. More complete discussions about selecting the dimensionality for a multidimensional scaling solution can be found in Shepard (1974), Rabinowitz (1975), and Borg and Groenen (2005, chap. 3).

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the scaling model that is fitted to them—they are not specified in advance by the researcher. So, the multidimensional scaling solution provides insights about the evaluative standards that people actually employ when thinking about the political figures.

Multidimensional scaling is a useful tool for social research, and it is particularly relevant for political science precisely because many important theories of political phenomena can be represented in spatial terms (Brady 2011). Over the years, there have been quite a few applications within the discipline. For example, MDS has been used to examine candidate evaluations within the mass public (Rabinowitz 1978; Weisberg and Rusk 1970); the spatial theory of voting (Cahoon, Hinich, and Ordeshook 1978; Enelow and Hinich 1984); individual differences in ideological thinking (Jacoby 1986; Marcus, Tabb, and Sullivan 1974) comparative partisanship and party systems (Gross and Sigelman 1984; Katz 1979); congressional roll-call voting (Hoadley 1980); comparative political culture (Andrews and Inglehart 1979); individual choices among basic values (Schwartz 2007); and the dimensionality of survey questions (Jacoby 1996). So, even though multidimensional scaling is not really a "mainstream" component of political methodology, it certainly has been used as a tool to investigate a number of substantive problems.

An Application: The Public's Candidate Perceptions in 2004

Here, we use MDS to address the substantive question that was raised earlier: does the American public actively differentiate political stimuli along ideological lines? Specifically, the information from survey respondents' affective ratings of the 13 political figures will be used to calculate perceptual dissimilarities among pairs of those figures. These dissimilarities are then used as input to a nonmetric multidimensional scaling routine in order to locate a set of points representing the candidates within a space of (hopefully) low dimensionality.

We begin with a multivariate data matrix, $\mathbf{V}$, that contains information from the 2004 CPS American National Election Study (ANES). Specifically, $\mathbf{V}$ has 711 rows and 13 columns, containing 711 NES respondents' feeling-thermometer ratings of the 13 prominent political figures listed earlier. The line-of-sight (LOS) procedure developed by Rabinowitz (1976) is used to create a dissimilarities matrix, $\boldsymbol{\Delta}$, from the information in $\mathbf{V}$. LOS is based on the assumption that the respondents' thermometer scores are generated by a process that is consistent with the multidimensional unfolding model. Using the geometric implications of this model, LOS rank orders all pairs of the political figures according to their dissimilarity from least to most dissimilar.[4]

The $\boldsymbol{\Delta}$ matrix of LOS dissimilarities is shown in Table 1. The zeroes in the main diagonal of this symmetric matrix indicate that each figure is not at all dissimilar to itself. The off-diagonal cell entries in $\boldsymbol{\Delta}$ are mostly integers that range from one to the total number of pairs (which is $k(k-1)/2$, or 78 in this case); larger values indicate greater dissimilarity for the pair of figures corresponding to the row and column defining the cell. Noninteger values indicate tied dissimilarities. Of course, there is far too much information in this matrix to be comprehensible in its "raw" form. Hopefully, MDS can be used to produce a graphical representation of these data that can be interpreted in substantive terms.

Nonmetric multidimensional scaling of the LOS dissimilarities from the 2004 data produces the two-dimensional point configuration shown in Figure 1.[5] This solution fits the data perfectly, in the sense that the Spearman correlation between the rank-ordered distances (i.e., between pairs of points) in the figure and the corresponding rank-ordered dissimilarities in $\boldsymbol{\Delta}$ is a perfect 1.0.[6] Furthermore, the relative positions of the points seem to be very reasonable in substantive terms. Points representing Democratic figures form a cluster in the lower-left part of the space; the points for Republicans are closer to the right side; and the point for Ralph Nader (who ran as an independent in 2004) is far removed from both partisan clusters near the top of the space.

The scaling solution determines the *relative* locations of the points (i.e., with respect to each other), but not the

[4]The LOS procedure assumes that (1) thermometer ratings are inversely related to the distances between respondents' ideal points and the candidates' points within a space, and (2) ideal points are distributed throughout the space (i.e., there are no "empty" regions devoid of ideal points). Rabinowitz (1976) showed that the geometric implications of these assumptions can be used to infer the rank order of the distances between candidate points, while still taking into account the presence of measurement error and discrete, ordinal values in the feeling thermometers.

[5]This particular MDS solution was obtained by using the PROC MDS routine in SAS. Virtually identical results are obtained with other multidimensional scaling routines, including ALSCAL in SPSS, the `smacofSym()` function in R, and `mdsmat` in Stata.

[6]The fit for a multidimensional scaling solution is often reported with a *badness*-of-fit statistic called Kruskal's Stress. Here, the Stress_1 coefficient is 0.04, a value that is generally regarded as indicating an excellent solution (Rabinowitz 1975). Note that we also tried a one-dimensional MDS solution. While the unidimensional point configuration is reasonable in substantive terms, the goodness of fit is quite a bit worse than the two-dimensional case, with a Spearman correlation of 0.91, and $\text{Stress}_1 = 0.22$. Of course, there is no need to try MDS solutions in three or more dimensions since the two-dimensional configuration provides a nearly perfect fit to the dissimilarities.

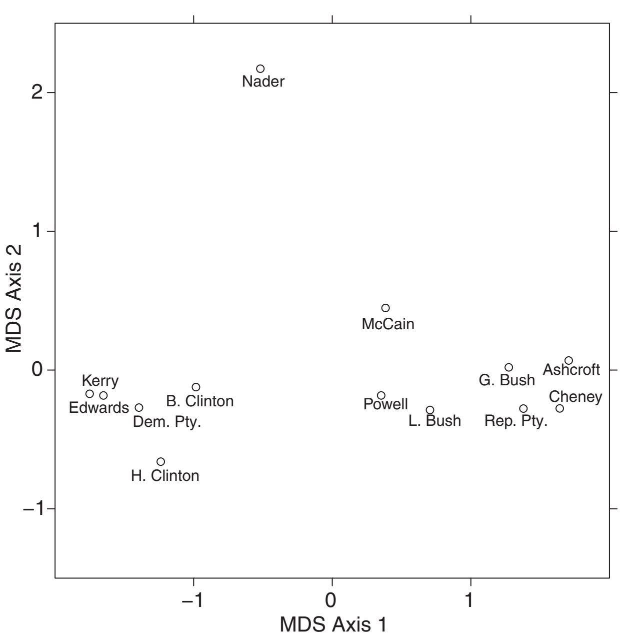
15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Matrix of Perceptual Dissimilarities among Political Figures from the 2004 Presidential Election

	G. W. Bush	John Kerry	Ralph Nader	Dick Cheney	John Edwards	Laura Bush	Hillary Clinton	Bill Clinton	Colin Powell	John Ashcroft	John McCain	Democ. Party	Repub. Party
G. W. Bush	0.0	73.0	62	8.0	68.0	20.0	51.5	41.0	24.0	7	25.5	50	5.0
John Kerry	73.0	0.0	56	78.0	1.0	54.0	15.0	17.0	47.0	77	37.0	2	74.5
Ralph Nader	62.0	56.0	0	72.0	59.0	53.0	60.0	49.0	58.0	70	39.0	57	71.0
Dick Cheney	8.0	78.0	72	0.0	74.5	25.5	65.0	51.5	29.0	12	30.0	66	4.0
John Edwards	68.0	1.0	59	74.5	0.0	44.0	14.0	16.0	46.0	76	38.0	3	69.0
Laura Bush	20.0	54.0	53	25.5	44.0	0.0	42.0	34.0	9.5	23	22.0	45	18.0
Hillary Clinton	51.5	15.0	60	65.0	14.0	42.0	0.0	19.0	32.0	67	40.0	13	55.0
Bill Clinton	41.0	17.0	49	51.5	16.0	34.0	19.0	0.0	31.0	61	36.0	11	48.0
Colin Powell	24.0	47.0	58	29.0	46.0	9.5	32.0	31.0	0.0	28	9.5	35	21.0
John Ashcroft	7.0	77.0	70	12.0	76.0	23.0	67.0	61.0	28.0	0	33.0	63	6.0
John McCain	25.5	37.0	39	30.0	38.0	22.0	40.0	36.0	9.5	33	0.0	43	27.0
Democ. Party	50.0	2.0	57	66.0	3.0	45.0	13.0	11.0	35.0	63	43.0	0	64.0
Repub. Party	5.0	74.5	71	4.0	69.0	18.0	55.0	48.0	21.0	6	27.0	64	0.0

Source: Line-of-sight procedure (Rabinowitz 1973, 1976) applied to feeling-thermometer ratings by 711 respondents from the 2004 CPS American National Election Study.

FIGURE 1 MDS Point Configuration for Political Figures from the 2004 Presidential Election



Note: Multidimensional scaling analysis of LOS proximities from Table 1. Stress_1 for this solution is 0.04, and the Spearman correlation between the dissimilarities and the distances is 1.00.

absolute position of the overall point cloud. In fact, this configuration has been rotated to a varimax solution. While this orientation is arbitrary from an analytic perspective, it does appear that the horizontal and vertical directions correspond to substantively interesting evaluative dimensions. Notice that relatively conservative figures like Ashcroft and Cheney are located farther to the right while the two most liberal candidates, Edwards and Kerry, are located to the left. Relatively moderate or nonpolitical figures (i.e., Powell, McCain, and Laura Bush) are closer to the center of the horizontal axis. Even Nader's horizontal coordinate falls in between the two partisan clusters, somewhat closer to the Democrats. Thus, it seems reasonable to interpret the horizontal axis as an ideological dimension.

The vertical direction in Figure 1 is a bit more ambiguous. It seems to represent something of a general "notoriety" characteristic or perhaps an insider-outsider distinction. The reasoning behind this interpretation is that 10 of the 13 points fall in a horizontal band located near the zero point on the vertical axis. The three points that fall outside this band represent candidates who were distinctive or unusual in 2004. The point for Hillary Clinton falls discernibly below the band, while the points for McCain and especially Nader fall above it. This suggests that people were thinking about these figures not only in terms of their ideologies but also by the degree to which they conform to the mainstream of the parties.

With respect to the substantive question motivating this example, the relative positions of the points in the horizontal direction within the multidimensional scaling solution seem to indicate the existence of an electorate that actively uses liberal-conservative ideology to differentiate political figures. However, it is a bit premature to conclude that this is actually the case. Even though the MDS point

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

configuration shows a very close fit to the data, there is no information about the stability of the scaling solution. As in any other statistical analysis, sampling error should be taken into account.

The matrix of dissimilarities shown in Table 1 and used in the multidimensional scaling analysis is based upon information from a single sample of 711 ANES respondents. Any other sample of the same size, drawn from the same population, probably would produce a different set of dissimilarities, leading to a different point configuration in the MDS solution. The results in Figure 1 might simply represent points drawn from extremely large sampling distributions. In that case, they might depict central tendencies in public opinion, but they would provide very little insight regarding the ways individual citizens view political figures. Just how much would we expect the scaling results to vary across repeated samples? And, to what extent does the point configuration in Figure 1 depict "real" differences among the political figures, rather than random noise?

Answers to the preceding questions are critical for evaluating the meaning and quality of the empirical results. If the point configuration in Figure 1 is highly unstable, then we could not place much stock in the relative positions of the respective points. As a consequence, we would be very hesitant to conclude that the electorate really does use ideology to think about political stimuli, despite the point estimates in Figure 1. The frustrating problem is that the analysis so far provides no insights one way or the other on this issue. So, we cannot really say that we have any answer to the substantive question that motivated the analysis in the first place.

Description Rather than Inference

The dilemma encountered at the end of the previous section certainly is not unique to our analysis of the public's political perceptions in 2004. Virtually all multidimensional scaling applications—in other disciplines as well as in political science—have employed scaling procedures to *describe* the proximity structure within a given dataset. There are few, if any, published attempts to assess the stability of MDS solutions in substantive applications. This omission is somewhat puzzling because inferential multidimensional scaling procedures do exist.

In fact, various inference strategies have been proposed. Maximum-likelihood MDS was developed for metric data by Ramsay (1977) and, separately, by Cahoon, Hinich, and Ordeshook (1978; also see Enelow and Hinich 1984, Appendix 9.1). Maximum-likelihood methods for nonmetric MDS were developed by Takane and his colleagues (e.g., Takane and Carroll 1981). Brady (1985) discusses statistical consistency and hypothesis testing in nonmetric MDS. More recently, Bayesian approaches for MDS of metric data have been proposed by Oh and Raftery (2001) and Bakker and Poole (2013). But, despite the existence of inferential methods, applications to substantive problems are almost entirely absent.[7]

There appear to be divided opinions among MDS researchers. On the one hand, some say that statistical inference is not particularly important. Young states, "Most often, the probabilistic (statistical) notion is irrelevant to the application (of MDS). . . . In these cases, MDS does just fine without the excess baggage introduced by inferential notions" (1984, 67). Others, however, emphasize the relevance of statistical inference: "For a long time, there has been a serious need for some error theory" (E. E. Roskam, quoted in Cox and Cox 2001, 108). Concerns have also been expressed about the stringency of the assumptions and the amount of data required for valid statistical inference with MDS (e.g., Weinberg, Carroll, and Cohen 1984; Young 1984).

Regardless of the reasons, inferential multidimensional scaling methods have just not caught on within the research community. One indicator of this (as well as a likely causal agent) is the fact that, while all of the major statistical software packages contain MDS routines, none of them have any provision for inference or stability assessment in their solutions. The net result is that multidimensional scaling is widely regarded (at least in political science) as a strictly exploratory procedure, with little relevance or immediate utility for theory testing (e.g., Bartels and Brady 1993; Brady 1990).

Resampling Strategies and MDS

Alternative approaches to inference in MDS, based upon resampling strategies, have made useful, but limited, progress. The most prominent resampling methodology is the bootstrap, in which the analyst samples randomly (with replacement) from the observed data matrix in order to estimate the probability distribution for some statistic of interest (Efron and Tibshirani 1993). Several

[7]The work by Cahoon, Hinich, and Ordeshook (1978) and Enelow and Hinich (1984) is something of an exception to this generalization, since their MDS methodology was specifically developed to test the spatial theory of voting. However, even these works emphasize the descriptive aspects of the scaling results. Apart from noting that the estimates of survey respondents' ideal points are heavily affected by error due to small degrees of freedom, the authors do not discuss sampling variability in the scaled point locations of their stimuli (presidential candidates).

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

researchers have used the bootstrap to assess the stability of points in scaling solutions obtained for multidimensional scaling models that assume several sets of dissimilarities as input data. Weinberg, Carroll, and Cohen (1984) apply bootstrap resampling to the weighted multidimensional scaling model, whereas Winsberg and De Soete (2002) use the bootstrap to assess stability in a latent-class MDS model. In both of these cases, the repeated samples are drawn from the data sources (e.g., subjects in an experiment or respondents on a survey). However, each data source provides a complete set of dissimilarities for the full set of $k(k-1)/2$ distinct stimulus pairs.

Unfortunately, the bootstrap cannot be applied directly to the most common multidimensional scaling situation, in which the researcher has one matrix of two-way, one-mode data. This is sometimes called the "classical multidimensional scaling" model, or CMDS (Schiffman, Reynolds, and Young 1981). It uses a single matrix of dissimilarities among all pairs of stimuli to produce a single configuration of points representing those stimuli. Each cell of the dissimilarities matrix contains information on a specific and distinct pair of stimulus objects. There is no way to "resample" that particular pairwise dissimilarity. Hence, in a CMDS analysis, it is impossible to obtain the randomly varying replications of the original dissimilarities matrix that are required for the bootstrap.

Taking a somewhat different approach, De Leeuw and Meulman (1986) use the jackknife to evaluate the stability of a multidimensional scaling solution across repeated trials wherein one stimulus is successively deleted from the estimation procedure. That is, the initial scaling of k objects is followed by k additional MDS estimates, each of which is carried out on only $k-1$ objects, omitting object i, with $i = 1, 2, \ldots, k$. However, De Leeuw and Meulman explicitly state that their approach is *not* an attempt to perform statistical inference on an MDS solution. Instead, it only assesses the stability of the scaled points' relative locations across perturbations of the dissimilarities matrix. The problem is that the relatively small number of systematically resampled dissimilarities matrices does not provide sufficiently reliable estimation of the sampling variability in the MDS point locations.

An Alternative Resampling Scheme for MDS

So far, resampling approaches have not been regarded as general strategies that should be applied in a routine manner to MDS solutions in order to make valid general statements based upon the observed data. The apparent hesitation about adopting resampling methodology for this broader purpose probably lies in practical questions about precisely what is the empirical distribution of data values that is resampled. Previous studies have treated the pairwise dissimilarities as the fundamental data of interest. In a typical MDS application, however, the number of stimulus objects and/or distinct data sources is limited. And, this causes problems for resampling, such as too little variability in a bootstrap sample drawn from multiple data sources, too few stimuli within a jackknifed sample, and the impossibility of using the bootstrap in the CMDS case.

In contrast to the preceding stumbling blocks, we believe that political science applications of multidimensional scaling may be particularly amenable to the use of resampling as an inferential tool, using a slightly different approach from those employed in earlier work. It is almost always the case that the dissimilarities between pairs of objects are *not* obtained through direct empirical observation. Instead, the pairwise dissimilarities are calculated from other empirical information about the stimuli. That is true for all but one of the political science applications cited earlier.[8]

For example, Weisberg and Rusk (1970), Andrews and Inglehart (1979), and Schwartz (2007) used correlations of survey respondents' evaluations across pairs of stimuli as measures of similarity between those stimuli. While the other studies used different similarity/dissimilarity measures, they all share a common characteristic: in every case, empirical observations on the individual stimulus objects are somehow "aggregated" in a pairwise manner to produce a single similarity/dissimilarity value for each pair of stimuli. It is the created dissimilarities that are input to a multidimensional scaling routine for analysis, rather than the original data on the respective stimulus objects. That suggests a strategy for applying the bootstrap:

1) Resample from the original multivariate data and create bootstrap replications of the complete dissimilarities data.
2) Perform MDS on each of the bootstrap replications.
3) Compare the replicated MDS solutions to the original MDS solution (i.e., based upon dissimilarities calculated from the full multivariate data) in order to assess stability directly.

[8]The exception is Marcus, Tabb, and Sullivan (1974), who have their experimental subjects rate the dissimilarity of pairs of political terms (e.g., "economic opportunity," "freedom of speech," "racial integration," etc.) on a scale ranging from 1 for "very similar" to 9 for "very dissimilar." In fact, their analysis estimates a weighted multidimensional scaling model from dissimilarities provided by multiple data sources (i.e., subjects). Therefore, the bootstrapping strategy employed by Weinberg, Carroll, and Cohen (1984) could be used to estimate sampling variability in this context.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The proposed strategy is analogous to a bootstrap procedure that Bollen and Stine (1992) developed for structural equation modeling (SEM). There, the original data are resampled to create bootstrap replications of the covariance matrix which is, in turn, used to assess the variability in the SEM parameter estimates. In both the MDS and SEM contexts, we cannot resample from the data matrix that is actually analyzed—the dissimilarities in the former case and the covariances in the latter—because the entries in that matrix are interdependent. So, the solution is to resample instead from the original data and construct replications of the full dissimilarities (or covariance) matrix.

A New Approach for Bootstrap MDS Confidence Regions

Let us assume that we are going to perform a classical multidimensional scaling analysis of two-way, one-mode data. Again, this produces a single configuration of points representing the stimuli. Building upon the nomenclature introduced earlier in the substantive example, the analysis begins with a two-way, two-mode, n by k matrix of multivariate data, $\mathbf{V}$. The n rows usually represent observations, and the k columns represent stimulus objects. The information in $\mathbf{V}$ is used to create $\boldsymbol{\Delta}$, a square matrix of order k containing the pairwise dissimilarities among the objects represented by the columns of $\mathbf{V}$. Cell entry δ_{ij} gives the dissimilarity between stimulus objects i and j, which were represented by columns i and j of $\mathbf{V}$.[9] Notice that we do not specify exactly how the δ_{ij} values are created from the i^{th} and j^{th} columns of $\mathbf{V}$. In fact, any available procedure for calculating dissimilarities can be used for this purpose.

An MDS is performed on the information in $\boldsymbol{\Delta}$, producing $\mathbf{X}$, a k by m matrix of coordinates locating k points (representing the stimulus objects) within m-dimensional space. In the MDS solution, the points are located such that, for any pair of points:

$$d_{ij} = \left[\sum_{l=1}^{m} (x_{il} - x_{jl})^2 \right]^{0.5} = f(\delta_{ij}) + \epsilon_{ij}$$

Where:

> d_{ij} is the distance between the points representing objects i and j within the m-dimensional space;[10]
>
> x_{il} and x_{jl} are the entries in $\mathbf{X}$ giving the coordinates for objects i and j, respectively, along the l^{th} dimension;
>
> δ_{ij} is the entry in $\boldsymbol{\Delta}$ giving the dissimilarity between objects i and j; and
>
> ϵ_{ij} is an error term.

The nature of the function, f, depends upon the researcher's assumption about the measurement level of the input dissimilarities. If the data are assumed to be measured at the interval or ratio level, then f is some parametric function, usually linear. If the data are regarded as ordinal, then f is some monotonic function, the precise shape of which is not known at the beginning of the analysis.

The error term, ϵ_{ij}, is assumed to be random but its distribution is not known. Within a specific MDS solution, the errors are not independent of each other. But, we will assume that separate replications of a specific ϵ_{ij} are independent.

A Bootstrap Procedure

The following bootstrapping procedure provides the data required to construct the confidence regions around the points:

1) Sample n rows, with replacement, from $\mathbf{V}$ to produce the first bootstrap replication of the multivariate data, $\mathbf{V}_1^*$.
2) Use $\mathbf{V}_1^*$ to produce the first bootstrap replication of the dissimilarities data, $\boldsymbol{\Delta}_1^*$.

[9] Most multidimensional scaling models assume that dissimilarities are symmetric, so that $\delta_{ij} = \delta_{ji}$. In the ideal situation, the $\boldsymbol{\Delta}$ matrix will contain the full set of $k(k-1)/2$ pairwise dissimilarities; in fact, the full matrix would contain each dissimilarity twice, once each in the upper and lower triangles. Most MDS software allows for missing cell entries in $\boldsymbol{\Delta}$. However, each missing δ_{ij} implies that there is one less constraint imposed on the relative positions of the points. Therefore, missing data do not preclude an MDS solution, but they do reduce its reliability. Note that "structural" missing data in Δ—for example, two stimulus objects that are never compared to each other—cannot be "fixed" through the resampling strategy proposed here. Nevertheless, the bootstrapping procedure does reveal the extent of the variability that exists in a scaling solution due to missing data and all other possible sources, such as sampling error.

[10] Here, we show this distance in simple Euclidean space. However, other types of distance can also occur in certain multidimensional scaling models. For example, nonmetric MDS permits Minkowski metrics, where the exponents on the distance formula can be other values than 2 and 0.5. And, individual difference models allow for distances within weighted Euclidean space. Nevertheless, the vast majority of MDS applications do rely on the simple Euclidean metric for constructing the point configuration.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3) Perform an MDS on $\boldsymbol{\Delta}_1^*$ to obtain the first bootstrap replication of the "raw" point coordinates, $\mathbf{Y}_1$.
4) Repeat Steps 1 through 3 for a large number (say, q) of bootstrap replications.

The bootstrapping process generates q different multidimensional scaling solutions. Now, MDS fits the interpoint distances to the input dissimilarities. But, the specific point coordinates that are used to generate these distances are arbitrary. The point configuration obtained in a CMDS can be reflected, rotated, dilated, and translated without affecting the fit to the input dissimilarities data. We need to find a transformation of each $\mathbf{Y}_p$ ($p = 1, 2, \ldots, q$) that brings it into maximum similarity (in the geometric sense) with the original $\mathbf{X}$.

For this purpose, we will use the Procrustean similarity transformation proposed by Schönemann and Carroll (1970). First, center the columns of $\mathbf{X}$ and $\mathbf{Y}_p$ so they sum to zero. Then calculate the product matrix, $\mathbf{X}'\mathbf{Y}_p$ and its singular value decomposition: $\mathbf{X}'\mathbf{Y}_p = \mathbf{A\Phi C}'$. Schönemann and Carroll (1970) show that:

The optimal rotation matrix is $\mathbf{T} = \mathbf{CA}'$.

The optimal dilation factor is $s = (\text{tr}\,\mathbf{X}'\mathbf{Y}_p\mathbf{T})/(\text{tr}\,\mathbf{Y}_p{}'\mathbf{Y}_p)$.

The optimal translation vector is $\mathbf{t} = n^{-1}(\mathbf{X} - s\mathbf{Y}_p\mathbf{T})'\mathbf{1}$.

Use the preceding results to obtain $\mathbf{Y}_p^* = s\mathbf{Y}_p\mathbf{T} + \mathbf{1t}'$, the optimally similar p^{th} bootstrap replication of the original MDS point coordinate matrix, $\mathbf{X}$.

Next, it is useful to rearrange (and relabel) the information from the bootstrapped MDS point configurations:

Extract the i^{th} row from each matrix, $\mathbf{Y}_p^*$, to form the m-element vector, $\mathbf{x}_{i,p}^*$, where $p = 1, 2, \ldots, q$.

Stack the $\mathbf{x}_{i,p}^*$ vectors atop each other to form $\mathbf{X}_i^*$, the q by m matrix of bootstrapped point coordinates for stimulus object i.

There will be k different $\mathbf{X}_i^*$ matrices, one for each of the stimulus objects in the MDS analysis.

Finally, we construct the $100(1-\alpha)\%$ confidence ellipsoid for the point representing stimulus object i (e.g., Johnson and Wichern 1998). To do so, we first calculate $\mathbf{S}_i$, the $m \times m$ variance-covariance matrix for the bootstrapped coordinates of i across the m dimensions. The $100(1-\alpha)\%$ confidence ellipsoid is defined by the set of points, $\mathbf{z}_j$ such that:

$$(\mathbf{z}_j - \mathbf{x}_i)\mathbf{S_i}^{-1}(\mathbf{z}_j - \mathbf{x}_i)' = \chi^2_{(m)}, \alpha,$$

where:

$\mathbf{x}_i$ is the row of coordinates for stimulus i from the original MDS point configuration, $\mathbf{X}$;

$\chi^2_{(m)}, \alpha$ is the α^{th} quantile of the chi-square distribution with m degrees of freedom.

The preceding steps are repeated to construct the confidence ellipsoid for each of the remaining $k-1$ points in the MDS solution.

Of course, if the multidimensional scaling solution under investigation has $m = 2$, the confidence ellipsoid becomes a regular two-dimensional ellipse, and it can be plotted directly. Otherwise, two (or, perhaps three) of the axes in the full MDS solution can be selected to construct confidence regions that represent a subspace of the full solution, but still can be easily depicted on a two-dimensional display medium. In any case, the resultant bootstrap confidence ellipsoid is interpreted like any other interval estimate of a population parameter. If the MDS were replicated on a large number of samples drawn randomly from the same population, the confidence ellipsoid for point i would contain the "true" relative position of stimulus i on $100(1-\alpha)\%$ of the replications. An alternative interpretation is that the confidence ellipsoid for stimulus point i is a region in the m-dimensional space that has a $(1-\alpha)$ probability of containing the "true" (relative) position of the stimulus point.

An Application: Confidence Regions for the Public's Candidate Perceptions

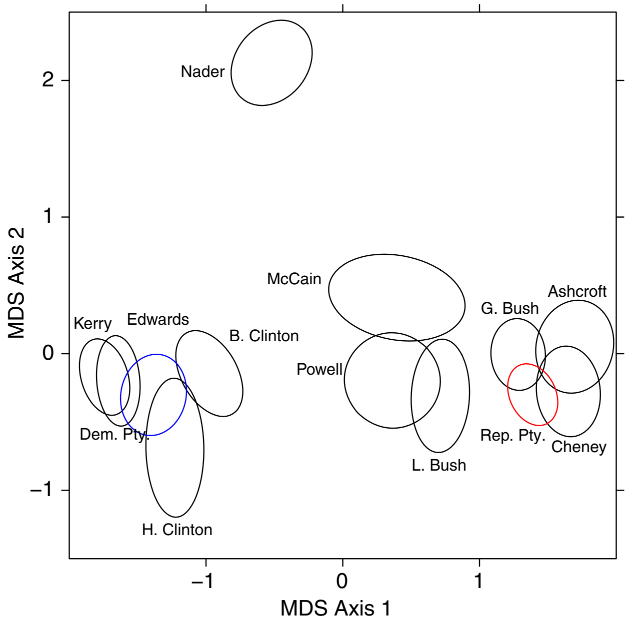
Here, we construct confidence regions for the point locations in the multidimensional scaling solution obtained from the LOS dissimilarities among the 13 political figures. The confidence regions will enable us to make statistical inferences about political perceptions within the population that generated the specific sample of respondents from the 2004 ANES used in the analysis. This goes far beyond the typical application of MDS, which is limited to simple description of proximity structure within a given dataset.

Recall that the matrix, $\mathbf{V}$, contains the ANES respondents' feeling-thermometer evaluations of the 13 political figures. We create 50 bootstrap replications of the original multivariate data, $\mathbf{V}_p^*$ ($p = 1, 2, \ldots, 50$), by sampling with replacement from the rows of $\mathbf{V}$.[11] Then, we apply the LOS procedure to each $\mathbf{V}_p^*$ to obtain the bootstrap

[11]The relatively small number of bootstrap replications is acceptable because we are only estimating the summary statistics for what appear to be well-behaved distributions of point coordinates (e.g., Booth and Sarkar 1998; Efron and Tibshirani 1993, 50–52).

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 2 Ninety-Five Percent Confidence Ellipses for MDS Solution



Note: Ellipses are constructed for 50 bootstrap replications of each point in the original MDS solution.

replications of the dissimilarities matrix, $\boldsymbol{\Delta}_p^*$. Next, we carry out the MDS on the 50 $\boldsymbol{\Delta}_p^*$ matrices, apply the Procrustean similarity transformation to each one, and rearrange the entries in the resultant $\mathbf{Y}_p^*$ bootstrapped coordinate matrices to obtain the $\mathbf{X}_i^*$ matrix for each of the 13 political figures. Finally, we calculate and plot the 95% confidence ellipse for each point in the MDS solution.[12]

Figure 2 shows the full configuration of confidence ellipses obtained from the bootstrap replications of the multidimensional scaling analysis. There are several interesting features in this figure. First, the ellipses all tend to be fairly compact, with each one occupying a relatively small proportion of the overall two-dimensional space. This shows that, even taking sampling variability into account, the public perceives differences among political figures. The spatial integrity of the point locations is particularly important, given the low levels of information and sophistication that are generally assumed to exist within the mass public. Despite these latter characteristics, people still differentiate among salient political figures in ways that are consistent with common understandings of liberal-conservative ideology.

Second, the ellipses form several distinct clusters within the space. As was the case with the point estimates, the Democratic stimuli form a cluster, and Ralph Nader falls at a widely separated position on his own. It appears, however, that the Republican stimuli actually fall into two different groups. George Bush, Cheney, Ashcroft, and the Republican party form one compact cluster near the right side of the space while Laura Bush, Colin Powell, and John McCain form another group closer to the center. Of course, this separation of point clusters anticipates the ideological bifurcation within the Republican party that has become very prominent in recent years; it is particularly telling that the separation of clusters remains very clear, even when we take the uncertainty in the candidate positions into account.

Third, the major axis for most of the ellipses is oriented in the vertical direction. This suggests that there was greater ambiguity in public beliefs about the viability or appropriateness of the figures than there was in perceptions of their ideological positions. There are two important exceptions to this general statement: John McCain's ellipse has its major axis oriented in a nearly horizontal direction, indicating greater public uncertainty about his ideology. And, Colin Powell's ellipse is nearly circular, showing that citizens had nearly equal variability in their beliefs about his viability and ideology.

Fourth, the degree of overlap between adjacent confidence ellipses reflects the degree to which the public differentiates between political figures. So, Kerry and Edwards, the two members of the Democratic presidential ticket, are viewed quite similarly. In contrast, there is very little intersection across the ellipses for George Bush and Richard Cheney; apparently, the public had a relatively clear view of their distinctiveness despite their connections in public office. It is also interesting that there is virtually no overlap between the ellipses for the two Clintons. This shows that Hillary Clinton definitely had developed a separate political identity of her own by 2004. And, Bill Clinton is largely distinct from the Democratic party, probably reflecting a more nonpartisan status that the public accords to an ex-president (although he is still clearly perceived as a liberal).

In terms of the substantive question motivating this MDS analysis, the 2004 electorate clearly differentiated political figures in a way that is consistent with the liberal-conservative continuum. If we regard the perpendicular

If we wanted to estimate the contours of the bivariate probability distribution for each point directly (e.g., using a convex hull), a much larger number of bootstrap replications would be necessary. Given the massive computing power available on even relatively small personal computers, it would certainly be possible to carry out such an analysis. But, if the coordinates are distributed bivariate normal (as they seem to be in this case), then the normal-theory ellipses provide identical information with much less effort.

[12] We did perform a test of multivariate normality and found no reason to assume that the bootstrap point distributions are markedly different from bivariate normal. A full report on the testing procedure is included in the online supplemental materials.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

projections from the ellipses onto the horizontal axis as interval estimates of the respective figures' ideologies, the ordering from the left side of the space to the right makes a great deal of sense. Just as we saw with the original point configuration in Figure 1, the ellipses fall in an array of sometimes overlapping positions from the most liberal candidates on the left (Kerry and Edwards) to the most conservative figures on the right (Cheney and Ashcroft). The widths of the intervals along the horizontal dimension formed by the ellipses generally are quite small, relative to the overall separation of the points. There are no instances where the ellipses are so wide that they extend from one side of the space to the other.

Even taking sampling variability into account, the data confirm that the public distinguished between liberal, centrist, and conservative stimuli in the political world. This is particularly striking because the ANES respondents were not instructed to use ideology (or any other specific criterion) when giving the feeling-thermometer ratings that were used to construct the pairwise dissimilarities between the candidates. In fact, to the extent that the public shows uncertainty about the candidate positions, it generally involves the viability dimension, rather than the axis that corresponds to liberal-conservative ideology.

To summarize, the basic multidimensional scaling analysis shows the structure of the public's candidate perceptions in 2004. This makes an important substantive contribution in itself because the spatial representation confirms that ideological polarization is a prominent feature in citizens' views of the political world. The bootstrap confidence regions obtained for the MDS solution actually make a separate contribution, by showing that the point locations for the various political figures are *not* vague central tendencies aggregated from large amounts of random noise. Instead, the American electorate was fairly precise in its perceptions of the candidates' and parties' liberal-conservative positions. This finding is rather surprising in view of the low sophistication levels and ideological naivete that are often attributed to the mass public.

Substantive Relevance of the Confidence Ellipses

The confidence regions obtained from the bootstrap procedure may have additional substantive implications that go beyond the empirical representation of model stability. The sampling error associated with each political figure is likely to be related to the public's uncertainty about that figure's political position relative to the other figures (e.g., Enelow and Hinich 1984). Such uncertainty may, in turn, affect the ways that citizens evaluate the respective figures. Previous research suggests that individual uncertainty about presidential candidates leads to more negative evaluations (e.g., Bartels 1986) and also decreases reliance on other judgmental criteria, such as ideology (e.g., Alvarez and Franklin 1993).

Here, we are conceptualizing uncertainty as a lack of clarity associated with a particular candidate. It is operationalized as the width of the confidence region relative to the ideology axis of the MDS solution. In order to see the consequences of this uncertainty, we propose a model in which it affects an individual's evaluation of a candidate, along with partisanship and ideology. The reduced form of the model is as follows:

$$\begin{aligned}\mathrm{Eval}_{ij} = \beta_0 + \beta_1 \mathrm{Copar}_{ij} + \beta_2 \mathrm{Dist}_{ij} + \beta_3 \mathrm{Width}_j \\ + \beta_4(\mathrm{Dist}_{ij} \times \mathrm{Width}_j) + \epsilon_{ij},\end{aligned} \quad (1)$$

where Eval_{ij} is person i's evaluation of political figure j; Copar_{ij} is a dichotomous indicator of copartisanship, coded 1 if i identifies with j's party and 0 otherwise; Dist_{ij} is the ideological distance between i and j; and Width_j is the width of the projection from j's confidence ellipse onto the horizontal axis in Figure 2.

All of the preceding variables can be operationalized with data from the 2004 ANES and the results from the MDS analysis. The dependent variable (Eval_{ij}) is just i's feeling-thermometer rating of j. For this analysis, the feeling thermometers are centered at the neutral point, so they range from -50 to 50. For purposes of the copartisanship variable, Democratic identifiers and leaners are copartisans with Kerry, Edwards, the Democratic party, Hillary Clinton, and Bill Clinton. Republican identifiers and leaners are copartisans with George W. Bush, Cheney, Ashcroft, McCain, Powell, the Republican party, and Laura Bush. Nonleaning independents are considered to be copartisans with Nader. In order to calculate ideological distance (Dist_{ij}), the point projections onto the horizontal axis of the MDS solution are rescaled to range from one to seven. Then, the distance term is the absolute difference between each figure's rescaled point coordinate and each respondent's ideological identification (also coded on a 1–7 scale). Finally, the width term (Width_j) is just the difference between the maximum and minimum projection from candidate j's confidence ellipse onto the rescaled horizontal axis.

Estimation of the model is complicated by the fact that each "observation" consists of a respondent-candidate pair. Thus, we need to take into account partially crossed random effects due to the two levels of nesting (i.e., candidates and respondents). A full report on

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Model Showing Individual Candidate Evaluations as a Function of Copartisanship, Ideological Distance, and Clarity in the Candidate's Ideological Position

	Coefficient Estimate	Standard Error
Copartisanship (Copar_{ij})	22.968*	0.478
Ideological distance (Dist_{ij})	−11.520*	2.392
Width of Confidence Ellipse (Width_j)	−10.131	7.125
Multiplicative term ($\text{Dist}_{ij} \times \text{Width}_j$)	8.291*	2.575
Intercept (β_0)	14.829*	6.704
Error variance (σ_ϵ)	20.502	
Number of observations	11,257	

Note: Table entries are maximum-likelihood estimates for the fixed-effect coefficients from a multilevel model with partially crossed random effects. Coefficients marked with an asterisk are statistically different from zero at the 0.05 level, in a two-sided test. Data sources are the 2004 ANES and the results from the multidimensional scaling analysis of the public's candidate perceptions. Each "observation" is a respondent-candidate pair. There are a total of 920 respondents and 13 candidates, although not all respondents evaluated all candidates and/or had usable data on all of the independent variables. The full multilevel model is discussed in a report included with the online supplemental materials.

the model is included in the online supplemental materials for this article. For present purposes, Table 2 presents the most relevant results, the maximum-likelihood estimates of the fixed-effect coefficients for the terms in equation (1).

Exactly as expected, copartisanship boosts a person's rating of a candidate; the effect is statistically significant and quite large, producing an average increase of almost 23 points on the feeling-thermometer scale, compared to ratings of candidates who are not in the respondent's own party. Similarly, ideological distance decreases the thermometer score, at least when there is no ambiguity about the candidate's liberal-conservative position (i.e., if a candidate's confidence region had zero width along the horizontal axis of Figure 2). For each unit increase in the separation between a person's ideological position and the candidate's position along the horizontal axis, there is an average decrease of 11.520 units in the thermometer rating.

These findings are not particularly surprising, given the well-known effects of partisanship and ideology on individual political evaluations. Instead, the really interesting results involve the impact of candidate-based uncertainty. Probably most important, the coefficient on the multiplicative term between distance and width is statistically significant and signed in the opposite direction from the coefficient for the distance term (i.e., it is positive, while the latter coefficient is negative). This shows that the net effect of ideological distance is attenuated when a political figure has a wider confidence ellipse. In substantive terms, this suggests that ambiguity in a candidate's liberal-conservative position makes it more difficult for citizens to use ideology as a standard for evaluation. Thus, the results are consistent with the assertion that uncertainty (i.e., wider confidence regions) reduces the impact of other judgmental criteria (i.e., the net effect of Dist_{ij} is smaller).

The coefficient on the width term is negative. This suggests that, all else equal, wider ellipses are associated with lower thermometer scores. However, this term is not statistically different from zero; therefore, it would be hazardous to suggest any particular substantive interpretation. Hence, the results do not provide any real support for the assertion that uncertainty, in itself, leads to more negative evaluations.

Of course, this brief analysis should not be taken as a conclusive investigation into the effects of uncertainty and ideological clarity on public evaluations of prominent political figures. It does demonstrate, however, that information about the stability of the multidimensional scaling solution has instrumental, as well as intrinsic, utility. Not only do the bootstrap confidence ellipses estimate the amount of sampling error and enable statistical inference, but they also reveal the relative precision with which the points are located relative to each other, a characteristic that may have important substantive implications of its own.

Cautions and Caveats

It is important to emphasize that bootstrapping is not a panacea for statistical inference in multidimensional scaling applications. There are at least five caveats that need to be kept in mind. First, the bootstrap approach for CMDS can only be used when the dissimilarities data are, themselves, constructed from multivariate data on the objects to be scaled. If the data consist of direct dissimilarity measures among the stimulus objects obtained from a single data source, then the resampling strategy we propose cannot be applied.

Second, it is important to make sure that the bootstrap replications are not affected by the various computational issues that can arise during the estimation of an

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MDS model, such as lack of convergence, local minima, and degenerate solutions (Davison 1984, chap. 5; Rabinowitz 1975). All three of the preceding problems can be diagnosed very easily. So, it is really just a matter of monitoring the relevant information across the resampling process. Any bootstrap replications on which the aforementioned problems occur should be discarded and replaced with additional replications of the MDS (Shepard 1974).

A third caveat is that the construction of the confidence ellipsoids assumes that the distribution of the coordinates from the bootstrap replications for each stimulus point is multivariate normal. The resampling procedure does not guarantee that this will be the case across the replicated MDS solutions. So, it is useful to test this distributional assumption in any application of the bootstrapping strategy. However, in our work so far, involving both simulated and real data, we have never found this assumption to be problematic.[13]

A fourth caveat involves the size of the multivariate data matrix, **V**, from which information is resampled. In order to obtain confidence regions for points representing, say, the k columns of **V**, we would draw repeated random samples of size n from the n rows of **V**. (Note that the roles of the n rows and k columns can be reversed.) But n must be fairly large in order to obtain reasonable estimates of the variance-covariance matrix for each point's coordinates. Chernick explains the problem that can occur when n is too small: "... with only a few values to select from, the bootstrap sample will underrepresent the true variability since observations are frequently repeated and bootstrap samples, themselves, can repeat" (2008, 173). In order to guard against this problem, Chernick recommends using bootstrap samples with $n \geq 50$ (2008, 174). While this limitation should not affect many political science applications (e.g., scaling dissimilarities based upon n survey respondents' evaluations of k political figures), it will be troublesome for analyses in which the objects to be scaled greatly outnumber the number of distinct pieces of information on those objects in the original multivariate data matrix (e.g., scaling a set of n nations based upon measures of k characteristics of those nations, where $k < 50$).

A fifth caveat involves missing values in the multivariate data matrix, **V**, from which the bootstrap replications are drawn. The literature on bootstrapping provides little guidance, because the general topic of missing data is either omitted from, or given rather cursory treatment in, the major texts (e.g., Chernick 2008; Efron and Tibshirani 1993). If there are empty cells (i.e., missing values) in **V**, then the amount of usable (i.e., nonmissing) data is likely to vary across the bootstrap replications, $\mathbf{V}_1^*$, $\mathbf{V}_2^*, \ldots, \mathbf{V}_q^*$. That, in turn, would cause fluctuating reliability in the bootstrap replications of the dissimilarities matrix, $\boldsymbol{\Delta}_1^*, \boldsymbol{\Delta}_2^*, \ldots, \boldsymbol{\Delta}_q^*$. The final outcome would be exaggerated variability in the bootstrapped point locations due to differences in the amount of information used to calculate the dissimilarities. Because the latter is an undesirable situation, we believe that the **V** matrix should contain no missing values in order to guarantee that the bootstrap replications are all drawn from the same source. Assuming that missing values do exist in the "raw" form of the data, then the researcher should use either listwise deletion or an imputation strategy to make sure that all the cells of **V** contain substantively meaningful entries.[14]

Despite these caveats and potential limitations, we remain very optimistic about the bootstrap MDS procedure. We have tested the methodology extensively, using many real and simulated datasets. In general, it works very well, and the results have always been highly satisfactory. While the procedure cannot be applied in every case, it does open up inferential possibilities for many situations in which multidimensional scaling could be a potential tool for testing political science theories.

Conclusions

In this article, we have laid out a bootstrap strategy for generating confidence regions in multidimensional scaling solutions. We are certainly not the first to recognize the relevance of resampling strategies like the bootstrap and the jackknife for MDS. Our approach, however, is quite different from earlier work in this area. Previous

[13]If the apparently unlikely situation does arise in which the bootstrapped replications depart markedly from a multivariate normal distribution, then a fully nonparametric confidence region could be constructed by taking the convex hull that just encloses the $100(1-\alpha)\%$ of the replicated points that fall closest to the point location in the MDS solution from the original, full, dataset. These nonparametric confidence regions would not have the nice geometric properties of ellipses, but they would still be completely suitable for assessing sampling variability in the scaled point locations. Note that a much larger number of bootstrap replications must be used to create the convex hull for each point in the scaling solution because we are estimating the sampling distribution itself, rather than merely the variance-covariance matrix of that distribution (Booth and Sarkar 1998).

[14]If a multiple imputation strategy is employed, then bootstrap replications could be drawn for each separate set of imputed values and combined to estimate the sampling distributions for the point locations in the MDS configuration. In that case, the confidence regions correctly would reflect the uncertainty due to the missing values, as well as that due to sampling error.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

researchers resampled the sources of complete dissimilarity matrices or took systematic subsamples from the full set of *k* stimulus objects. We resample from the data that are used to construct the dissimilarities, then create a complete dissimilarities matrix for each bootstrap replication. In this manner, we obtain independent replications of the data-generation process for the MDS. And that, of course, is exactly the kind of information that is needed for an empirical assessment of sampling variability in the multidimensional scaling solution.

Our primary motivation for developing this strategy is to assess sampling error in the MDS point configuration and thereby facilitate statistical inference from an observed scaling solution to unobservable population parameters. But, resampling is also useful for dealing with other aspects of a multidimensional scaling analysis. For example, bootstrapped MDS solutions could be used to investigate variability in scaled point locations due to a small number of stimulus objects relative to the number of dimensions; the effects of varying assumptions about the characteristics of the data being scaled or the precise MDS model that is being fitted to the dissimilarities; and changes in an MDS solution that might occur when stimulus objects are added to, or deleted from, the data being scaled. Thus, the bootstrap can be used to build upon Gifi's (1990, chap. 12) reminder that sampling variability is only one out of several types of stability that might be of interest in a scaling analysis. At the same time, the analysis in the preceding section shows that variability in the precision of the point estimates within the MDS solution may provide useful substantive information that can be exploited by the researcher.

The great advantage of our approach is its flexibility. As long as the analysis begins with multivariate data in which there is a sufficiently large number of observations with respect to the stimulus objects, the bootstrap strategy can be used. That is the case regardless of the type of dissimilarity measure, the assumptions about measurement levels in the data, the estimation procedure, and the precise multidimensional scaling model that is being fitted. The flexibility is particularly important because all of these characteristics can vary independently and in nontrivial ways from one MDS analysis to the next. Nevertheless, the bootstrap resampling strategy for obtaining the confidence ellipses could be applied to studies with any possible combination of these features.

Another advantage of our approach is its simplicity. The bootstrap MDS routine can be programmed very easily using tools available in any of the major statistical software systems. To make applications of this methodology even more convenient, we have written both a SAS macro and an R package to perform bootstrap resampling of a multidimensional scaling solution.[15] These are available on the authors' websites and the *AJPS* Dataverse. The R package can also be downloaded from CRAN. We hope that other researchers will find these tools useful for their own purposes.

In conclusion, the bootstrap approach provides a useful strategy for examining the stability of a multidimensional scaling solution. This is something that simply has not been considered in the vast majority of previous MDS applications. Our approach is relevant to a wide variety of specific research contexts. It can be used to enhance the utility of multidimensional scaling as a tool for testing substantive theories while still maintaining the flexibility in assumptions and analysis that make MDS so useful for exploring structure in data. In so doing, we hope that the bootstrap multidimensional scaling strategy will help to restore this family of methods to greater prominence in political science research.

References

Abramowitz, Alan I., and Kyle L. Saunders. 2008. "Is Polarization a Myth?" *Journal of Politics* 70(3): 542–55.

Alvarez, R. Michael, and Charles H. Franklin. 1993. "Uncertainty and Political Perceptions." *Journal of Politics* 56(3): 671–88.

Andrews, Frank M., and Ronald F. Inglehart. 1979. "The Structure of Subjective Well–Being in Nine Western Societies." *Social Indicators Research* 6(1): 73–90.

Bakker, Ryan, and Keith T. Poole. 2013. "Bayesian Metric Multidimensional Scaling." *Political Analysis* 21(1): 125–40.

Bartels, Larry M. 1986. "Issue Voting under Uncertainty: An Empirical Test." *American Journal of Political Science* 30(4): 709–28.

Bartels, Larry M., and Henry E. Brady 1993. "The State of Quantitative Political Methodology." In *Political Science: The State of the Discipline II*, ed. Ada R. Finifer. Washington, DC: American Political Science Association, 121–59.

Bollen, Kenneth A., and Robert A. Stine. 1992. "Bootstrapping Goodness–of–Fit Measures in Structural Equation Models." *Sociological Methods and Research* 21(2): 205–29.

Booth, James G., and Somnath Sarkar. 1998. "Monte Carlo Approximation of Bootstrap Variances." *American Statistician* 52(4): 354–57.

Borg, Ingwer, and Patrick Groenen 2005. *Modern Multidimensional Scaling: Theory and Applications.* 2nd ed. New York: Springer–Verlag.

[15]The SAS macro and the R function carry out the resampling scheme and organize the results from the bootstrap replications. In each case, the actual MDS is performed by a routine already included in the software, PROC MDS in SAS, and the `smacofSym()` function from the `smacof` package in R.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Brady, Henry E. 1985. "Statistical Consistency and Hypothesis Testing for Nonmetric Multidimensional Scaling." *Psychometrika* 50(4): 509–37.

Brady, Henry E. 1990. "Traits versus Issues: Factor versus Ideal–Point Analysis of Candidate Thermometer Ratings." *Political Analysis* 2(1): 97–129.

Brady, Henry E. 2011. "The Art of Political Science: Spatial Diagrams as Iconic and Revelatory." *Perspectives on Politics* 9(2): 311–31.

Cahoon, Lawrence, Melvin J. Hinich, and Peter C. Ordeshook. 1978. "A Statistical Multidimensional Scaling Method Based on the Spatial Theory of Voting." In *Graphical Representation of Multivariate Data*, ed. Peter C. C. Wang. New York: Academic Press, 243–78.

Chernick, Michael R. 2008. *Bootstrap Methods: A Guide for Practitioners and Researchers.* Hoboken, NJ: Wiley–Interscience.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press, 212–42.

Cox, Trevor F., and Michael A. A. Cox. 2001. *Multidimensional Scaling.* 2nd ed. New York: Chapman and Hall.

Davison, Mark L. 1984. *Multidimensional Scaling.* New York: Wiley.

De Leeuw, Jan, and Jacqueline Meulman. 1986. "A Special Jackknife for Multidimensional Scaling." *Journal of Classification* 3(1): 97–112.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction.* Cambridge: Cambridge University Press.

Fiorina, Morris P., Samuel A. Abrams, and Jeremy C. Pope. 2008. "Polarization in the American Public: Misconceptions and Misreadings." *Journal of Politics* 70(2): 556–60.

Gifi, Albert. 1990. *Nonlinear Multivariate Analysis.* Chichester, UK: Wiley.

Gross, Donald A., and Lee Sigelman. 1984. "Comparing Party Systems: A Multidimensional Approach." *Comparative Politics* 16(4): 463–80.

Hoadley, John F. 1980. "The Emergence of Political Parties in Congress, 1789–1803." *American Political Science Review* 74(3): 757–79.

Jacoby, William G. 1986. "Levels of Conceptualization and Reliance on the Liberal–Conservative Continuum." *Journal of Politics* 48(2): 423–32.

Jacoby, William G. 1996. "Testing the Effects of Paired Issue Statements on the Seven-Point Issue Scales." *Political Analysis* 5, 1993–1994. Ann Arbor: University of Michigan Press, 61–111.

Jacoby, William G. 2002. "Liberal–Conservative Thinking in the American Electorate." In *Research in Micropolitics: Political Decision Making, Participation, and Deliberation*, ed. Michael X. Delli Carpini, Leonie Huddy, and Robert Y. Shapiro. Greenwich, CT: JAI Press, 97–147.

Johnson, Richard A., and Dean W. Wichern. 1998. *Applied Multivariate Statistical Analysis.* 4th ed. Upper Saddle River, NJ: Prentice-Hall.

Katz, Richard S. 1979. "The Dimensionality of Party Identification: Cross-National Perspectives." *Comparative Politics* 11(2): 147–63.

Mann, Thomas E., and Norman J. Ornstein. 2012. *It's Even Worse Than It Looks: How the American Constitutional System Collided with the Politics of Extremism.* New York: Basic Books.

Marcus, George, David Tabb, and John L. Sullivan. 1974. "The Application of Individual Differences Scaling to the Measurement of Political Ideologies." *American Journal of Political Science* 18(2): 405–20.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches.* Cambridge, MA: MIT Press.

Oh, Man-Suk, and Adrian E. Raftery. 2001. "Bayesian Multidimensional Scaling and Choice of Dimension." *Journal of the American Statistical Association* 96(3): 1031–44.

Rabinowitz, George B. 1975. "An Introduction to Nonmetric Multidimensional Scaling." *American Journal of Political Science* 19(2): 343–90.

Rabinowitz, George B. 1976. "A Procedure for Ordering Object Pairs Consistent with the Multidimensional Unfolding Model." *Psychometrika* 41(3): 349–73.

Rabinowitz, George B. 1978. "On the Nature of Political Issues: Insights from a Spatial Analysis." *American Journal of Political Science* 22(4): 793–817.

Ramsay, J. O. 1977. "Maximum Likelihood Estimation in Multidimensional Scaling." *Psychometrika* 42(2): 241–66.

Schiffman, Susan S., M. Lance Reynolds, and Forrest W. Young. 1981. *Introduction to Multidimensional Scaling: Theory, Methods, and Applications.* New York: Academic Press.

Schönemann, Peter H., and R. M. Carroll. 1970. "Fitting One Matrix to Another under Choice of a Central Dilation and a Rigid Motion." *Psychometrika* 35(2): 245–56.

Schwartz, Shalom H. 2007. "Basic Personal Values: A Report to the National Election Studies Board." ANES Pilot Study Report nes011882.

Shepard, Roger N. 1974. "Representation of Structure in Similarity Data: Problems and Prospects." *Psychometrika* 39(4): 373–421.

Takane, Yoshio, and J. D. Carroll. 1981. "Nonmetric Maximum Likelihood Multidimensional Scaling from Directional Rankings of Similarities." *Psychometrika* 46(4): 389–405.

Weinberg, Sharon L., J. Douglas Carroll, and Harvey S. Cohen. 1984. "Confidence Regions for INDSCAL Using the Jackknife and Bootstrap Techniques." *Psychometrika* 49(4): 475–91.

Weisberg, Herbert F., and Jerrold G. Rusk. 1970. "Dimensions of Candidate Evaluation." *American Political Science Review* 64(4): 1167–85.

Winsberg, Suzanne, and Geert De Soete. 2002. "A Bootstrap Procedure for Mixture Models: Applied to Multidimensional Scaling Latent Class Models." *Applied Stochastic Models in Business and Industry* 18(4): 391–406.

Young, Forrest W. 1984. "Scaling." *Annual Review of Psychology* 35: 55–81.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

I: Assessing Multivariate Normality in Bootstrap Replications of the MDS Point Coordinates.
Figure S1: $\chi^2_{(2)}$ quantile plots for testing multivariate normality of bootstrap replications from multidimensional scaling analysis of 2004 public perceptions of political candidates.
II: A Multilevel Model Showing the Effects of Uncertainty on Individual Candidate Evaluations.
Figure S1: The effect of ideological distance on candidate evaluation, conditional on the horizontal width of a candidate's confidence ellipse.
Table S1: Model showing individual candidate evaluations as a function of copartisanship, ideological distance, and clarity in the candidate's ideological position.

15405907, 2014, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ajps.12056 by Consorci De Serveis Uni De, Wiley Online Library on [01/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License