# A Statistical Reduction of the United States Senate 

Kelan John Farrell-Smith, M.S., M.B.A., M.A<br>Claremont Graduate School


#### Abstract

"I would like to describe a field, in which little has been done, but in which an enormous amount can be done in principle. This field is not quite the same as the others... but... it might tell us much of great interest about the strange phenomena that occur in complex situations. Furthermore, a point that is most important is that it would have an enormous number of technical applications." - Richard Phillips Feynman

Caltech Engineering and Science, Volume 23:5, February 1960


#### Abstract

U.S. Senate roll call vote data were collected and converted into dichotomous numerical data for the years 1989-2013. 25 separate Factor Analyses were performed in SPSS: one on each year. Unconventionally, the senators were treated as the "variables" and the roll call votes were treated as the "subjects," such that the senators loaded onto extracted factors. The number of factors extracted from each year was plotted over time. The number of extracted factors decreased over time, indicating increased political partisanship and consolidation of political authority.


## Introduction

In an abstract mathematical sense, how many senators are there really? And does this number change over time based on certain political (or other) conditions? What I mean by "how many are there, really?" is, simply, that I imagine that there are underlying relationships between certain senators, or among certain groups of senators, that would seemingly not make it out of the question to clump those senators together as if they are, for the most part, one voting entity. This paper attempts to explore the nature of these "voting entities," and strives to uncover and explain their pattern of existence over time.

## Literature Review

Scholars too numerous to cite have used United States Senate roll call votes as their data sets in order to make quantitatively based assessments, which have, largely, dealt with senatorial partisanship. Many of these scholars have used mathematically straightforward, albeit labor-intensive, models in order to achieve reasonably robust measures of the increasingly partisan nature of the voting behavior of U.S. senators. Most of the strategies employed by academics have revolved around one scheme or another of calculating the percentage of time that senators votes "along party lines," particularly on "critical votes," a cumbersome (not to mention subjective) task that involves delving deep into the partisan complexities of each bill voted on, of each Senate trial adjudicated (and voted "Guilty" / "Not Guilty" on), of each presidential appointment confirmed by the Senate, and so forth, which is a research project in and of itself and is beyond the scope of this study.

This research is of a different form - it does not attempt to distinguish bills from one another or apply subjectively determined weighting schemes to individual roll call votes, in order to attain percentages-based results that support a priori hypotheses made by the researcher. On the contrary, because of the nature of this analysis (and the nature of "listwise" missing value deletion used in this analysis), it was optimal to weight each roll call vote equally. No votes were deemed "critical votes;" no bills were deemed "critical bills." The results of this research reveal not the tendencies of senators to vote along predetermined partisan lines, rather, they reveal the very party lines themselves and who falls into them, not to mention various "sub-parties," sub-groups, and cliques, which have been elucidated by extensively data-mining all roll call votes of the Unites States

Senate, and by taking advantage - creatively - of mathematical analysis and modern statistical reduction techniques typically reserved for use by clinical psychologists. "Factor analysis need not be strictly exploratory or strictly confirmatory," (Cudeck \& MacCullum, 2007, 58), and as such, many of the hypotheses (and conclusions) for this study were derived from the statistical analysis in a largely a posteriori fashion, despite the fact that some rigid adherents to the scientific method may find this fact distasteful.

To introduce the reader more clearly to the subject matter: "The basic idea of factor analysis is the following. For a given set of manifest response variables $x_{1}, \ldots, x_{p}$ one wants to find a set of underlying latent factors $\xi_{1}, \ldots, \xi_{k}$, fewer in number than the observed variables. These latent factors are supposed to account for the intercorrelations of the response variables in the sense that when the factors are partialed out from the observed variables, there should no longer remain any correlations between these. If both the observed response variables and the latent factors are measured in deviations from their means, this leads to the linear factor analysis model... the objective of factor analysis is to determine the number of factors $k$ and estimate the factor loadings $\left(\lambda_{i j}\right)$," (Cudeck \& MacCullum, 2007, 48). For the purposes of this research study, I have chosen to treat - in a somewhat radical and counterintuitive fashion - senators as the "variables" to be dimension reduced - that is, something like "grouped" - onto "underlying latent factors," while I have chosen to treat the individual roll call votes themselves as the "subjects." In other words, the each roll call vote can be thought of as having registered a senator, rather than each senator as having registered a roll call vote for the purposes of this research. This does not violate any mathematical or statistical (or logical) assumptions.

An important point of discussion is that there is not current scholarly consensus on whether or not it is mathematically acceptable to use Factor Analysis techniques on dichotomously scored variables. Some scholars, in fact, are fervently wary of this practice: Klaus Kubinger points out in his academic work, "This paper serves to remind the reader, that factor analysis in the case of dichotomous variables will often lead to artificial factors," (Kubinger, 2003, 106), and that, "[a]ccording to Fergerson (1941), Smith (1950), Stouffer, Guttman, Suchman, and Lazarfeld (1950), and Guttman (1955), the evidence already existed some years ago that by using dichotomous items... factor analyses will most likely result in as many factors as there are items," (Kubinger, 2003, 107). He goes on to remark - rather caustically towards statistical neophytes and novices that, "the aim of [his] paper is not to disclose any conditions when concurrent algorithms are superior; the purpose of this paper is simply to prevent [social scientists] who do not have a thorough knowledge of statistics from using conventional SPSS-default factor analysis for dichotomous variables," (Kubinger, 2003, 109). Conversely, other scholars take the position opposite Kubinger. Nearly thirty years prior to Kubinger's remarks, the article titled "An Argument in Support of Ordinary Factor Analysis of Dichotomous Variables," (Percy, et. al, 1976) showed that performing a Factor Analysis on a dichotomously-scored variables yields the same statistical results as performing a Factor Analysis on variables scored in other ways. They offer a very cogent "argument... for the use of ordinary factor analysis in dichotomous variables by comparing the factor solutions generated from 2-point Likert scale and 5-point Likert scale representation of the same item," (Percy, et. al, 1976, 143), and go on to explicitly state that, "one is not losing information... by employing dichotomous scaling," (Percy et al., 1976, 144). So,
despite the fact that academic consensus has not been reached on this issue, this research study operated under the assumption that it is absolutely fine to run a Factor Analysis on dichotomously-scored variables, and furthermore, because the results section of this study "worked out" - that is, the results of this study made intelligible sense, one example of this being that Republican senators tended to load together onto the same factors, as did Democrats - this study refutes Kubinger's claims that the Statistical Program for the Social Sciences (SPSS) cannot handle dichotomously scored variables in the context of a regular Factor Analysis - on the contrary, the results of the Factor Analyses performed in this research study exceeded expectations and created little (if any) technical problems. Mr. Kubinger, seems to be, simply, flat-out wrong.

There is no unanimously agreed-upon procedure for factor analyzing dichotomous data. The majority of authors of statistical textbooks seem to simply suggest that each researcher perform their due diligence when applying this statistical procedure in an unusual way, insofar as they should interpret their results conservatively and acknowledge the fact that they are playing with - and mining - data in an unconventional way. Yet boldness is part of new science. Why not flip a dichotomous-variable-based factor analysis on its axis and treat what would typically be understood as the "variables" instead as the "subjects" and treat what would typically be thought of as the "subjects" instead as the "variables," so that we may understand the traditionally conceived of "subjects" in a more reduced way? There is no reason this is not procedurally allowed, despite the fact that it has not been tried before.

Whereas "voting blocs" relate more to how senators cluster on one individual roll call vote or to alliances that last for fairly short periods of time over a few roll call votes,
the type of factor analysis employed in this study reveals a more long-term conceptualization of a voting blocs, by clustering senators based on a full year's worth of roll call votes - something that would be impossible to do with a simple visual inspection. Quantitative approaches to examining voting blocs in Congress based on roll call votes go back at least to James L. Sundquist's Dynamics of the Party System (1953) [See Table 6-2 and Table 7-1 in that text]. There are examples of quantitative analyses of Senate roll call votes in chapter 7 of Dodd and Oppenheimer's Congress Reconsidered (2013) [See Fig. 7-4 and 7-5 in that text], as well as in many other scholarly places. Yet, no one prior to this study - has applied Factor Analysis to senatorial voting by treating senators as variables to be loaded onto latent extracted factors.

In this unconventional kind of Factor Analysis - that is, one that treats people as "variables" and response scores as "subjects" - the groupings revealed by the extracted factors can be considered to amount to something akin to "stereotyped groups of people." I do not recommend this statistical procedure for use on the members of the general population or in psychological contexts because of this fact. However, if used strictly to group United States senators by roll call vote, then it seems less like "stereotyping senators," and instead seems more like "analyzing the voting behavior of elites and grouping them into clusters," which has a more scientific (and politically correct) ring to it.

## Method

## Data

Dichotomous (in this case "Yea" / "Nay") data were collected from the official United States Senate website for all United States Senate roll call votes from the years 1989 through 2013, as well as for 2014 (through April $15^{\text {th }}$ ), which were all years available [refer to Chart 1 in the appendix section of this paper for confirmation that this was all available years]. Votes on this website were listed in "Yea" / "Nay" form (and, occasionally in "Guilty" / "Not Guilty" form); the U.S. government does not provide these votes in spreadsheet form [refer to Chart 2 in the Appendix section]. As a result, these data had to be converted by hand into zeros and ones ("Yea" $=1$, "Nay" $=0$ ), so that they would be in a usable form for subsequent mathematical analysis. Votes where senators registered a "No Vote" were left blank and treated as "missing values" - this will be discussed more in the analysis section of this paper. A total of 8,198 roll call votes occurred in the U.S. Senate between the years 1989 and 2013 [see Figure 1, below], amounting to the conversion, by hand, of 819,800 individual roll call votes [refer to Chart 3 in the appendix section for a sample of a Microsoft Excel data set spreadsheet produced for this study]. Typically, with a dataset this large, a second, identical, data set would be created, and the two would be statistically checked against one another in order to ensure accuracy. Because of time constraints, an alternative method of ensuring accuracy in the data-entry phase of this study had to be sought out, because it was unfeasible to input the data twice (the entry of the first complete data set took approximately 500 hours of labor over a six-and-a-half week period from late-February 2014 until mid-April 2014). The solution to this issue was the creation of a "Yea Counter" in the Microsoft Excel
spreadsheet, which tabulated the zeros and ones entered for each roll call vote in a final column [refer to Chart 4 in the appendix for a visual example of the "Yea Counter"]. These "Yea Counters" were matched up with the numbers of Yeas listed for a vote, as listed on the official U.S. Senate website [refer to Chart 2 (once again) in the appendix for a visual of a Yeas listing on the Senate website].

Prior to converting the completed Microsoft Excel data sets to SPSS format for statistical analysis, an extensive data cleaning process was undertaken. Specifically, all votes where a senator registered a "present," a "present, giving live pair," or other rare vote besides a "Yea" or "Nay" were converted into "missing values," (and therefore treated as a "No Vote") in order to keep the types of responses to an intelligible minimum and eliminate noise. Thus, upon upload into SPSS, the data contained only three values, " 0, " " 1, " and "missing value."

Fig. 1: Number of Senate Roll Call Votes by Year, (1989-2013)

| RCV |  |
| :--- | :--- |
| Year | $\boldsymbol{n}$ |
| 1989 | 312 |
| 1990 | 326 |
| 1991 | 280 |
| 1992 | 270 |
| 1993 | 395 |
| 1994 | 329 |
| 1995 | 613 |
| 1996 | 306 |
| 1997 | 298 |
| 1998 | 314 |
| 1999 | 374 |
| 2000 | 298 |
| 2001 | 380 |
| 2002 | 253 |
| 2003 | 459 |
| 2004 | 216 |
| 2005 | 366 |
| 2006 | 279 |
| 2007 | 442 |
| 2008 | 215 |
| 2009 | 397 |
| 2010 | 299 |
| 2011 | 235 |
| 2012 | 251 |
| 2013 | 291 |

Number of U.S. Senate Roll Call Votes
Years 1989-2013


## Analysis

There are no independent or dependent variables in Factor Analysis. As mentioned before, rather than treat the senators as "subjects" and treat Senate roll call votes as "variables" to be analyzed and grouped onto latent factors, in this research design, I treated the roll call votes as "subjects" and the individual senators as "variables" to be loaded onto to latent extracted factors, in order to reduce the large number of Senators into a smaller number of "latent senator groupings." However, before Factor Analyses of each year could be performed, precautions had to be taken to address the problem of "missing values."

Because statistical programs cannot take "missing values" into account when performing Factor Analyses, deletion algorithms are often employed. "Listwise deletion" is a technique that that excludes any subject that has left a scoring item blank. Because roll call votes are treated as the subjects in this research, this means that if any senator failed to vote on a given roll call vote, then that particular roll call vote would be excluded from analysis for that year. This is considered to be an overly aggressive technique to handle "missing values" by most scholars, because it excludes quite a bit of data from analysis. However, more conservative techniques for dealing with missing values, such as "pairwise deletion," and coding the missing values "999" (a common practice), have their own theoretical drawbacks and limitations. Further, it was impossible to determine if a senator had chosen not to participate on a particular roll call vote as a means of abstention - giving their "No Vote" the feel of a "quasi-Nay" vote (which should perhaps be scored ".5" or a little bit less) - versus if a senator had not participated in a particular roll call vote simply because he was sick or out of town that
day (giving the "No Vote" the feel of a "missing value"). Because of these problems in determining the meaning of each "No Vote," it was decided that they should be treated as "missing values" in order to avoid interpretive error on the part of the researcher. As such, any roll call vote with a "missing value" was omitted from analysis by SPSS via listwise deletion. In years where a senator missed more than $20 \%$ of all roll call votes, rather than subjecting those senator's "missing values" on roll call votes to listwise deletion and eliminating a good deal of the entire year's votes, the individual senator was omitted from the year's analysis instead, in order to preserve more of the data. In this way, every possible attempt was made to reduce and randomize data omissions, and to preserve "subjects" and "variables."

Twenty-five separate, partially exploratory, partially confirmatory, Factor Analyses (or to be more mathematically specific, "Principal Component Analyses" due to the nature of the variables) were performed on all United States Senate roll call votes from 1989 through 2013. The "Kaiser Stopping Criterion" was used: factors with an eigenvalue of greater than 1 - that is, factors that controlled for more than $1 \%$ of the variance of the year's roll call voting - were extracted. Factors that did not meet this criterion were not extracted, and thus not considered to be groupings of senators; they were not considered to be "voting entities."

Additionally, data was collected and converted for the year 2014 (through April $15^{\text {th }}$ ), and a Factor Analysis was performed on that year as well, for reasons of posterity, however, the year 2014 was not included in final trend analyses because of its incomplete status.

After the Factor Analyses were performed on each year, the number of factors extracted (or more mathematically specifically, the number of "dimensions that the variables reduced to") from each year were plotted and analyzed over time using simple time-series regression analysis in Microsoft Excel. Two variants of this final trend analysis of the number of groups found in the U.S. Senate over time were created, and are discussed further (as well as pictured) in the final results section of this research paper.

## Factor Analyses (1989-2013, 2014*)

"There is no unique way to determine the number of factors. This is best done by the investigator who knows what the variables are supposed to measure. Then the number of factors can be specified at least tentatively," (Cudeck \& MacCullum, 2007, 49). The following year-by-year interpretations are, to an extent, objective; to an extent, subjective - this is the nature of Factor Analysis and Principal Component Analysis. It is not the goal of this research to attain $100 \%$ interpretive accuracy, rather it is simply to shed light on the extracted factors that exist in the Senate, and to perhaps, provoke follow-up studies and stimulate discussion (and probably disagreement) based on the analysis provided.

For detailed visual looks at each year's Factor Analysis, please refer to the supplemental "Tables" manual provided in conjunction with this paper. In the supplemental manual, "Missing Value Analysis" tables, "Extraction" tables, and "Total Variance Explained" tables are included for each year, as well as detailed, color-coded "Component Matrices" that show which senators loaded onto which factors, are included for each year. The senators themselves have also been color-coded red/blue if they were deemed to be partisan actors in a given year, and are color-coded gray if they were deemed to be impartial administrative actors, based on the how they loaded onto factors.

Next, the first of two results sections contains brief summaries of the Factor Analyses for each year, 1989-2013, as well as for the yet-to-be-completed year 2014. Again, the supplemental "Tables" manual should be consulted while reading this section if visual clarification is required, as it was impractical (and off-putting) to include its vast contents within this document.

Note to reader: In the ensuing year-by-year analysis, the terms "factor," "component," "group," and "dimension" are used virtually interchangeably - they mean the same thing, for all intents and purposes.

## 1989

In 1989, Senator Matsanuga (D-HI) was excluded from the analysis because he was absent for $29.3 \%$ of the roll call votes for the year. The remaining 99 senators loaded onto 15 factors with an eigenvalue greater than 1 . Of those factors, 7 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $35.05 \%$ and $19.91 \%$ of the variance in the year's voting, respectively. Factor 1 seems to be a "General Administrative Factor" [color-coded gray in the supplemental manual] - in other words, almost all of the senators for the year 1989, regardless of party affiliation, loaded onto the same factor, and no senator negatively loaded onto that factor. It is because of this loading pattern that the interpretation was made that Component 1 in 1989 was indeed a "General Administrative" or "General Senatorial" factor - it represents the group of senators that are, simply put, behaving like government administrators by putting general administrative concerns above partisan ones. Factor 2 seems to be a more specific kind of partisan liberal factor, with many Democrats loading positively onto the factor, and many Republicans loading negatively onto the factor. Factor 3 in 1989, which controls for
$4.53 \%$ of the variance of the years roll call voting, seems to be a bipartisan component, with a Democrat and two Republicans loading significantly positively onto it. It is interesting to note that the two Democratic Louisiana senators load significantly negatively onto Factor 6, a small, seemingly conservative factor. This is a bit of proof that not only did the factor analysis employed detect general administrative tendencies and party lines, but it also detected more subtle allegiances and grouping of senators. Further evidence of this exists with regards to Factor 4, which two Democrats without military backgrounds loaded significantly positive onto, as well as with regards to Factor 5, which two Democrats with military backgrounds loaded significantly onto.

Overall, the majority of senators in 1989 loaded onto the General Administrative Factor, whereas only about $60 \%$ of the senators load onto the Partisan factor. Only two senators who loaded significantly onto the partisan factor failed to load onto the General Administrative factor. Unfortunately, as this paper will show, this will not be the case in most of the subsequent years, as the Partisan factor begins to explain the majority of variance of each year's roll call voting in the early 1990s, and takes over - nearly entirely - beginning in 1995 with the Republican Revolution.

## 1990

Senators Wilson (R-CA) and Matsanuga were scrubbed in 1990, because they missed $22.4 \%$ and $23.6 \%$ of the year's roll call votes respectively. The remaining 98 senators loaded onto 16 factors with an eigenvalue greater than 1 . Of those factors, 6 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $29.48 \%$ and $24.45 \%$ of the variance in the year's voting, respectively. Factor 1 seems to, once again, be a "General Administrative Factor," however, not nearly as many senators loaded onto
the Administrative factor as did in 1989, indicating growing partisan divide in the Senate. Factor 2 seems to be a conservative factor, with many Republicans loading positively onto the factor, and many Democrats loading negatively onto the factor. The partisan factor controlled for more variance relative to the administrative factor than it did in 1989, indicating that senators were voting along party lines more often than they did the year prior, yet were still acting more as government officials than as partisan politicians. Interestingly, Senator Jeffords (I-VT) loaded, by himself, onto Factor 6, which controlled for $2.3 \%$ of the variance of the year's voting - this was deemed to be the "Independent" Factor in 1990.

## 1991

In 1991, Senators Pryor (D-AR), Harkin (D-IA), and Kerrey (D-NE) were not included in the Factor Analysis because they missed $56.1 \%, 24.6 \%$, and $22.5 \%$ of the year's roll call votes respectively. The remaining 97 senators loaded onto 14 factors with an eigenvalue greater than 1 . Of those factors, 6 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $37.44 \%$ and $21.38 \%$ of the variance in the year's voting, respectively. In stark contrast to the first two years of analysis, Factor 1 seems to be the Partisan factor, whereas Factor 2 seems to be the General Administrative factor. This indicates a switch in the mentality of the Senate as a whole. Whereas in 1989 and 1990 it could be argued that the entirety of the Senate was more focused on administrative business rather party politics, in 1991, this was clearly not the case. Furthermore, we see that many Republicans loaded both onto the Partisan factor as well as onto the Administrative factor, whereas most of the Democrats failed to load significantly onto the Administrative Factor but loaded heavily onto the Partisan Factor.

Because of this, it seems that the Democrats perhaps catalyzed, with their roll call voting, the switch from an administratively focused Senate to a partisan one. However, this change was likely solidified in 1995 by the Republicans, as this analysis will soon show.

## 1992

Five senators were left out of the Factor Analysis in 1992 due to missing more than $20 \%$ of the year's votes. It was a presidential election year, and as such, Senators Wirth (D-CO), Harkin (D-IA), Helms (R-NC), Burdick (D-ND), and Gore (D-TN) missed $20.7 \%, 25.9 \%, 37.4 \%, 24.1 \%$, and $44.1 \%$ of the year's roll call votes respectively. The remaining 95 senators loaded onto 15 factors with an eigenvalue greater than 1 . Of those factors, 10 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $34.36 \%$ and $21.36 \%$ of the variance in the year's voting, respectively. Factor 1 was once again the Partisan factor, and Factor 2 was once again the General Administrative Factor. There were several partisan liberal factors, most interestingly, a state-line liberal factor that the two Nevada senators (and no one else) loaded onto. The state-line Nevada Factor accounted for nearly $4 \%$ of the variance of the year's voting, and as such, the two Democratic Nevada senators (Reid and Bryan) accounted for about twice the amount of variance that they probably should have. Once again, in 1992, most of the Democrats loaded primarily onto the Partisan Factor without loading onto the General Administrative factor. About half of the Republicans loaded solely onto the General Administrative factor, whereas about half of them loaded significantly negatively onto the Partisan factor and loaded onto the General administrative factor. Generally speaking, in 1992, the Democrats were acting together like partisans, and the Republicans were acting together like administrators.

## 1993

No senators missed more than $20 \%$ of roll call votes; therefore, no senators were excluded from the Factor Analysis performed on 1993. The 100 senators loaded onto 16 factors with an eigenvalue greater than 1 . Of those factors, 7 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $38.34 \%$ and $19.88 \%$ of the variance in the year's voting, respectively. Factor 1 represented a Partisan (liberal) Factor, and Factor 2 was once again the General Administrative factor. In 1993, nearly every Senator loaded heavily (positively or negatively) onto the Partisan Factor, whereas only about half of them loaded onto the General Administrative factor. Of the senators that loaded onto the General Administrative Factor, about half were Democrats and about half were Republicans. Independent Senator Jeffords loaded heavily on Factor 4, a conservative factor that accounted for about 4\% of the variance of the year's votes. Senator Campbell (R-CO) loaded positively onto the Partisan factor, despite all other Republicans loading significantly negatively onto it. Many senators voted in a partisan fashion in 1993, yet many still voted and behaved as government administrators.

## 1994

In 1994, no Senators missed more than $20 \%$ of roll call votes; therefore, once again no senators were excluded from the Factor Analysis. The 100 senators loaded onto 17 factors with an eigenvalue greater than 1 . Of those factors, 8 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $28.91 \%$ and $23.03 \%$ of the variance in the year's voting, respectively. Factor 1 represented a Partisan liberal factor, and Factor 2 was once again the General Administrative Factor. Interestingly, there was little overlap between the Partisan group and the Administrative group in 1994, indicating
that some senators were acting purely as partisans, while other, arguably more responsible senators were acting purely as proper administrators. Factor 6 was an "Independent" Factor, which only Senator Jeffords (I-VT) significantly loaded onto - this factor accounted for $2.7 \%$ of the variance in the year's voting, a fairly significant amount considering this "group" contained only one person.

## 1995

The Republican Revolution occurred in 1994, and as such, the Factor Analysis yielded far different results than in 1994. Senator Packwood missed $21.9 \%$ percent of the votes that year, and was omitted from the analysis. The 99 remaining senators loaded onto 8 factors with an eigenvalue greater than 1 . Of those factors, just 3 had an eigenvalue of greater than 2 . These values represent the greatest (negative) change from year-to-year in terms of the amount of groupings that the senators loaded onto in the entire study. In essence, there was a consolidation of power in 1995, as far fewer groups of senators accounted for far more of the total variance of the year's roll call votes than they had in years past. Extracted Factors 1 and 2 accounted for $52.98 \%$ and $15.91 \%$ of the variance in the year's voting, respectively. Factor 1 was a Partisan (conservative) Factor that almost every single Republican loaded heavily onto, and that almost every single Democrat loaded significantly negatively onto. About half of Republicans and Democrats loaded heavily onto the General Administrative Factor (Factor 2), however, this factor accounted for less than three times less of total year's voting variance than the Partisan Factor accounted for. What this means is this: 1995 was a markedly different year in the Senate than previous years. Partisanship had been on the rise from 1989-1994, and indeed, party line voting seems to have overtaken administrative good-sense
sometimes around 1992. However, in the early 1990's the Administrative Factor and the Partisan Factor seemed to be weighted somewhat equally, regardless of which one was on top in a given year. In 1995, all that changed, as partisanship crushed neutral administrative governance when the Republicans gained a supermajority in Congress. The number of groups of senators reduced from 8 factors with an eigenvalue over 2 in 1994, to 3 in 1995. That is, 8 different groups of senators accounted for at least $2 \%$ of the variance in roll call voting in 1994, whereas only 3 did in 1995. Increased partisanship in the Senate seems to rely on the ability to unilaterally pass legislation, and seems to result in a consolidation of the number of groups that have influence of any real value.

## 1996

In 1996, no senators missed more than $20 \%$ of roll call votes; therefore no senators were excluded from the Factor Analysis. The 100 senators loaded onto 12 factors with an eigenvalue greater than 1. Of those factors, 5 had an eigenvalue of greater than 2 . This was a slight rebound from the drastic decrease in the number of groups in 1995. Extracted Factors 1 and 2 accounted for $44.75 \%$ and $22.68 \%$ of the variance in the year's voting, respectively, indicating a partial return to administrative concerns over partisan ones on the parts of some senators. Factor 1 represented a Partisan (liberal) Factor, with most Democrats loading positively onto it and most Republicans loading negatively onto it; Factor 2 was once again the General Administrative Factor. Factor 3 seems to be a conservative factor, with senators Cohen (R-ME) and Simpson (R-WY) loading heavily onto it.

## 1997

No senators missed more than $20 \%$ of roll call votes in 1996 - no senators were omitted from the Factor Analysis. The 100 senators loaded onto 13 factors with an eigenvalue greater than 1 . Of those factors, just like 1996, 5 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $44.13 \%$ and $20.84 \%$ of the variance in the year's voting, which indicates that 1997 was not much different that 1996 in terms of the way the senators aligned themselves. Factor 1 represented a Partisan (conservative) Factor; Factor 2 represented a General Administrative Factor. Most Republican senators loaded significantly positively onto Factor 1; most Democratic Senators loaded significantly negatively onto Factor 1. About two thirds of the senators in 1997 loaded onto the General Administrative Factor, despite also loading onto the Partisan Factor, indicating that many senators behaved in a quasi-partisan, quasi administrative fashion in 1997.

## 1998

Senator Helms (R-NC) and Senator Glenn (D-OH) were excluded from the Factor Analysis of 1998, having missed $22.3 \%$ and $23.6 \%$ of the year's roll call votes respectively. The 98 remaining senators loaded onto 10 factors with an eigenvalue greater than 1. Of those factors, 5 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $\mathbf{4 2 . 5 1 \%}$ and $24.12 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (conservative) Factor. Nearly every Republican loaded heavily positively onto it, and nearly every Democrat loaded heavily negatively onto it. The majority of senators also loaded onto Factor 2, the Administrative Factor, in 1998, however the Administrative Factor did not account for as much variance of the year's roll
call voting as the Partisan Factor, and as such 1998 should be classified as another partisan year in the U.S. Senate. A side component of note was Factor 4, which Senator Jeffords (and a Republican) loaded onto. This Independent Factor accounted for 2.28\% of the variance of the year's roll call votes, indicating, once again, that Independent Senator Jeffords has a disproportionate amount of influence, considering he is just 1 of 100 members.

1999

Senator McCain (R-AZ), likely due to his presidential campaigning, missed $35.3 \%$ of the roll call votes in 1999, and was thus not included in the factor analysis for the year. The 99 remaining senators loaded onto 10 factors with an eigenvalue greater than 1 . Of those factors, only 3 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $58.13 \%$ and $16.72 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (conservative) Factor. Nearly all senators loaded heavily onto Factor 1, positively for Republicans and negatively for Democrats. Only about half of Senators loaded onto the General Administrative Factor. Presidential campaign years seem to drive the member of the Senate to vote more along party lines and less along general administrative lines than is typical.

## 2000

Senator McCain was once again absent for more than $20 \%$ of votes in 2000, as was another presidential contender, Senator Lieberman (D-CT), missing 22.5\% and $21.1 \%$ of all roll call votes respectively. They were not included in the Factor Analysis. The 98 remaining senators loaded onto 11 factors with an eigenvalue greater than 1 . Of those factors, 4 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted
for $46.27 \%$ and $27.92 \%$ of the variance in the year's voting respectively. Factor 1 was a once again a Partisan (conservative) Factor. Factor 2 was once again the General Administrative Factor. Nearly all senators loaded onto both factors, indicating some attempt to return to administrative business during the presidential election year. Interestingly, no senators loaded onto any other factors with an eigenvalue over 2 in 2000 besides Factors 1 and 2, which indicates that the party lines had been consolidated into two rigid groups. It appears that there were no within-party divisions in 2000-not on the Democratic side, and not on the Republican side.

## 2001

No senators missed more than $20 \%$ of roll call votes in 1996 - so, no senators were excluded from the Factor Analysis that year. The 100 senators loaded onto just 7 factors with an eigenvalue greater than 1 . Of those 7 factors, 3 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $49.15 \%$ and $25.86 \%$ of the variance in the year's roll call voting. Factor 1 was a Partisan (liberal) Factor, with nearly all Democratic senators loading positively onto it, and nearly all Republican senators loading negatively onto it. Factor 2 was a General Administrative Factor, with nearly all senators loading positively onto it. Although party-line voting seems to have predominated in terms of variance explained, because almost all senators loaded onto the Administrative Factor as well, 2001 seems to have been a year when the government was also focused on its administrative business, at least to a certain extent.

2002

In 2002, Senator Helms (R-NC) missed $62.5 \%$ of the roll call votes. He was therefore excluded from the Factor Analysis for the year. The 98 remaining senators
loaded onto 15 factors with an eigenvalue greater than 1 . Of those factors, 7 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $37.02 \%$ and $22.43 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (liberal) Factor nearly all Democrats loaded positively onto it, while nearly all Republicans loaded negatively onto it. Senator Jeffords loaded heavily Democratic onto Factor 1, revealing his "liberal-ness." The vast majority of senators from both parties loaded positively onto Factor 2, the General Administrative Factor. Factor 1 and Factor 2 had less of a gap between them than in many immediately preceding years - partisanship seems to have subsided a bit in 2002, perhaps in the wake of the $9 / 11$ tragedy, although partisan concerns continued to trump administrative concerns, as they had since the early 1990s.

## 2003

A number of senators set out on the campaign trail in 2003 - Senators Lieberman (D-CT), Kerry (D-MA), and Edwards (D-NC) missed over 20\% of the roll call votes, as well as did Senators Graham (D-FL) and Miller (D-GA). They were absent 53.8\%, $63.8 \%, 38.8 \%, 32.5 \%$, and $22.7 \%$ of the time, respectively. The remaining 98 senators loaded onto just 4 factors with an eigenvalue greater than 1 . Of those factors, 3 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $61.27 \%$ and $15.96 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (conservative) Factor. Nearly every Republican loaded heavily positively onto it, and nearly every Democrat loaded heavily negatively onto it. The majority of senators also loaded onto Factor 2, the Administrative Factor. Only Senator Nelson (D-NE) loaded significantly (negatively) onto Factor 3. Nearly all of the senators fell into the partisan group, whereas only about half of them fell into the General Administrative group as well. A return to
consolidated partisan politics in the campaign years seems to have disrupted the temporary return to administrative functioning in the period following 9/11. In fact, 2003 produced the fewest amount of extracted factors of any year, indicating that power was more concentrated in the hands of the party leaders than in any other year, because there was a complete lack of party sub-factions in the Senate that year.

## 2004

2004 continued to be a presidential campaign year for Senators Kerry and Edwards, as both missed out on a majority of Senate roll call votes, with those two men missing $89.8 \%$ and $58.8 \%$ of the year's votes respectively. They were thus omitted from the Factor Analysis. The 98 remaining senators loaded onto 11 factors with an eigenvalue greater than 1. Of those factors, 6 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $49.65 \%$ and $25.16 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (conservative) Factor. Nearly every Republican loaded heavily positively onto it, and nearly every Democrat loaded heavily negatively onto it. Only about half of the senators loaded onto Factor 2, the Administrative Factor. This was a year of partisanship, with relatively little neutral administrative governing taking place. This seems to be the pattern in presidential election years.

## 2005

Senator Corzine (D-NJ) was the only senator excluded from analysis in 2005, as he missed $37.2 \%$ percent of the roll call votes. The 99 remaining senators loaded onto 9 factors with an eigenvalue greater than 1 . Of those factors, 4 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $47.49 \%$ and $24.27 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (conservative) Factor. Factor 2 was a

General Administrative Factor. Factor 3, which accounted for $2.55 \%$ of the years roll call vote variance, was a Republican group composed of Senator McCain and Senator Chafee (R-RI). Nearly all senators loaded onto the General Administrative Factor, which indicates a return to administrative function after the partisan presidential election year.

## 2006

Senator Rockefeller (D-WV) was the only senator excluded from analysis in 2006 - he was absent for $40.5 \%$ of the roll call votes for the year, and was not included in the Factor Analysis for 2006. The remaining 99 senators loaded onto 11 factors with an eigenvalue greater than 1 . Of those factors, 6 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $47.49 \%$ and $24.27 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (conservative) Factor that nearly everyone loaded onto positively or negatively along party lines. Factor 2 was a General Administrative Factor. About 60\% of all senators loaded positively onto this factor. 2006 seems to be as administrative a year as possible, considering it took place in an extremely partisan era.

## 2007

Numerous senators were once again on the presidential campaign trail in 2007, with 7 senators missing more than $20 \%$ of roll call votes for the year. Senators McCain (R-AZ), Dodd (D-CT), Biden (D-DE), Obama (D-IL), Brownback (R-KS), Clinton (DNY), and Johnson (D-SD) were absent for $55.9 \%, 37.6 \%, 38.9 \%, 37.6 \%, 30.5 \%, 23.3 \%$, and $70.4 \%$ percent respectively. The 93 remaining senators loaded onto 9 factors with an eigenvalue greater than 1 . Of those factors, 4 had an eigenvalue of greater than 2.

Extracted Factors 1 and 2 accounted for $44.93 \%$ and $18.13 \%$ of the variance in the year's
voting respectively. Factor 1 was a Partisan (liberal) Factor. Factor 2 was a General Administrative Factor. Much fewer senators loaded onto the General Administrative Factor than in the previous year, yet just as many loaded onto the Partisan Factor. This indicates a rise in partisanship heading into the 2008 presidential election year.

2008

President Barack Obama (63.7\%), Ted Kennedy (41.4\%), John McCain (80\%), Hillary Rodham Clinton (48.8\%), and Robert C. Byrd ( $25.1 \%$ ) all were absent for more than $20 \%$ of the roll call votes in 2008, and as such, they were excluded from the Factor Analysis. The 95 remaining senators loaded onto 11 factors with an eigenvalue greater than 1. Of those factors, 5 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $47.58 \%$ and $20.39 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (liberal) Factor, with the vast majority of Democrats loading positively onto it, and the vast majority of Republicans loading negatively onto it. Factor 2 was a General Administrative Factor, with about $75 \%$ of senators loading onto it. Factor 3, which accounted for $4.16 \%$ of the variance in roll call voting, seems to be a Bipartisan Factor. Factor 4 is a Partisan (liberal) factor composed of two Democrats Senator Tester (D-MT) and Senator Landrieu (D-LA) - accounting for $3.29 \%$ of the year's roll call voting variance.

## 2009

In 2009, three senators were excluded from the factor analysis: Kennedy (D-MA), Franken (D-MN), and Byrd (D-WV) because they missed more than $20 \%$ of the votes. They were absent $72.8 \%$ of the time, $54.7 \%$ of the time, and $40.1 \%$ of the time respectively. The remaining 97 senators loaded onto 7 factors with an eigenvalue greater
than 1. Of those factors, 4 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $57.99 \%$ and $13.45 \%$ of the variance in the year's voting respectively. Factor 1 represented a liberal Partisan Factor, with nearly all senators loading heavily onto it one way or another. Very few senators loaded onto Factor 2, as the partisan divide seems to have grown very wide in 2009, perhaps as a result of the Republican reaction to the election of Barack Obama. In fact, senators loaded far more heavily onto the Partisan Factor than they had in any previous year, so, not only have the majority of senators become partisan rather than administrative, they have become extremely partisan rather than administrative. Indeed, in 2009, the Partisan Factor accounted for almost five times as much variance of the year's roll call voting than did the General Administrative Factor!

## 2010

Senator Isakson (R-GA) and Senator Byrd (once again) were absent for more than $20 \%$ of the votes in 2010, and so, they were left out of the Factor Analysis for the year. They missed $20.4 \%$ of the votes and $44.5 \%$ of the votes respectively. The 98 remaining senators loaded onto 7 factors with an eigenvalue greater than 1 . Of those factors, 4 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $69.38 \%$ and $11.22 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (liberal) Factor. Nearly all senators loaded heavily positively over heavily negatively onto it. Barely any senators loaded onto the General Administrative Factor, meaning almost all legislation dealt with was of a highly partisan nature, and not much day-to-day government business was being taken care of by either party. The Partisan Factor accounted for 7 times as much variance in the year's voting as the Administrative Factor.

This is a far cry from the structure of the initial year in the analysis, 1989, when the Administrative Factor dominated the Senate scene.

## 2011

No senators missed more than $20 \%$ of the roll call votes in 2011, and as such, none were excluded from the Factor Analysis. The 100 senators loaded onto 9 factors with an eigenvalue greater than 1 . Of those factors, 6 had an eigenvalue of greater than 2 . Extracted Factors 1 and 2 accounted for $45.20 \%$ and $26.82 \%$ of the variance in the year's voting respectively. Factor 1 represented a liberal Partisan Factor, with nearly all senators loading heavily onto it positively or negatively along party lines, albeit less heavily so than in 2010. Factor 2 represents the Administrative grouping, with about $80 \%$ of senators loading onto this factor. Factor 3 was a Partisan (conservative) Factor consisting of Senators Paul (R-KY) and DeMint (R-SC). This seems to be a Tea Party Factor. There are also two minor liberal factors, one made up of 3 Democratic Senators, and one made up of 2 Democratic senators. Determining these small allegiances, and small allegiances like these, should be the subject of future research in this area. For now, they are available for examination by interested persons in the supplemental Tables manual that accompanies this study.

## 2012

Senator Kirk was absent for $100 \%$ of the Senate roll call votes in 2012 - he basically excluded himself from the Factor Analysis. The remaining 99 senators loaded onto 11 factors with an eigenvalue greater than 1 . Of those factors, 4 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $53.23 \%$ and $14.67 \%$ of the variance in the year's voting respectively. Factor 1 was a Partisan (liberal) Factor that
nearly everyone loaded onto, positively or negatively. Factor 2 was a General Administrative Factor. Hardly any senators loaded onto the General Administrative Factor in this year, indicating a return to partisanship that has been typical of presidential election years.

## 2013

In 2013, Senator Lautenberg (D-NJ) was excluded from the Factor Analysis because he was absent for $42.3 \%$ of the roll call votes for the year. The remaining 99 senators loaded onto 7 factors with an eigenvalue greater than 1 . Of those factors, 4 had an eigenvalue of greater than 2. Extracted Factors 1 and 2 accounted for $65.34 \%$ and $12.59 \%$ of the variance in the year's voting respectively. Factor 1 was again a Partisan (liberal) Factor that nearly everyone loaded onto one way or another. Very few senators loaded onto Factor 2, the Administrative Factor, indicating that partisanship was high, and little general government legislation was being dealt with. The Republicans hate Obamacare, and the Democrats hate that the Republicans hate Obamacare - their partisan divide and unwillingness to work together is clearly revealed by the Factor Analysis of the U.S. Senate roll call voting from 2013.

## 2014 (*through April 15 ${ }^{\text {th }}$ )

Through the first 107 votes of this year, Senator Coburn (R-OK) and Senator Rockefeller have been absent more than $20 \%$ of the time, missing $31 \%$ and $25.9 \%$ of all roll call votes respectively. They were therefore excluded from the Factor Analysis performed on this year's votes. The 98 remaining senators loaded onto 11 factors with an eigenvalue greater than 1 . Of those factors, 7 had an eigenvalue of greater than 2.

Extracted Factors 1 and 2 accounted for $48.24 \%$ and $26.48 \%$ of the variance in the year's
voting respectively. Factor 1 represented a liberal Partisan Factor, with most senators loading heavily onto it. About half of the senators loaded onto Factor 2, the Administrative Factor. Perhaps 2014 is shaping up to be more of an administrative year than a partisan year...

## Results

The final, hard-earned, results section of this study is very short in comparison to first results section containing the detailed breakdown of the Factor Analysis of each year. There are only two trend lines two look at here: the number of extracted factors from each Senate year with an eigenvalue greater than 1, over time [Figure 2, pictured below], and the number of extracted factors from each Senate year with an eigenvalue greater than 2, over time [Figure 3, pictured below]. Both of these trend lines show that the number of extracted factors that senators have loaded onto has decreased over time. Further, both graphs show that the drastic consolidation in the number of groups that Senators fall into coincides with the Republican Revolution of 1995. It seems, from these Factor Analyses, that partisanship solidified in 1995, never to be overtaken by neutral administrative concerns again. Influential sub-groups of Senators vanished virtually overnight. After 1995, Senators from both parties loaded heavily and often onto Partisan factors, while only occasionally loading onto Administrative factors, while also very rarely loading significantly onto any other factors beside those two.

Fig. 2: Number of "voting entities" over time; Kaiser Stopping Criterion (Eigenvalue > 1)

| Eigen $>1$ |
| :--- |
| Year $\boldsymbol{\xi}_{\text {n }}$ <br> 1989 15 <br> 1990 16 <br> 1991 14 <br> 1992 15 <br> 1993 16 <br> 1994 17 <br> 1995 8 <br> 1996 12 <br> 1997 13 <br> 1998 10 <br> 1999 10 <br> 2000 11 <br> 2001 7 <br> 2002 15 <br> 2003 4 <br> 2004 11 <br> 2005 9 <br> 2006 11 <br> 2007 9 <br> 2008 11 <br> 2009 7 <br> 2010 6 <br> 2011 9 <br> 2012 11 <br> 2013 7 |

Dimension Reduction of U.S. Senators, Years 1989-2013 (Eigenvalue > 1)


Fig. 3: Number of "voting entities" over time; Smith Senate Stopping Criterion (Eigenvalue > 2)


## Discussion

In the late 1980s and early 1990s, Senators tended to act in an administrative capacity in such a way that most of their voting behavior was determined by a general factor that all senators loaded onto. This grouping is what I have termed the Administrative Factor, or "A" Factor, and it accounted for most of the variance of the roll call voting in 1989 and in 1990. During these years, most senators also loaded onto what I termed a Partisan "P" Factor - positively by members of one party, and negatively by members of the other. However, during these years, the P Factor did not account for as much of the overall roll call vote variance as did the A factor. In 1994 the P Factor accounted for more variance than did the A Factor, but this difference was small compared to 1995, when the P Factor took over almost entirely, and many senators stopped loading on to the A Factor at all. The partisan effects of 1995 seem to still be with us.

Not only did the Partisan Factor overwhelm the Administrative Factor in 1995, but also, the number of groups that senators fell into reduced significantly in the same year. They all seem to have decided that they must unite along party lines, or inevitably perish. Perhaps, in 1995, members of the Senate finally got around to reading Anthony Downs and James Sundquist, as they seem to have come to terms with, and mastered, the ideas of maximization of voter utility and drawing bold lines of party cleavage for the sake of retaining their seats in the Senate and furthering their party's platform. Or perhaps the consolidation of senatorial groups is the natural effect of one party taking unilateral control of Congress.

## Conclusion

"Most studies are to some extent both exploratory and confirmatory because they involve some variables of unknown composition. The former should be chosen with great care in order that as much information as possible about the latter be extracted. It is highly desirable that a hypothesis that has been suggested by mainly exploratory procedures should subsequently be confirmed, or disproved, by obtaining new data and subjecting these to more rigorous statistical tests," (Cudeck \& MacCullum, 2007, 58).

Future research is needed is the nascent field of "senators-as-the-variables-not-as-the-subjects-roll-call-vote-Factor-Analysis" in order to build on this rudimentary piece of scholarly work. However, it is my most sincere hope that this research has shed at least enough light on this creative application of Factor Analysis to provoke academicians more gifted than myself (and with more research assistants than myself) to replicate this analysis and to take a deeper look at the subtle senatorial relationships that fell beyond the scope (and outside the time constraints) of this initial study.

## References

Cattell, R. C., (1978). The Scientific Use of Factor Analysis in Behavioral and Life Sciences. Plenum Press, New York. vii-618

Cudeck, R. C. \& MacCallum, R. C., (2007). Factor Analysis at 100: Historical Developments and Future Directions. Lawrence Erlbaum Associates, Inc., New Jersey. vii-381

Dodd, L. C. \& Oppenheimer, B. I., (2013). Congress Reconsidered - Tenth Edition [Chapter 7: The Dynamics of Party Government in Congress (Smith, S. S. \& Gamm, G.), and Chapter 8: Richard Fenno's Theory of Congressional Committees and the Partisan Polarization of the House (Aldrich, J. H., Perry, B. N., \& Rohde, D. W.)]. CQ Press. 167189, 193-216

Harman, H. H., (1976), Modern Factor Analysis - Third Edition Revised. The University of Chicago Press. 24

Kubinger, K. D., (2003). On Artificial Results Due to Using Factor Analysis for Dichotomous Variables. Psychology Science. Vol. 45, No. 1. 106-110

Percy, L., et al., (1976). An Argument in Support of Ordinary Factor Analysis of Dichotomous Variables. Advances in Consumer Research. Vol. 3, 1976. 143-148

Sundquist, J. L. (1973). Dynamics of the Party System. The Brookings Institution Press.
Thurstone, L. L., (1947). Multiple Factor Analysis. The University of Chicago Press. v535

United States Senate Roll Call Tables, Years 1989-2014
http://www.senate.gov/pagelayout/legislative/a_three_sections_with_teasers/votes.htm

## Appendix

Chart 1: All Roll Call Votes Available (1989-2014)


Chart 2: Format of Votes on Official Senate Website (Pre-coding)


## Chart 3: Example of Microsoft Excel Data Set



Chart 4: Example of "Yea Counter" (the black column) with formula exposed


