Electoral Studies 49 (2017) 155–172

Contents lists available at ScienceDirect

Electoral Studies

journal homepage: www.elsevier.com/locate/electstud

# Modeling spending preferences & public policy

J. Alexander Branham, Stephen A. Jessee*

*The University of Texas, Austin, United States*

ARTICLE INFO

*Article history:*
Received 27 February 2017
Received in revised form
8 June 2017
Accepted 30 June 2017
Available online 24 July 2017

*Keywords:*
Public opinion
Elite opinion
Public policy
Bayesian analysis
Item response models

ABSTRACT

Understanding preferences over government spending is important for understanding electoral behavior and many other aspects of the political world. Using data on relative preferences for more or less spending across different issue areas, we estimate the general spending preferences of individuals and congressional candidates along a left-right spending dimension. Our modeling approach also allows us to estimate the location of policies on this same dimension, permitting direct comparison of people's spending preferences with where they perceive policy to be. We find that public shows very low levels of polarization on spending preferences, even across characteristics like partisanship, ideology, or income level. The distribution of candidates' spending preferences shows much more sorting by party, but candidates are significantly less polarized than is contemporary voting in Congress.



## 1. Introduction

Deciding how much to spend and on what is one of the most consequential tasks of a modern day government. The United States government, despite spending a relatively small percent of the country's GDP in comparison to many advanced industrialized democracies, still spends an amount roughly equal to one fifth of the nation's economic output. At the same time, the size of government, which is closely linked to spending, is commonly seen as one of the most prominent issue dimensions dividing the two major political parties in modern U.S. politics, playing a significant role in electoral politics. This makes understanding preferences for spending, particularly in relation to spending levels on specific issues, a particularly important task for scholars.

However, the usual instrument for measuring public opinion — the survey question — has some difficulties measuring spending preferences, which limits the study of spending preference and policy in the electoral arena. While it is easy to imagine that survey respondents can provide meaningful answers to questions on non-spending issues, such as "Do you believe that same-sex marriage should be legal?" or "Under what circumstances do you think that abortion should be allowed?," spending policy is denominated on a scale that is virtually unfathomable to all but the most informed policy wonks.[1] Therefore, surveyors usually ask a less demanding question about respondents' relative preferences — whether they would like to see spending increased, decreased, or kept about the same.[2] A notable application of this logic is the thermostatic model of public opinion and policy (Wlezien, 1995; Soroka and Wlezien, 2010; Wlezien and Soroka, 2012; Pacheco, 2013). In this model, citizens' relative preferences represent the difference between

[1] Previous work has highlighted the difficulty that citizens have in estimating quantities such as the inflation or unemployment rates (Conover et al., 1986) or overall economic conditions (Holbrook and Garand, 1996). Gilens (2001) shows that perceptions about the percent of the federal budget devoted to foreign aid are often very far from the true values. Spending levels would seem to be an order of magnitude more difficult to comprehend. Even knowing whether spending on most areas is measured in millions, billions, or trillions is likely beyond the capacity of many Americans. Ansolabehere et al. (2012) show that survey respondents can understand familiar economic quantities, particularly when provided with benchmarks. This work, however, focuses on numbers that respondents are likely to come into direct contact with in the course of their daily lives such as the price of gasoline. Our focus on federal spending levels seems quite different from these quantities.

[2] There are, of course, other ways of measuring preferences related to spending. One such way is to ask about the general level of spending or taxation rather than spending on a given policy (for examples, Hansen, 1998; Krimmel and Rader, 2017). Using relative preferences gives us the advantage of being able to use multiple questions to jointly scale preferences of the public and congressional candidates together (see Ansolabehere et al., 2008, for a discussion of why using multiple measures of preferences is especially useful).

* Corresponding author. Department of Government, The University of Texas, Austin, United States.
*E-mail address:* sjessee@utexas.edu (S.A. Jessee).

http://dx.doi.org/10.1016/j.electstud.2017.06.007
0261-3794/

the citizen's preferred, or ideal, policy position and the actual location of policy on a given issue. Other scholars have focused on determining how spending preferences on specific issues influences voters' electoral choice (Williams, 2015) or how personal experience with welfare benefits can affect vote choice (Orriols, 2010).

While relative preferences are interesting, they are limited in what they can tell us by themselves. We cannot, for example, measure *distance* — that is, when two people both say that they prefer greater spending, we cannot say whether one of them prefers much more and the other just a little more, or whether they both want a great deal more. Similarly, when a respondent answers that spending is "about right," we have no way of knowing whether spending is exactly right for them or whether they would prefer a little more or a little less. This is important if we want to compare how well represented different groups of the public, if we seek to understand the role of spending preferences in electoral decisions, or if we are interested in studying polarization. Relative preferences also do not give us information on respondent's *overall* spending preferences. While ideology represents a simplification of politics into a left-right space, we lack a comparable measure for spending preferences.[3]

In the next section, we develop a model that uses respondents' stated relative spending preferences to estimate an overall spending preference for each respondent. Previous research has already shown that it is reasonable to scale some issues together to measure underlying spending preferences (Jacoby, 1994; Schneider and Jacoby, 2005; Jacoby, 2008). Our model also estimates the position of spending policy on each specific issue on the same scale as respondent preferences. Following this, we use data from the 2014 General Social Survey (GSS) to estimate the model and discuss the parameter estimates.[4] In addition to constructing a measure for spending preferences and policy location, we also contribute to two debates in the literature.

First, our estimates of citizens' ideal points and policy positions suggest that spending on most policies is lower than many individuals' preferences. There may be systematically lower spending levels than a majority prefers, though limitations of the data make this difficult to say with certainty. These results are in line with scholars who suggest that the government budget is too small (Downs, 1960). Since increased spending is usually associated with liberalism, our results are also in line with studies that find that policy is oftentimes to the right of what people want (Lax and Phillips, 2012; though they focus at the state level).

Second, we also show that there is little polarization in the public, at least with regard to spending preferences. Although the public may be polarized on other issues, it does not appear to be polarized by spending preferences.[5] Additionally, there is virtually no difference in spending preferences across income levels, and only a little across party lines or self-reported ideology. This suggests that, with regard to spending preferences at least, there is little polarization in the public.

From there, we apply our framework to estimate spending preferences of citizens and candidates in congressional elections on the same scale. This is possible because the 1998 GSS and the 1998 National Political Awareness Test, a survey fielded to candidates running for election to the U.S. Congress, used identical or nearly identical questions about spending preferences. These results allow for the direct comparison of spending preferences of the mass public and political elites.

We show that while there is very little partisan polarization among the spending preferences of ordinary citizens, congressional candidates show relatively strong divergence by party in terms of their preferred level of government spending (this is in line with other literature on the subject; see for example Theriault, 2006, 2013). Again, comparing spending levels with spending preferences we find that spending on most policies is lower than median preferences.

## 2. An item response model of spending preferences

Because measuring absolute spending preferences directly through survey questions is infeasible, we propose a model that uses data on relative preferences across specific spending areas to estimate absolute preferences for overall spending. Our approach is related to that of Richman (2011), who combines DW-NOMINATE scores (Poole and Rosenthal, 2011) with legislators' expressed relative preferences in order to estimate the positions of status quo locations. Instead of using exogenous preference estimates, however, we estimate both the preferences of individuals (and later, candidates) and the locations of spending policy in specific areas on a common overall spending dimension.

We build on the ideal point framework commonly used to measure ideology and other latent attitudes in political science (see, for example, Poole and Rosenthal, 1991; Heckman and Snyder, 1996; Clinton et al., 2004). Let $x_i$ represent individual $i$'s ideal point along a spending dimension.[6] Since we are dealing with spending issues, $x_i$ represents a respondent's overall preference for government spending.

Under our model, person $i$'s preferred spending level in policy area $j$ is given as:

$$y^*_{ij} = x_i\beta_j + \varepsilon_{ij} \tag{1}$$

where $x_i$ is individual $i$'s overall spending preference, $\beta_j$ is an issue-specific discrimination parameter, and $\varepsilon_{ij} \sim N(0,1)$ is a disturbance term assumed to be independent across respondents $i$ and issues $j$.

We do not directly observe $y^*_{ij}$, but instead observe the response $y_{ij}$, referred to in the literature as a "relative preference," a trichotomous outcome of either "too much," "about right," or "too little," assumed to be generated according to:

$$y_{ij} = \begin{cases} \text{"too much"} & \text{if } y^*_{ij} < \kappa_{1j} \\ \text{"about right"} & \text{if } \kappa_{1j} \le y^*_{ij} < \kappa_{2j} \\ \text{"too little"} & \text{if } \kappa_{2j} \le y^*_{ij} \end{cases} \tag{2}$$

where $\kappa_{1j}$ and $\kappa_{2j}$ are question-specific cutpoints between the three response options.

[3] Of course, the work on the public's "mood" is related to this (Stimson, 1991; Erikson et al., 2002; Enns and Kellstedt, 2008). Stimson, however, includes non-spending information in his measure. There are also other attempts at creating a spending-specific mood measure (Ura and Ellis, 2012), which we discuss more below.

[4] We also analyze a question wording experiment embedded in the GSS for a majority of the spending issues, showing that most estimates are unaffected by changes in wording.

[5] The literature on polarization is vast and somewhat polarized itself. Abramowitz and Saunders (2008); Baldassarri and Gelman (2008); Webster and Abramowitz (2017), for example, argue that the public is more polarized. Fiorina and Abrams (2008); Levendusky (2009) argue otherwise. Others argue that polarization is more complicated. Perhaps it has occurred in some issue areas, like climate change (McCright and Dunlap, 2011), or only among partisans (Lelkes, 2016).

[6] Others have argued that individuals' spending preferences ($x_i$ in our model) are related and unidimensional (for example Jacoby, 1994, 2008). If preferences in a certain policy are either only weakly related or unrelated to this single dimension, as some previous research finds, then the associated discrimination parameter β will be at or near zero.

Under this model, people are more likely to say there is too little spending on a specific issue when their spending preference $x_i$ is higher. Equivalently, they are more likely to say that there is too much spending on a specific issue when their spending preference $x_i$ is lower. The cutpoints $\kappa_{1j}$ and $\kappa_{2j}$ indicate the thresholds between the three response types ("too much," "about right," and "too little"), which can vary across policy areas. These cutpoints can be thought of similarly to those in an ordered probit model.

Of central interest here are the locations of spending policy, at least as understood by citizens, on each issue area. It should be noted that we are not attempting to estimate actual dollar amounts of spending, which would be readily available from government reports. Instead, we seek to estimate a spending preference dimension structured based on the preferences of people. Whether one prefers more or less spending on a given issue is determined by one's overall spending preference as well as the characteristics of the specific spending area in question (Soroka and Wlezien, 2010). In this framework, for example, there would be nothing wrong with the spending policy position for defense being estimated to be to the left of that for space exploration despite the fact that the country spends many times more on defense than space exploration in dollar terms. What we seek to estimate is one's overall preferences for spending, rather than the specific dollar amount one wishes the government would spend either overall or in any one area.

Although there is not a specific parameter in the model representing the spending locations for each policy area, by explicitly laying out an assumed structure for the relationship between spending preferences and responses, we can produce direct estimates of these policy positions. Specifically, it is possible through a simple transformation of parameters to obtain estimates of these quantities. For issue $j$, we calculate the location of spending policy ($p_j$) as

$$p_j = \frac{\kappa_{1j} + \kappa_{2j}}{2\beta_j} \tag{3}$$

which is simply the average of the two cutpoints on issue $j$ divided by the issue's discrimination parameter.

This location is a sensible estimate of each policy's locations for a few reasons. First, it represents the spending preference value at which a respondent would be equally likely to say "too much" or "too little" is spent on issue $j$. Second, it is also the value at which the probability of saying that spending on issue $j$ is "about right" is maximized.[7] Finally, respondents with overall spending preferences to the left (right) of $p_j$ are more (less) likely to say that there is "too much" than "too little" spending on the policy in question.

The estimated policy positions are not meant to measure the dollar amount of spending on each issue. To the contrary, they represent perceived spending positions according to respondents on a relative, not absolute, scale. Specifically, the spending policy locations indicate the position of each policy's spending level relative to the spending preferences of individuals. Policies on which only the respondents with the highest spending preferences want more spent will be estimated to have higher spending levels, while those on which most respondents want more spent (in other words, only those with the lowest spending preferences prefer that less be spent) will have lower estimated spending locations. Estimating these spending positions reveals the locations of policy in each area relative to the distribution of individuals' overall spending preferences.

## 3. Estimating citizen's spending preferences

As described above, our model uses expressions of relative spending preferences — whether citizens prefer that spending in a given area be increased, decreased, or kept about the same — to estimate the absolute positions of both citizen spending preferences and policies on an overall spending preferences dimension. Therefore, to estimate this model we require survey data in which respondents are asked whether they want more, less, or about the same amount of spending on each of a variety of different spending areas.

The GSS provides just such a dataset. In this section, we analyze the 2014 GSS, which was fielded to 2538 respondents between February and April of 2014. The 2014 GSS targeted English or Spanish speaking people 18 years or older, living in non-institutional arrangements within the United States. These data are particularly well suited for our purposes because they include 18 spending questions across a diverse set of spending areas. Table 1 lists the wordings for these questions as well as the percent of respondent giving each of the three response options — "too much," "about right," and "too little."[8]

Of course, a limitation of these data is that the questions do not explicitly ask respondents to make tradeoffs, though the question text does mention expense. Presumably, individuals' (and later, candidates') responses would change if we forced them to choose between increased spending and higher taxes (or more deficit spending, etc). If there is a lot of arbitrariness to individuals' answers (i.e. if everyone generally wants more "free" spending on all policies), we might worry about how much information we actually have in the data to estimate an underlying spending preference. After all, if we force individuals to make tradeoffs (including in the survey question that increasing spending leads to increased taxes, for example) and everyone then answers that spending levels are about right or too high we may worry how much we can learn from questions without tradeoffs.

Previous research at least somewhat assuages this fear for three reasons. People do not seem to want something for nothing, generally speaking (Welch, 1985). Second, the thermostatic model seems to work. That is, when spending goes up, fewer people want additional spending. This leads us to conclude that people are not viewing spending increases in these questions as "free." Third, people's responses to these spending questions tend to reflect a guns/butter tradeoff (Wlezien, 1995). In other words, when support for defense spending increases (decreases), support for social spending tends to decrease (increase). This indicates that individuals *are* taking into account the fact that money is limited, at least in some sense. This is in line with what scholars have found more recently when they force individuals to decrease spending in one area if they increase spending in another (Bonica, 2015).

We estimate our model in a Bayesian framework using JAGS (Plummer, 2003), called through the rjags package in R (Plummer, 2015), to implement a Gibbs sampler that produces draws from the joint posterior distribution over all of the model's unknown

[7] Though note that this does not mean that "about right" will necessarily be the most likely answer on policy $j$ for a respondent whose spending preference is located at $p_j$.

[8] The question text is "We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount."

**Table 1**
2014 GSS spending items.

| | | | Response percent | | |
|---|---|---|---|---|---|
| Policy | GSS Variable | Wording | too much | about right | too little |
| Education | nateduc | improving the nation's education system | 6.0 | 22.2 | 71.7 |
| Environment | natenvir | improving and protecting the environment | 9.6 | 31.7 | 58.8 |
| Race | natrace | improving the conditions of Blacks | 15.4 | 48.7 | 35.9 |
| Health | natheal | improving and protecting the nation's health | 13.4 | 27.6 | 59.0 |
| Big cities | natcity | solving the problems of big cities | 16.2 | 36.5 | 47.3 |
| Child care | natchld | assistance for childcare | 9.0 | 39.8 | 51.2 |
| Welfare | natfare | welfare | 49.2 | 30.4 | 20.4 |
| Energy | natenrgy | developing alternative energy sources | 10.5 | 32.1 | 57.4 |
| Drugs | natdrug | dealing with drug addiction | 11.9 | 28.7 | 59.4 |
| Mass transportation | natmass | mass transportation | 9.3 | 51.0 | 39.7 |
| Parks | natpark | parks and recreation | 6.4 | 61.9 | 31.7 |
| Science | natsci | supporting scientific research | 12.2 | 46.2 | 41.6 |
| Crime | natcrime | halting the rising crime rate | 8.6 | 30.1 | 61.3 |
| Foreign Aid | nataid | foreign aid | 64.1 | 28.5 | 7.4 |
| Social Security | natsoc | social security | 6.3 | 37.6 | 56.1 |
| Roads | natroad | highways and bridges | 12.1 | 41.1 | 46.8 |
| Space | natspac | space exploration program | 29.0 | 45.0 | 25.9 |
| Defense | natarms | the military, armaments, and defense | 31.1 | 32.4 | 36.4 |

parameters given the data.[9] Independent vague normal priors (mean zero, variance 100) are used for discrimination parameters ($\beta_j$) and cutpoints ($\kappa_{1j}$ and $\kappa_{2j}$), while independent standard normal priors are used for spending preferences $x_i$. The model is run in an unidentified state and each iteration of the sampler is post-processed to impose the identifying restriction that respondent spending preferences have mean zero and variance one, and that higher values of spending preferences represent preferences for more spending in general (although this will not necessarily be true for all individual policies, a point discussed in more depth below).[10]

The estimated discrimination parameters ($\beta_j$) for the 2014 GSS are shown in Fig. 1. These parameters indicate how strongly and in what direction respondents' overall spending preferences ($x_i$) are related to responses on each question. Seventeen out of the eighteen $\beta_j$ estimates are positive, with only one (space) having its 95% highest posterior density region (HPD) overlap zero. This indicates that on almost all spending areas, respondents with higher overall spending preferences are more likely to prefer more (and less likely to prefer less) spending on specific issues.

Some of these positive discrimination parameters, such as those for education, environment and race, are larger in magnitude, indicating that individuals' overall spending preferences are strongly related to spending preferences on that issue.[11] Others, such as those on roads, social security and foreign aid, are estimated to be smaller in magnitude, suggesting that they are not as central to overall spending preferences.

There are two values among the estimated discrimination parameters that merit further explanation. First, space shows little evidence of discrimination along the overall spending dimension. This means that relative preferences over spending on space exploration are largely unrelated to overall spending preferences. Second, defense spending has a discrimination parameter that is estimated to be slightly negative. This implies that those who have higher overall spending preferences are more likely to prefer *less* spending on defense. Equivalently, individuals who prefer less spending generally are more likely to prefer *more* spending on defense. While this may seem somewhat counterintuitive at first glance, defense is one of the few issues in modern American politics on which conservatives, who traditionally prefer lower levels of overall government spending, often argue for more spending than liberals.

The cutpoint parameters for each question ($\kappa_{1j}$ and $\kappa_{2j}$) indicate the regions of the scale for $y_i^*$ in which respondents are likely to give each of the three response options for each question. These parameters, whose estimates are shown in Fig. 2, can be interpreted similarly to the cutpoints in an ordered probit model.[12]

As discussed above, the model is identified by restricting $x_i$ to have mean zero and variance one across all respondents at each iteration of the sampler. Thus we can interpret zero on the spending scale as "moderate," at least relative to the distribution of preferences in the public. The majority of respondents have estimated preferences between −1 and 1, with virtually all the estimates falling between −2 and 2. Overall, the distribution of spending preference estimates is unimodal, and very close to a standard normal.[13]

Fig. 3 plots the estimated location of each of the policies on the same overall spending preferences dimension on which respondent spending preferences ($x_i$) are estimated. Posterior medians are represented by dots and 95% HPDs are represented by lines.[14] The numbers under the policies represent the 95% HPD of the proportion of people whose ideal spending point is to the right of that policy's location.

These estimates represent where each policy falls on the primary dimension structuring citizens' overall preferences for government spending. The highest estimated policy location is for

[9] The sampler was run for 250,000 iterations, discarding the first 100,000 as burn in and saving every 25th after that for a total of 6000 stored samples. Following Gill (2008), multiple diagnostics were calculated. All indicate that the sampler reached convergence fairly quickly and that the effective sample sizes for all parameters were large enough for reliable inference. See Appendix B for details.

[10] This last restriction is operationalized by flipping the scale so that $\beta_j$ is positive for the majority of policies.

[11] Of course, their specific response will also depend on the values of the cutpoints $\kappa_{1j}$ and $\kappa_{2j}$, which are presented in Fig. 2 below.

[12] In fact, conditioning on spending preferences $x_i$ reduces the model to a set of ordered probit models, one for each spending preference question.

[13] See Fig. 4 and the discussion below.

[14] Posterior medians, rather than means, are used for $p_j$ because the division by $\beta_j$ in Equationfootnote:footnote 3 means that for iterations where the discrimination parameter is near zero, the value of $p_j$ becomes extremely large in magnitude, resulting in very unstable estimates. This is only consequential for policies such as space where the posterior likelihood of the parameter being near zero is somewhat large.

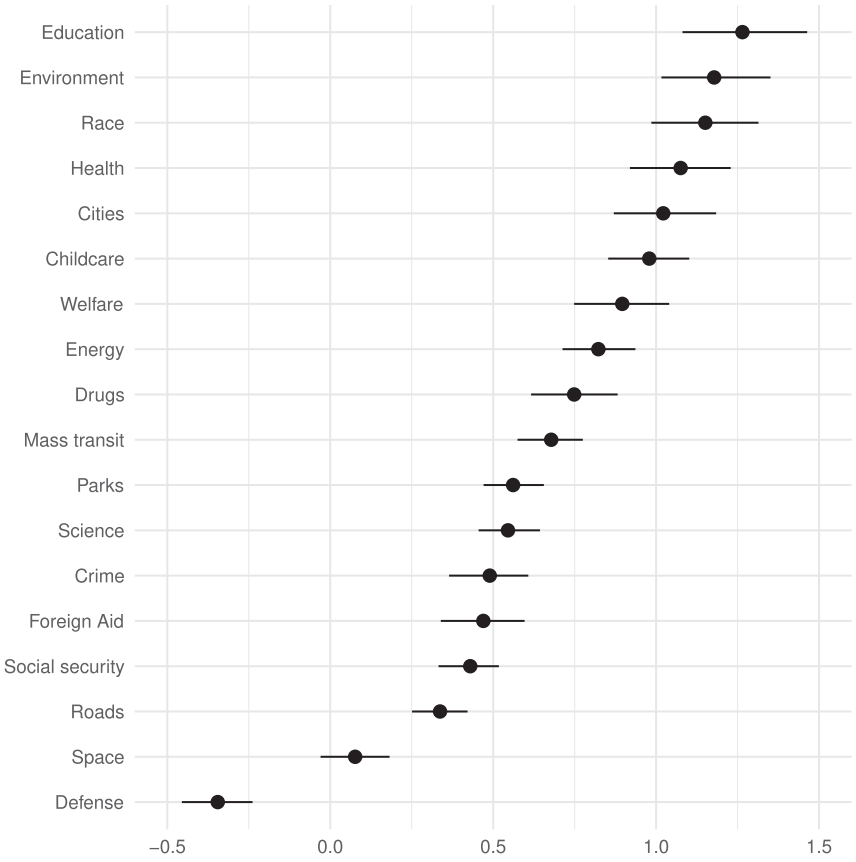



**Fig. 1.** Discrimination Parameter Estimates, 2014 GSS. *Dots represent posterior means, horizontal bars indicate* 95% *HPDs.*

foreign aid with a posterior median of 2.84, indicating that this policy is located to the right of virtually all respondents' overall preferences. Note that this does not mean that we would predict that virtually all respondents would say that the government is spending "too much" on foreign aid, but rather that the probability of saying there is "too little" spent on foreign aid is low for all respondents. The three spending areas other than foreign aid that have a posterior median above zero are defense, welfare, and space. The first two of these are estimated fairly precisely as having values only slightly above zero, while the spending level for space exploration is estimated with a huge amount of uncertainty.[15]

Most other policy positions are estimated to be quite a bit lower than zero, indicating that most respondents prefer increased spending. Social security, the policy with the lowest posterior median, is estimated to be at −2.89 with over 99% of respondents' ideal points estimated to be to the right of the policy (the 95% HPD for the percentage of respondents with spending preferences above this policy's location is [98.9, 100]). Even policies with higher estimated positions are estimated to be lower than large majorities of respondents' preference. The estimated policy location for race, for example, is estimated to be lower than the spending preferences of 71.6% of respondents (95% HPD [67.6, 75.3]). In fact, fourteen out of the eighteen policy areas are estimated to be less than zero, meaning that for the vast majority of areas surveyed in these data, spending levels are below the average spending preference of the American public.[16] Previous work finds that a large segment of the American public is in favor of increased spending, so we view this as a positive check of the face validity of our measure.

Given that the size of government is a central divide between the platforms of the modern Democratic and Republican parties, we might expect there to exist large differences between Democratic and Republican identifiers in the American public. Surprisingly, there is an extremely high amount of overlap between the spending preferences of Democratic, Republican, and independent identifiers. Fig. 4 plots the distribution of estimated spending preferences by party identification.[17] The average spending preference among Democratic identifiers is 0.24 (95% HPD [0.21, 0.26]) while for Republicans it is −.32 (95% HPD [−.35, −.28]). The average independent spending preference is −.02 (95% HPD [−.06, 0.03]). This means that not only are most spending policy positions lower than the average American's preferences, but 14 out of the 18 policies are estimated to be lower than the average Republican citizen's overall spending preferences. Finding that Republicans are

[15] The reason for this uncertainty becomes clear when recalling the estimated discrimination parameter for space from Fig. 1. Space is the only spending area whose $\beta_j$ is not estimated to be clearly to one side of zero. Because policy positions $p_j$ are calculated by dividing the average of the two cutpoints on an issue by its discrimination parameter (Equationfootnote:footnote 3), policies for which the discrimination parameter is close to zero will have policy positions estimated very imprecisely. The logic is that if responses regarding spending levels on a specific issue, such as space here, are unrelated to respondents' overall spending preferences, then these responses do not provide information about the spending level on that specific issue.

[16] Of course, this could, and most likely would, change dramatically if we change the question as to make the tradeoffs between, for example, more spending and higher taxes more stark, as discussed above.

[17] Leaning independents are coded as partisans following Keith et al. (1992).

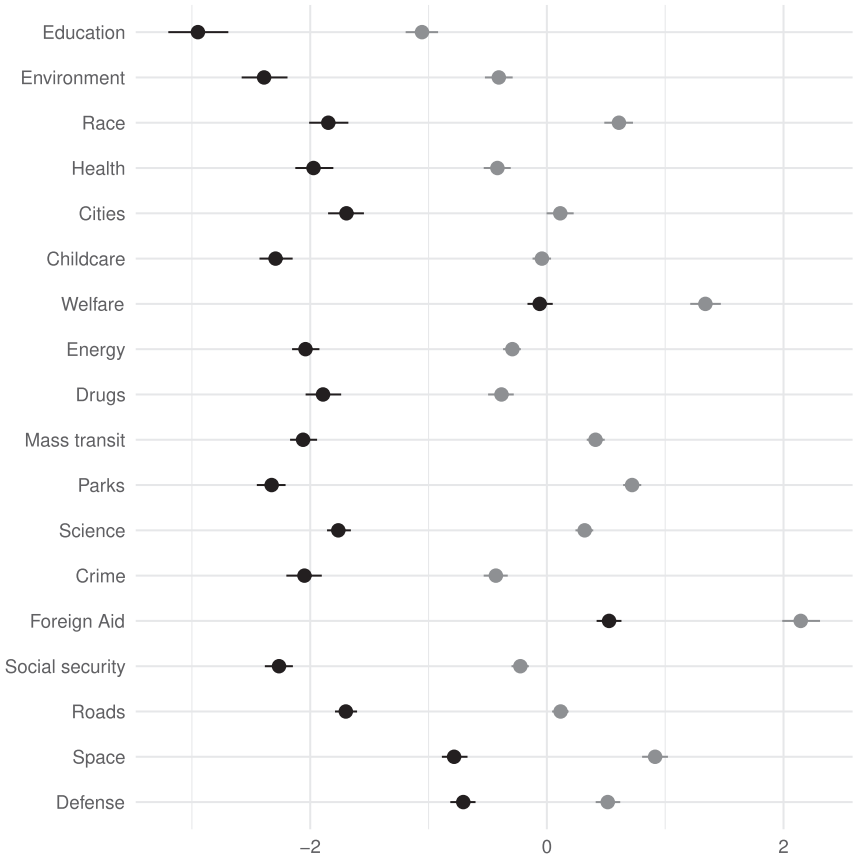



**Fig. 2.** Cutpoint Estimates, 2014 GSS. *Dots represent posterior means for* $\kappa_1$ *and* $\kappa_2$*, horizontal bars represent* 95% *HPDs.* $\kappa_2$ *is in the lighter shade.*

in favor of increased spending on a wide range of policy issues is in line with previous research (see, for example Ellis and Stimson, 2012).

Spending preferences show a slightly stronger relationship with self-placed ideology than with party identification, shown in Fig. 5a. Still, the average spending preference among self-identified "extremely conservative" respondents is −.54, which is still above fifteen out of the eighteen policies included in the 2014 GSS. In fact, defense is the only policy position whose location is estimated to be between the average preferences of extreme liberals and extreme conservatives.

Income seems to be more or less uncorrelated with spending preferences. Fig. 5b shows the breakdown of spending preferences by income. The distribution of spending preferences by income level is in line with other recent work on this subject (Soroka and Wlezien, 2008; Ura and Ellis, 2008; Branham et al., 2017). Even though income inequality historically high and still rising (Brandolini and Smeeding, 2006), overall spending preferences are strikingly similar across income levels.

Embedded in the 2014 GSS is a question-wording experiment in which half of the respondents, randomly selected, were shown alternate question wordings for eleven of the eighteen items as well as the remaining seven items for which there was only one wording. Some of these changes (e.g. "space exploration" versus "space exploration program") seem unlikely to make a difference. Others, however, are quite different such as the two wordings on welfare: "welfare" and "assistance to the poor," paralleling classic examples of question wording effects (Schuman and Presser, 1981; Rasinski, 1989). In Appendix Appendix A, we analyze this, including the standard and alternate wordings for each of these items as well as the questions with only one wording, for a total of 29 items. For the majority of these alternate wordings, item parameters were all similar. Four questions, however, showed significant differences: welfare, cities, race, and crime.[18] As Appendix Appendix A shows, the overall results presented above are similar whether using only the standard question wordings or including alternate wordings as separate items.

This section's findings demonstrate that the mass public is not highly polarized in terms of their spending preferences. Citizens show only minor differences in spending preferences by party identification, and only slightly larger differences by self-placed ideology. There are virtually no differences in spending preferences by income level. While this lack of polarization might be expected to produce spending policies that are highly representative of the median voter, this is not the case, at least judging by the spending policy locations implied by respondents' views. For the vast majority of areas, spending policy is estimated to lie below, usually far below, the average respondent's preferences, and few of the policy locations are estimated to be close to the center of the distribution of citizen preferences. This apparent skew in representation on spending policy begs the question of whether political elites have different preferences from those of ordinary citizens,

[18] Welfare comes as no surprise, given the difference in question wording. One of crime's wordings is "halting the increasing crime rate," which is an odd way to phrase the question given that crime rate in the U.S. has been on the decline for a few decades. This rather odd question phrasing may be to blame for the different item parameters. It is unclear why the parameters with cities are different. Aside from these three, there were some other very minor differences with the cutpoints for two questions (drugs and foreign aid).

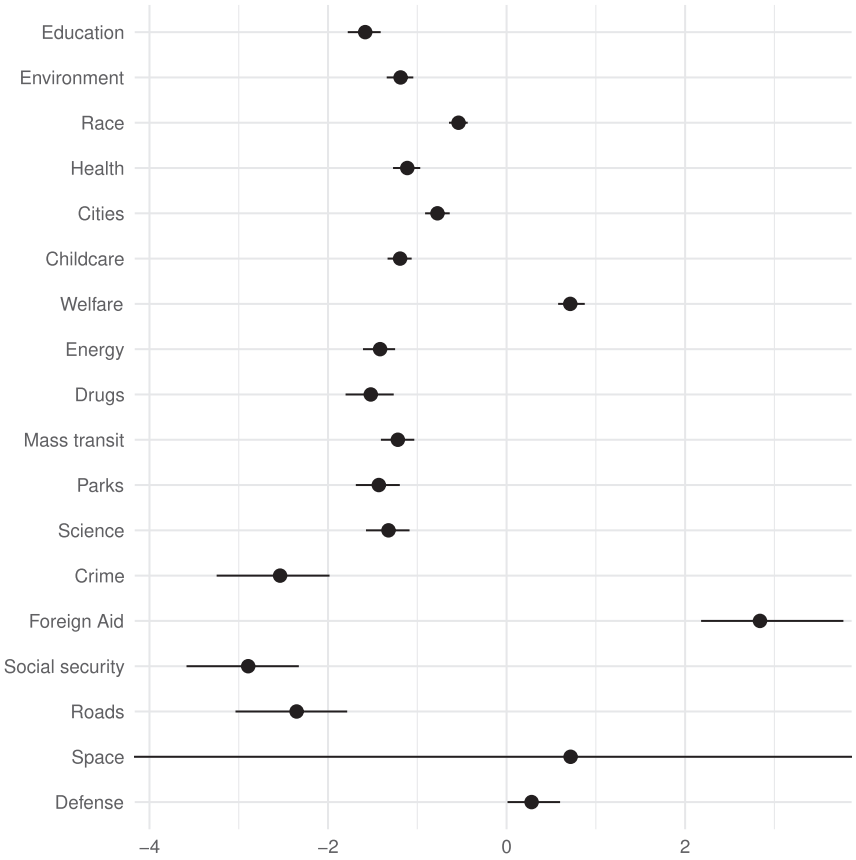



**Fig. 3.** Policy Position Estimates, 2014 GSS. *Dots represent posterior medians for policy's location as defined in Equation eq:policy-location. The HPD for space's location is not fully contained in the figure—it ranges from around −6 to 8, but is not reliably estimated given that the discrimination parameter for this item is often sampled near zero. See footnote 14 for more discussion of this.*

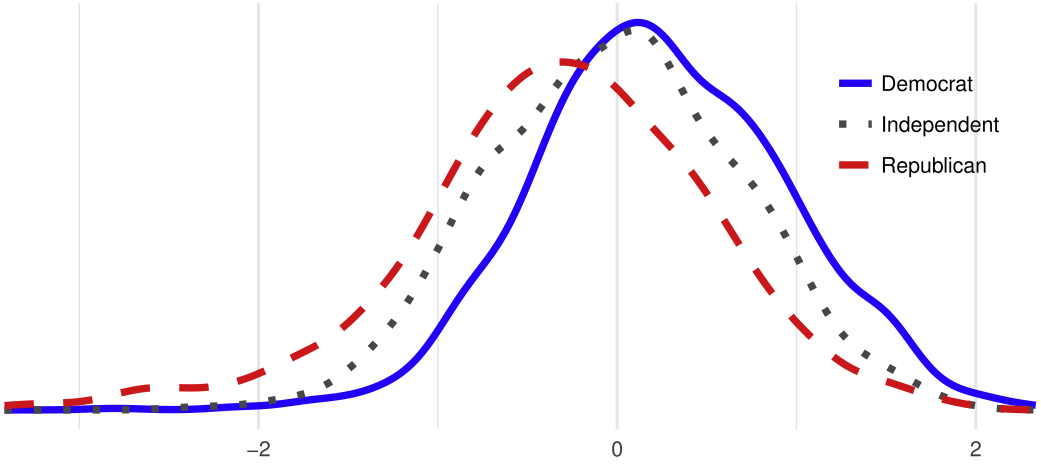
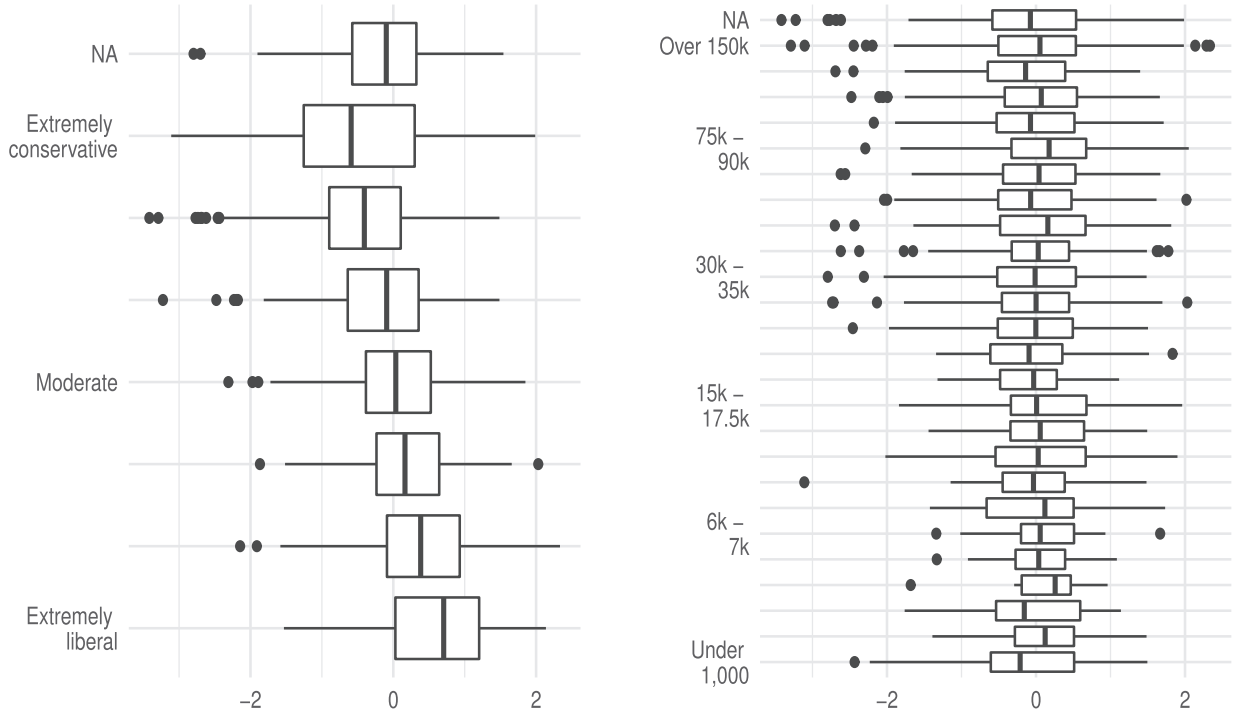



**Fig. 4.** Spending Preferences by Party ID, 2014 GSS. *The blue solid line indicates Democrats, red dashed Republicans, and gray dotted independents. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)*

including whether they are similarly homogeneous or whether they are more polarized by party in terms of their spending preferences.

## 4. Adding candidates' preferences

The above results shows that policies are not where most people want them — in fact, they are oftentimes far below the median citizens' preferences. There are several possible explanations for this, one of which being that elites' spending preferences do not look like citizens' spending preferences and since political elites make policy they are simply creating policies that are more suited to their own preferences. Additionally, as we show in Fig. 4, spending preferences by party are not very polarized. Is this also




(a) Spending preferences by ideology

(b) Spending preferences by income

**Fig. 5.** Spending Preferences by Ideology and Income, 2014 GSS. *Boxplots show distribution of estimated spending preferences $x_i$ by self-placed ideology (left pane) and income (right pane). The box begins at the 25th percentile and extends to the 75th percentile. The bar represents the mean. The "whiskers" extend to the most extreme value within 1.5 IQRs. Points more extreme than that are plotted as dots.*

the case for elites, or are elites more sorted by party?

We focus here on estimating spending preferences of citizens and candidates for Congress on the same scale. This requires data that asks the same (or at least similar) questions of both sets of actors. For this, we rely on the National Political Awareness test (NPAT), which is a survey fielded to candidates for office by the organization Project Vote Smart (see Ansolabehere et al., 2001; Shor and McCarty, 2011; Richman, 2011; for previous research using these data.)[19] Most importantly for our purposes, some waves of this survey include questions about candidates' relative spending preferences. The 1998 wave of the NPAT was chosen because it provides the best combination of candidate response rate and question overlap with the corresponding year for the GSS.[20] Specifically, candidates are asked across a range of different issues to "Indicate what levels of funding you support for the following categories." We drop all third party and independent candidates, leaving 449 major-party candidates who ran in the 1998 general election for congressional office.

Table 2 lists the specific policies candidates were asked about in the 1998 NPAT as well as the percentage breakdown of responses. The NPAT and GSS provide slightly different response options for spending questions, which we recode to correspond to obtain comparable scales.[21] The questions span a fairly wide range of policies across different issue areas.

In order to estimate the spending preferences of candidates alongside those of citizens, we must match some of the questions in the GSS and NPAT. Fortunately, the question wordings from the NPAT and GSS are very similar or identical for 7 out of 13 questions on the NPAT. We were able to match questions about spending on the arts, education, the environment, crime, space, and welfare. NPAT items on AIDS programs, housing projects, job training programs, medicaid, medicare, and student loan programs did not have any corresponding item in the GSS and therefore were not matched, but instead were included as separate items that candidates responded to, but citizens did not. The second column of Table 2 indicates which GSS questions the NPAT items are combined with.

Table 3 shows the proportion of respondents in each response category of the 1998 GSS. The 1998 GSS, like the 2014 version, included two question wordings for several of its items with respondents randomized to either the standard or alternate wordings. In order to include all respondents from the 1998 GSS in the combined analyses (rather than dropping roughly half of them), we had

[19] One potential objection to using NPAT data is that perhaps candidates who fill out these data are unrepresentative of the larger group. Later, we compare candidates who completed the NPAT to those who did not. Results do not indicate that NPAT respondents are dramatically different from non-respondents. In 2010 this survey was renamed the Political Courage Test.

[20] In recent years, candidate response rates have dropped precipitously. Although the 1996 NPAT was completed by more major-party candidates, it included fewer questions that matched well with GSS items (5 for the 1996 NPAT vs. 7 for the 1998 wave).

[21] We recode the "greatly increase" and "slightly increase" questions from the NPAT to correspond to the GSS response option of "too little." We recode the "eliminate," "greatly decrease," and "slightly decrease" on the NPAT to correspond to the GSS's "too much" response, and we recode NPAT responses of "maintain status" to match the GSS's "about right" response.

**Table 2**
Question wordings and proportion of candidates in each category for 1998 NPAT.

| | | Response percent | | |
|---|---|---|---|---|
| NPAT Question | GSS match | too much | about right | too little |
| AIDS programs | N/A | 15.2 | 44.4 | 40.5 |
| Arts funding | natarts | 44.0 | 37.7 | 18.4 |
| Education K-12 | nateduc, nateducy | 9.2 | 18.4 | 72.4 |
| Environmental programs | natenvir, natenviy | 16.1 | 40.8 | 43.1 |
| Foreign aid | nataid, nataidy | 52.7 | 40.9 | 6.5 |
| Housing projects | N/A | 24.5 | 47.7 | 27.8 |
| Job training programs | N/A | 15.0 | 30.9 | 54.1 |
| Law enforcement | natcrimy | 5.6 | 30.9 | 63.6 |
| Medicaid | N/A | 4.9 | 62.5 | 32.6 |
| Medicare | N/A | 2.8 | 53.2 | 44.0 |
| NASA | natspace, natspacy | 24.2 | 56.7 | 19.1 |
| Student loan programs | N/A | 6.9 | 35.3 | 57.8 |
| Welfare | natfare | 46.1 | 45.0 | 8.9 |

**Table 3**
1998 GSS items.

| | | Response percent | | |
|---|---|---|---|---|
| Policy | GSS variable name | too much | about right | too little |
| Parks | natpark | 6.4 | 57.9 | 35.7 |
| Roads | natroad | 10.1 | 49.6 | 40.3 |
| Social security | natsoc | 6.7 | 34.0 | 59.4 |
| Mass transportation | natmass | 10.7 | 53.5 | 35.9 |
| Education (alt) | nateducy | 6.1 | 19.3 | 74.6 |
| Crime (alt) | natcrimy | 8.1 | 36.5 | 55.4 |
| Welfare (alt) | natwelfy | 11.5 | 26.2 | 62.3 |
| Health (alt) | nathealy | 7.8 | 22.0 | 70.2 |
| Environment (alt) | natenviy | 7.9 | 26.6 | 65.5 |
| Foreign Aid (alt) | nataidy | 72.0 | 21.5 | 6.5 |
| Defense (alt) | natarmsy | 32.1 | 49.2 | 18.7 |
| Education | nateduc | 6.4 | 22.6 | 71.0 |
| Health | natheal | 5.8 | 25.7 | 68.6 |
| Drugs (alt) | natdrugy | 12.1 | 31.2 | 56.7 |
| Space (alt) | natspacy | 42.3 | 45.6 | 12.1 |
| Crime | natcrime | 7.2 | 29.5 | 63.3 |
| Drugs | natdrugy | 9.8 | 29.3 | 60.9 |
| Environment | natenvir | 8.0 | 29.3 | 62.7 |
| Welfare | natwelf | 45.4 | 37.8 | 16.8 |
| Defense | natarms | 32.2 | 49.3 | 18.6 |
| Foreign Aid | nataid | 64.2 | 28.7 | 7.1 |
| Space | natspac | 42.2 | 47.0 | 10.8 |
| Race (alt) | natracey | 25.0 | 44.8 | 30.2 |
| Race | natrace | 18.0 | 44.4 | 37.6 |
| Big cities | natcity | 13.8 | 34.1 | 52.1 |
| Arts | natarts | 22.4 | 52.5 | 25.2 |
| Big cities (alt) | natcityy | 33.2 | 43.8 | 23.0 |

to make decisions about whether to combine different wordings into a single item. In some cases, e.g. “space exploration program” versus “space exploration,” it seems obvious that the questions can be combined. Others, however, such as “welfare” versus “assistance to the poor” seem potentially problematic. In order to assess whether it is reasonable to combine GSS question wordings, we estimated the model for the GSS spending preferences data alone, treating each question wording as a separate item. We then asked whether the parameter estimates between different wordings of the questions showed notable differences. If they did not show differences, it seems reasonable to combine those different question wordings. We combined question wordings for seven out of the eleven items with multiple wordings.[22]

Among the questions where wordings were not combined, only crime and welfare have corresponding NPAT items. In both of these cases, the choice of which GSS wording to match to the NPAT question was clear because the NPAT question wordings —“welfare” and “law enforcement” — match perfectly to a GSS wording, while the other question wording for these items —“assistance to the poor” and “halting the rising crime rate” — seem quite different. In both of these cases, the remaining GSS wording was treated as a separate item, with no NPAT item matched to it.

As above, the model is identified by post-processing each iteration of the sampler to impose the restriction that spending preferences ($x_i$) for citizens and candidates together have mean zero and variance one and that higher values represent preferences for more spending. The sampler is run for 250,000 iterations with the first iterations 100,000 dropped and each 25th iteration thereafter saved. Examination of multiple diagnostics showed strong evidence of convergence (see Appendix Appendix B).

Fig. 6 shows the estimated discrimination parameters for each item. As for the 2014 GSS, virtually all of these values are positive, indicating that higher overall spending preferences are positively associated with desire to increase spending levels. Only defense spending is estimated to clearly have a negative discrimination parameter, while the discrimination parameter for space overlaps zero, meaning that preferences for spending on space are unrelated to overall spending preferences. The cutpoints, shown in Fig. 7 also show similar characteristics, generally speaking, to those from the 2014 GSS.

Fig. 8 shows the estimated locations for spending policy for each item. As before, the dots represent posterior medians and the lines indicate 95% HPDs. The estimated spending policy locations from the 1998 data are mostly concentrated between negative two and zero as they were for the 2014 GSS. Only four policy areas — arts, foreign aid, welfare, and big cities (alternate version) — have estimated spending locations to the right of zero. As before, the spending location for space is estimated with a huge amount of uncertainty due to the fact that the posterior for its discrimination parameter is concentrated near zero. The largest outliers in spending policy locations are foreign aid, which is positioned far to the right, and roads and crime (alternate version), which are estimated to be far to the left.

Fig. 9 shows the distribution of estimated spending

[22] For five of these items (defense, law education, environment, health and space), the 95% HPDs for all item parameters overlapped across question wordings. Only $\kappa_1$ failed to overlap between wordings for the foreign aid item and only $\kappa_2$ failed to overlap (by 0.001) for the drugs item. Because of these relatively minor differences, we combined the wordings of these two questions as well. The remaining four items showed notable differences for most or all parameters and therefore were not combined.

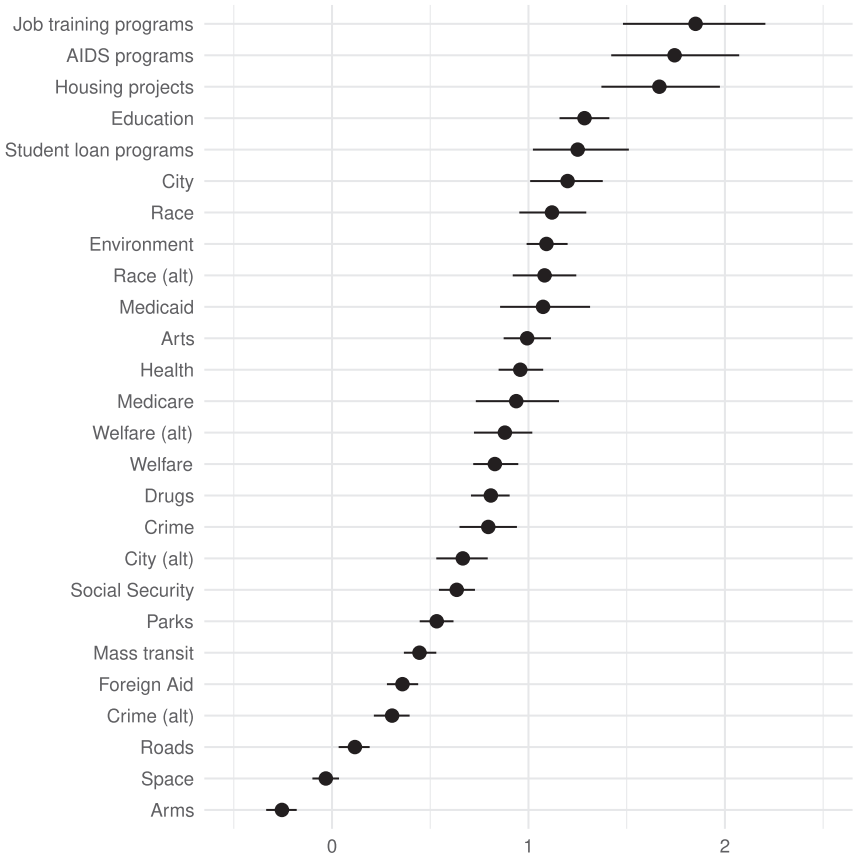



**Fig. 6.** Discrimination Parameter Estimates, 1998 GSS & NPAT. *Dots indicate posterior means, horizontal bars represent* 95% *HPDs.*

preferences for congressional candidates and respondents to the GSS, separated by party. As for the 2014 GSS, citizen spending preferences show little evidence of polarization. The preferences of Democrats, independents, and Republicans differ by a small amount on average and there is a very high amount of overlap between the distributions for each of these groups. Only seven percent of the variation in spending preferences is explained by party identification.

Among congressional candidates, there is evidence of more polarization, but the spending preference distributions are still relatively similar. While thirty-five percent of the variation in candidate spending preferences is explained by party, there is still a much higher degree of overlap between Democratic and Republican candidates' spending preferences than there is between, for example, the DW-NOMINATE scores of the survey's candidates (Poole and Rosenthal, 2000) or the cfscores estimated by Bonica (2013b).

More than just examining the respective shapes of citizen and candidate spending preference distributions, however, our joint scaling allows us to answer questions about the preferences of these two groups compared to each other. The overall variation in candidate spending preferences is only slightly larger than that for citizens (standard deviations of $x_i$ are 1.16 and 0.97, respectively). Perhaps even more surprisingly, the within-party variation for citizens and candidates is similar.[23]

This result contrasts with those of Bafumi and Herron (2010), who look not at spending preferences but at overall policy ideology estimated for members of Congress and the public. Bafumi and Herron find that the distribution of legislator ideology is quite bimodal, with most citizens holding ideological positions in between these two modes. Our results show that on spending preferences, at least those stated by candidates in the NPAT survey, candidates positions are much more unimodal (note that if pane (b) of Fig. 9 were plotted for all candidates rather than separately by party, the distribution would look roughly normal, in contrast to Bafumi and Herron's distribution of estimated ideology in which Democratic legislators are almost completely separated from Republican legislators with very few estimated to be in the middle range near the median American).[24]

A potential objection to looking at NPAT respondents is that they are unrepresentative of the overall population of candidates. Although other research (Ansolabehere et al., 2001; Shor and McCarty, 2011; Richman, 2011) suggests this is not the case, especially when response rates were still relatively high as in the 1998 NPAT, we investigate whether non-respondents look different than candidates who respond to the NPAT.

Using data from Bonica (2013a), we were able to successfully merge nearly all of the candidates in our dataset with data

[23] The standard deviations for Democratic and Republican candidate $x_i$ distributions are 0.92 and 1.03, respectively. For respondents, these values are 0.91, 0.95, and 0.97 for Democrats, Independents, and Republicans, respectively.

[24] Note also that estimated candidate spending preferences do not show large differences for incumbents, challengers, and open seat candidates.

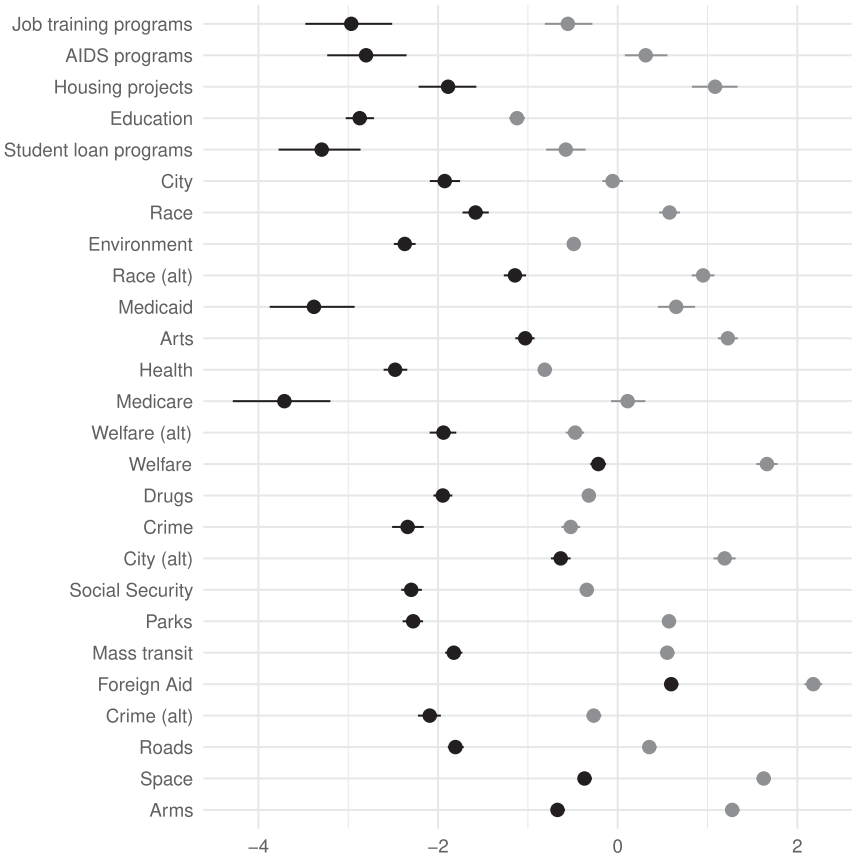



**Fig. 7.** Cutpoint Estimates, 1998 GSS & NPAT. *Dots represent posterior means, horizontal bars indicate 95% HPDs.*

including election results and other candidate information.[25] Evidence suggests that NPAT respondents and non-respondents are very similar. We ran differences in means tests between candidates who responded to the NPAT vs those who did not across several candidate characteristics. Insignificant results included cfscores ($p = 0.39$), DW-NOMINATE ($p = 0.65$), party ($p = 0.91$), and gender ($p = 0.92$).

Variables that had a statistically different mean between respondents and non-respondents included whether the candidate won or lost ($p = 0.03$, 47% of people who responded were winners vs 40% of non-responders) and the number of contributors to a candidates' campaign ($p = 0.003$, candidates who responded have more givers on average (mean = 342.2) than those who did not (mean = 229.9)). Overall, these results suggest that while there are some candidate characteristics that correlate with nonresponse, the candidates responding to the NPAT do not look dramatically different from those who did not.

We also looked at how different our spending preference estimates are from ideal points that are designed to capture overall ideology. Evidence suggests that, while correlated, spending preferences and overall ideology are different. Our point estimates are correlated with Poole and Rosenthal's DW-NOMINATE scores (based on voting behavior) at $r = -0.76$. They are correlated with Bonica's estimates (based on campaign contributions) at $r = -0.66$. Even with the relatively high correlations between these measures, they tell different stories, particularly about partisan polarization. That spending preferences are different from ideology may also explain why we find only slight polarization at the elite level. Polarization here is estimated to be minimal with regard to spending preferences, while other measures that include a mix of different issues (e.g. gay marriage, abortion, immigration, etc) show a much wider gulf between parties.

## 5. Examining joint scaling assumptions

A key assumption of the joint scaling of citizen and candidate data above is that the same single-dimensional structure underlies the views of both groups. Obviously, this is implausible in an exact sense, but the important question for our purposes is whether the two groups' preferences are primarily explained by a similar enough underlying dimension to make jointly estimating their preferences on the same scale a useful and interesting exercise.

In order to assess this, we conduct three sets of analyses, each directed at a different question. First, we separately examine the variance explained by the principal components of the GSS and NPAT datasets. This gives a sense for how much of the variation in citizens' and in candidates' spending preferences are explained by a single spending dimension. Second, we conduct an exploratory factor analysis of both of these groups' spending responses and compare the factor loadings from each one. This allows us to compare the way individual spending items (e.g. "Defense" or "Welfare") relate to the underlying spending dimension for each of these groups and whether these relationships are similar between the two groups. Finally, we relax the model estimated in section 4

[25] Only 54 major-party candidates from the 1998 NPAT did not appear in Bonica's data. Many of these are candidates who did not perform well (under 30% of the popular vote, for example).

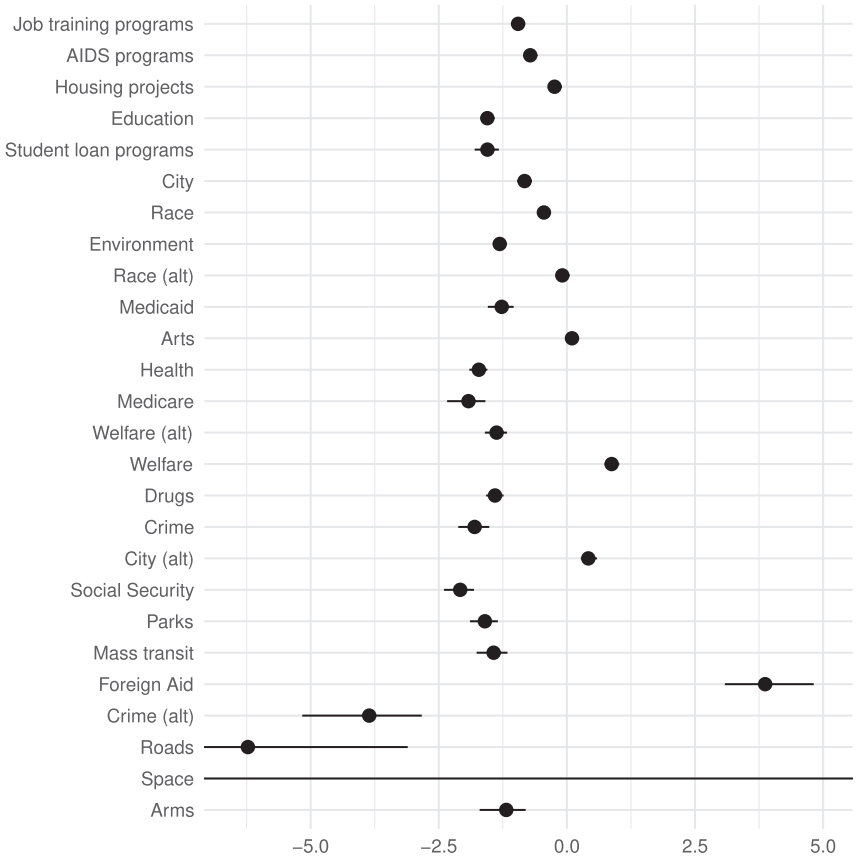



**Fig. 8.** Policy Position Estimates, 1998 GSS & NPAT. *Dots indicate posterior medians, lines indicate* 95% *HPDs. The lower HPD bound for roads is not included, it is* −14.3. *The median for space is* −13.2 *with an HPD of* −197.6 *to 194.1.*

to allow citizens to have different utility error variances than candidates in order to determine if this changes our central findings.

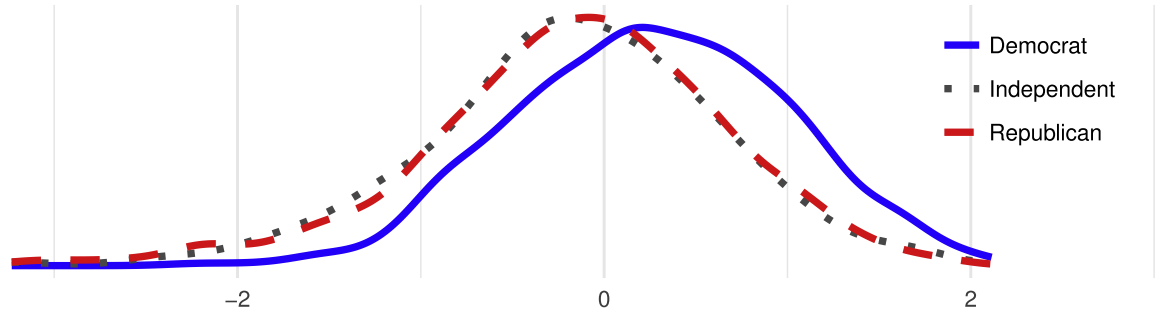
### 5.1. Dimensionality of NPAT and GSS data

Fig. 10 presents scree plots of the GSS and NPAT data based on a principal components analysis.[26] Both plots show a high value for the first principal component, with a relatively large decline for the second, and smaller drops after that. This "leveling out" of the scree plot after a single value is typically interpreted as evidence of a strongly single-dimensional structure for the data. Although other cutoffs (e.g. which principal components have values greater than one) are sometimes used, there are no hard and fast rules for interpreting scree plots. What these plots both show, however, is that a single spending dimension can explain a large proportion of the variance in both citizens' and candidates' responses to spending questions. Although subsequent dimensions can add explanatory power, each one contributes much less than does the first dimension. So although these datasets are not exactly single-dimensional, a single-dimensional analysis is likely to contribute quite a bit to our understanding of the spending preferences of these two groups.

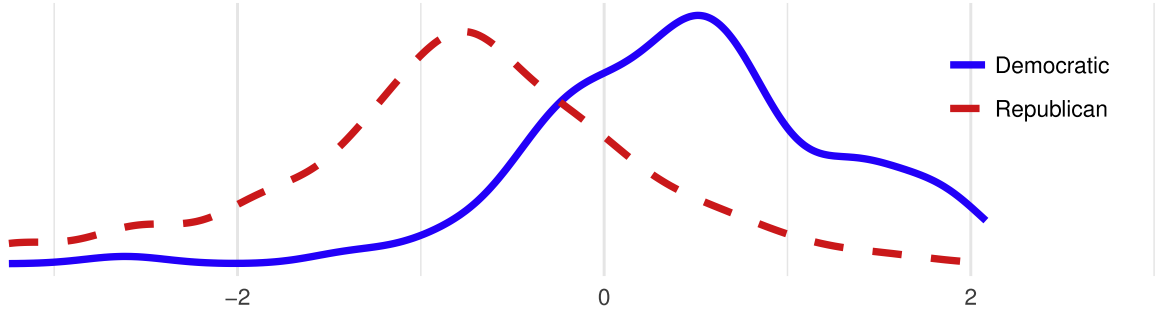
### 5.2. Structure of citizen and candidate spending preferences

Another worry we may have is whether congressional candidates and individuals have differently-structured spending preferences. Jessee (2016), for example, discusses this issue in the context of estimating the policy ideologies of citizens and members of Congress. One way of examining whether the single dimensional structures underlying each of these two datasets is similar is by looking at factor loadings from factor analysis. Fig. 11 shows the factor loadings from separate unidimensional factor analyses of merged items from the NPAT and GSS data (see Table 2). Although the separate analyses don't produce factor loadings that are directly comparable on the same numerical scale (for example, a value of 0.6 for an NPAT factor loading doesn't necessarily mean the same thing as the same value for a GSS factor loading), we can see whether the general pattern of loadings is similar. It is clear from Fig. 11 that these loadings are indeed very similar overall, with a relatively linear relationship existing between the two sets of estimates. The items with factor loadings near zero (such as "Space" or "Defense"), which do not strongly relate to other spending items in the NPAT, also tend to be the ones that have factor loadings close to zero in the GSS data. Conversely, those with larger factor loadings in the NPAT also tend to have larger factor loadings in the GSS. This suggests that the primary dimension underlying citizens' and candidates'

[26] This analysis was done using the **princomp** function in R. Because GSS respondents were randomized on questions with two versions to either get all standard question wordings or all alternate wordings (not independent randomization across questions), we dropped all alternate wordings from this analysis. This is done since no respondents answered both a standard and an alternate wording and therefore the correlations between these items, which are needed for the principal components analysis, cannot be calculated. Results from instead dropping the standard wordings of items with two versions were very similar.




(a) Spending preferences of individuals




(b) Spending preferences of candidates

**Fig. 9.** Spending Preferences of Individuals and Candidates. *Panes plot density of estimated spending preferences $x_i$ for 1998 GSS respondents (top pane) and congressional candidates from the 1998 NPAT (bottom pane) separated by partisanship from joint scaling of these two datasets.*

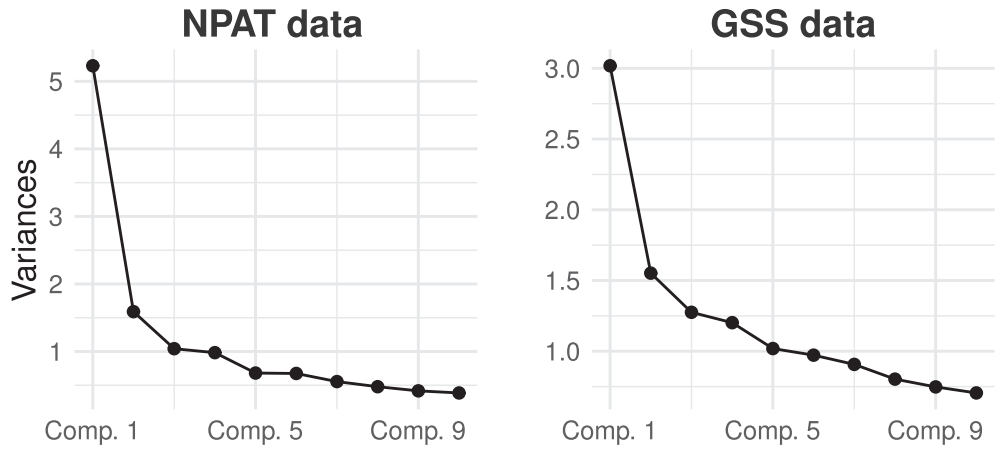



**Fig. 10.** Scree Plots for GSS and NPAT Items. *Dots represent variance explained for the first ten principal components of 1998 GSS and NPAT data. Because of lack of overlap between questions asked, the alternate wordings for GSS questions are dropped from the analysis.*

spending preferences in these common items is structured in a similar way. To the extent that there are some differences between these two groups' preference structures, the much larger number of GSS respondents as compared to the number of NPAT candidates should make the estimated dimension closer to that for citizens than for candidates. Thus our findings can be interpreted as answering questions about how citizens perceive spending policy levels and candidates' stated spending positions.

### 5.3. Heteroskedastic errors for citizens and candidates

As a final robustness check for our item response theory model, we estimate a model in which the variance of the latent-scale disturbance (error) term is allowed to be different for citizens and candidates.[27]

[27] Formally, we fix the standard deviation of the errors $\varepsilon_{ij}$ from Equationfootnote: footnote 1 at 1 for all candidates and estimate a common standard deviation for $\varepsilon_{ij}$ for all respondents, using a Cauchy prior with scale 1000 truncated below zero for this parameter. The model is otherwise identical to the main one used in section 4.

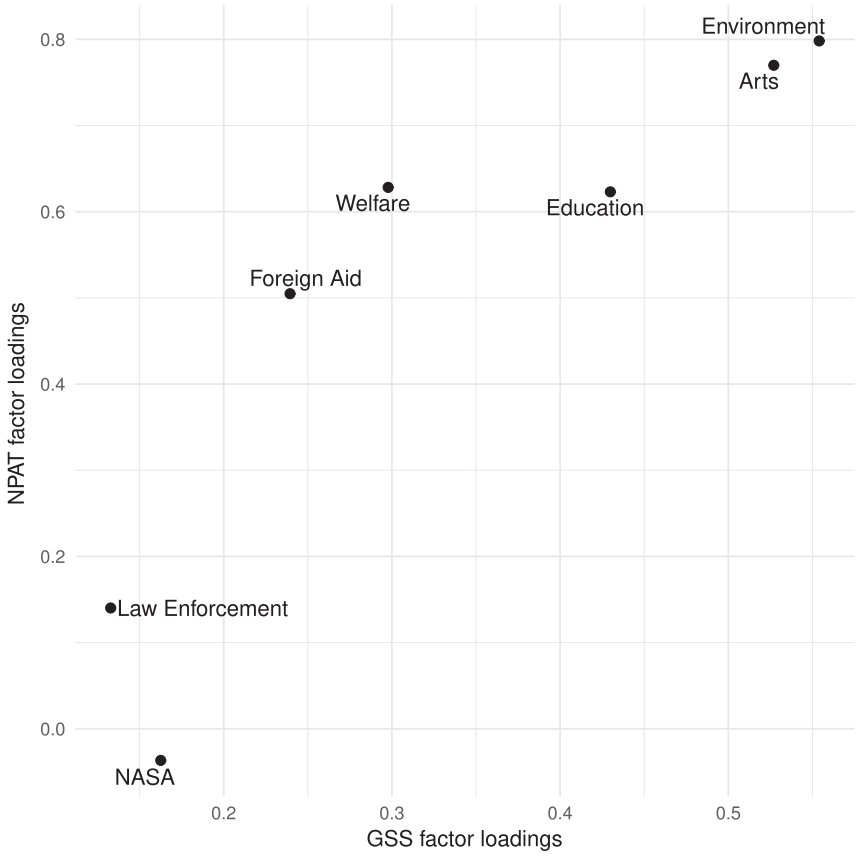



**Fig. 11.** Factor Loadings for 1998 GSS and NPAT Items. *Dots represent factor loadings from separate factor analyses of matched items from both datasets.*

Following previous work by Jessee (2009) and Lauderdale (2010), this allows for the possibility that ordinary citizens might take positions on individual policies with more "noise" than candidates who have otherwise identical overall spending preferences. This might not be unexpected given that candidates are essentially professional position takers.[28] Although they aren't strictly comparable as noted above, the factor loadings from the NPAT and GSS data shown in Fig. 11 also provide some suggestion that candidates discriminate more precisely along this spending dimension.

The results of this heteroskedastic model indeed suggest that ordinary citizens have a higher standard deviation in their position taking than do candidates. The posterior mean for respondent error standard deviation is 1.41, with a 95% HPD of $[1.31, 1.52]$. Although this is not a large difference, we may wonder whether the discrepancy between citizens and candidates affects any of the findings above. In all cases, however, estimates of the model's other parameters (the distribution of individuals' spending preferences, item discrimination parameters, item cutpoints, policy locations) are nearly identical between the homoskedastic and heteroskedastic models and all of the substantive findings above are the same regardless of which of these two models is used.

## 6. Discussion

The United States federal government spends each year an amount of money that is unfathomable for many people. Hundreds of billions of dollars are spent on defense alone. The same goes for spending on various welfare programs. In total, the U.S. federal government spends about one of every five dollars spent in the United States each year. Thus, if we are interested in studying public opinion on spending, it is important to measure spending preferences correctly. In this paper, we develop an item response model that allows us to scale multiple questions about spending preferences together in order to measure individuals' overall spending preferences. We find that people and candidates generally prefer increased spending on many, but not all, policy areas, consistent with other research in this area. We also find that party polarization in the public is very low, nearly nonexistent. Candidates show more polarization than do citizens, but they are not nearly as polarized on spending as they are on measures of overall ideology.

Our model has several advantages over previous attempts at scaling spending preferences. Whereas some previous papers construct an additive index (Ura and Ellis, 2012; for example) or make assumptions about which items scale and which do not (Jacoby, 2000), our method allows issues to be related to overall spending preferences at different strengths. We show that it is in fact the case that some spending on some issues (like education) is very closely related to overall spending preferences, whereas spending on other issues (like space) is not. Furthermore, this method is flexible enough to indicate whether an issue (defense spending, in our case) is actually negatively related to overall spending preferences.

The flexibility of the model allows us to jointly scale various actors as long as they answer the same questions. In our analysis, we concentrated on the public and congressional candidates, but this could be extended to, for example, interest groups, so long as they answer survey questions. Placing interest groups in the same

[28] It is also possible that some kinds of individuals have a larger or smaller error variance. While a full, systematic examination of that possibility is outside the scope of this analysis, previous research suggests that citizens actually respond quite similarly to policy regardless of their level of political sophistication (Enns and Kellstedt, 2008).

space would, subject to some assumptions, allow researchers to compare spending preferences of these groups with their members and each other. Additionally, our model is flexible enough to apply across different years and different question sets. A final advantage of this method is that it generates measures of uncertainty, something not possible with a simpler indexing method.

We take advantage of this to include candidates' spending preferences as well as individuals' in order to study polarization at the elite and mass level. We find that citizens exhibit very low levels of polarization by partisanship. Congressional candidates are more polarized than citizens, but their level of polarization on spending preferences is quite a bit lower than what is seen in most general policy-based ideology estimates. This suggests that, while voters often face choices in terms of overall ideology between Democratic candidates who are much more liberal and Republican candidates who are much more conservative than their own views as shown by Bafumi and Herron (2010), the choices between candidates in terms of spending tend to be less stark.

Our model also includes a point at which individuals who respond "about right" are estimated to be indifferent over increasing or decreasing spending.[29] In other words, individuals are acting as if policy is located there. According to our analysis there is strong sentiment among both citizens and Democratic candidates for higher levels of spending than currently exist in most areas. Republican candidates also usually prefer more spending, though less strongly than Democratic candidates. In a sense this represents a puzzle — if most citizens and most candidates from both parties would prefer to see spending in an area increased, why doesn't it happen? The likely explanation comes from the common tendency to want have one's cake and eat it too. Voters typically desire, and candidates often promise, more spending, lower taxes, and smaller deficits despite the mathematical impossibility of many of these claims (see also Citrin, 1979). In light of this, one might dismiss these results as simply reflecting "cheap talk" by candidates about what policies they would support if elected. But the fact that candidates find it advantageous to support these types of policies highlights the importance of studying the spending positions of candidates and the spending preferences of citizens.

One limitation of our modeling strategy compared to previous strategies is that it is more difficult to make comparisons across time. For example, we cannot estimate the model across years nor directly compare ideal points across years without making implausible assumptions.[30] So we cannot directly compare how our model relates to the "mood" measure developed by Stimson (1991), for example. What we can compare with some additional assumptions are differences within years over time. So, for example, if we were interested in comparing the difference between Democratic and Republican spending preferences over time using our modeling strategy, we could estimate the model separately for each year, compute the average distance between Democrats and Republicans, and compare these numbers over years. This would give us insight into the degree to which the parties are polarized with regard to spending preferences and how that polarization varies over time. Similar methods could be used to study various other phenomena with regard to spending preferences such as the differences between candidates and voters, or winning candidates and losing candidates. Finally, our approach could be used to assess the degree to which the estimated policy locations respond to significant changes in spending levels that result from major legislation or other things.

## Appendix A. 2014 GSS question wording analysis

In addition to the questions listed in Table 1, half of the respondents in the 2014 GSS were randomized to receive alternate question wordings for eleven of these items. This survey experiment provides a unique opportunity to examine the dependence of our model's estimates on the specific wording of questions used to estimate the model.

Table A.4 lists these alternate question wordings. Some of these changes (e.g. "space exploration" versus "space exploration program") seem likely to be inconsequential. Others, however, such as "welfare" versus "assistance to the poor" seem more likely to produce estimates that meaningfully differ between the two wordings.

**Table A.4**
Alternative question wordings

| | | | Response Percent | | |
|---|---|---|---|---|---|
| Policy | Variable | Wording | too much | about right | too little |
| Foreign Aid | nataidy | assistance to other countries | 74.1 | 20.4 | 5.5 |
| Defense | natarmsy | national defense | 31.5 | 36.1 | 32.4 |
| Big cities | natcityy | assistance to big cities | 38.4 | 41.7 | 19.9 |
| Crime | natcrimy | law enforcement | 13.7 | 39.2 | 47.1 |
| Drugs | natdrugy | drug rehabilitation | 13.9 | 33.8 | 52.3 |
| Education | nateducy | education | 7.5 | 17.7 | 74.8 |
| Environment | natenviy | the environment | 12.2 | 26.9 | 60.9 |
| Welfare | natfarey | assistance to the poor | 11.8 | 24.4 | 63.8 |
| Health | nathealy | health | 20.7 | 22.2 | 57.1 |
| Race | natracey | assistance to Blacks | 24.1 | 46.9 | 29.0 |
| Space | natspacy | space exploration | 32.6 | 42.5 | 24.9 |

In order to understand the effects of these different question wordings, we estimate our model treating each question wording as a different item. This produces discrimination parameter and cutpoint estimates for twenty-nine items (the standard wordings for the same eighteen items considered in section 3 as well as the alternate wordings for eleven items). A key question will be how these item's parameters, in particular $\beta_j$ and $p_j$ differ across the two phrasings.

[29] In fact, the model is general enough to be applied to many situations with an ordered trichotomous response. For example, survey questions with the common "like", "neutral", "dislike" response option would also be good fits for this model, assuming multiple questions that all rely on some latent trait.

[30] For example, we could assume that either the distribution of ideal points or the locations of spending policies are constant across time, both of which seem problematic.

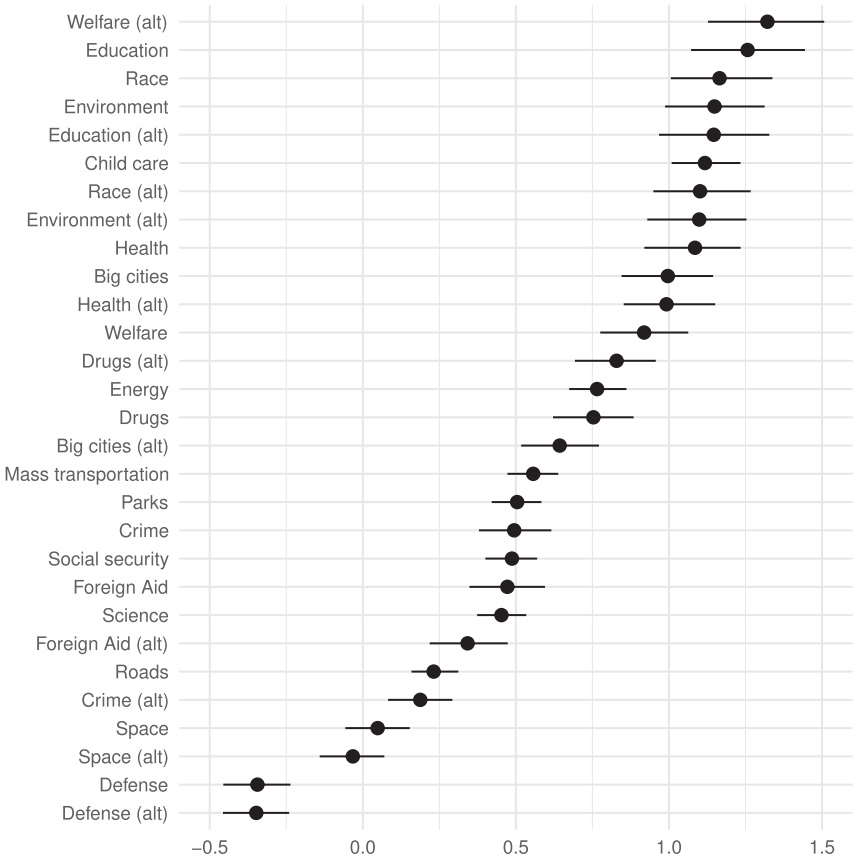



**Fig. A.12.** Discrimination Parameter Estimates, 2014 GSS with Alternate Wordings.

Figure A.12 is comparable to Fig. 1 and shows the estimated discrimination parameters. Estimates for the standard wordings, plotted in black, are quite similar to those in Fig. 1. Furthermore, the discrimination parameter estimates for the alternate wordings are similar for eight out of the eleven items. The three items for which 95% HPDs differ between the two versions are cities, welfare, and crime.


**Fig. A.13.** Estimated cutpoints, 2014 GSS with Alternate Wordings.

Figure A.13 is comparable to Fig. 2 in that it plots the estimated cutpoints for each policy. The estimated cutpoints do not overlap for six policies: welfare, race, health ($\kappa_1$ only), drugs ($\kappa_2$ only), big cities, and foreign aid ($\kappa_1$ only).

Because of the large number of parameters estimated (for example, 18 $\beta$s, 36 $\kappa$s, and 2,538 ideal points for the 2014 GSS data), it is virtually impossible to visually inspect the traceplots for each parameter to asses convergence. We did manually inspect many traceplots, none of which indicated a problem with convergence.

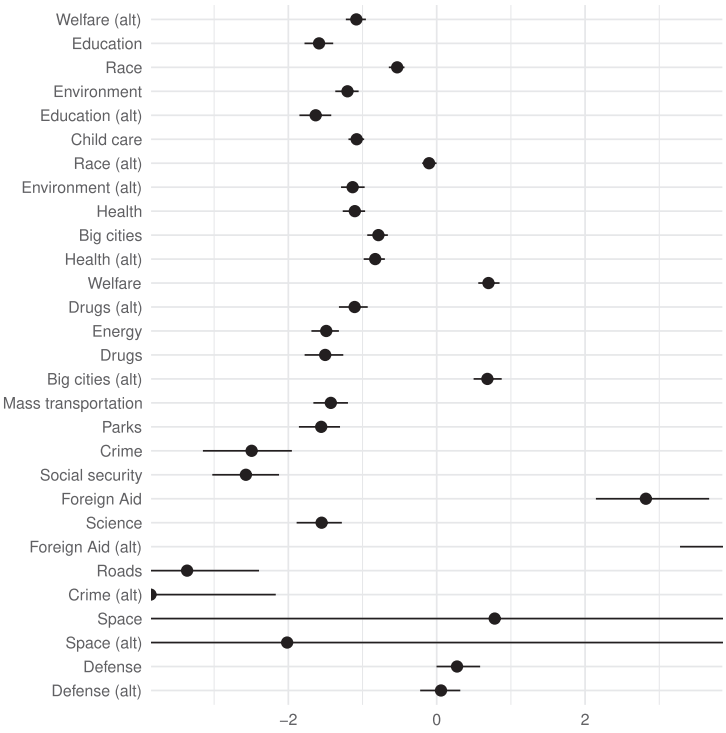



**Fig. A.14.** Ideal Point and Policy Position Estimates, 2014 GSS with Alternate Wordings.

Figure A.14 plots the estimated policy positions for standard and alternate wordings for items in the 2014 GSS. The policy with the strongest evidence of divergence between the two wordings is welfare. This is unsurprising given that dramatically different levels of support for "welfare" and for "aid to the poor" are often cited as classic examples of question wording effects. Figure A.14 shows that the estimated position of spending on "aid to the poor" is well to the left of that for "welfare," with the former estimated to the left of zero and the later to the right. The policy positions for the two wordings of the race item also showed clear, albeit relatively small, differences. Even though the wordings on this issue — "improving the condition of Blacks" versus "assistance to Blacks" — do not seem dramatically different. Finally, spending on cities, for which the standard and alternate wordings, respectively, are "solving the problems of big cities" and "assistance to big cities," show fairly large differences.

Overall, the model's estimates appear fairly robust to changes in wording, but with some important exceptions. This might be thought to imply that the spending preferences of ordinary Americans are meaningful things, at least when measured based on expressed relative preferences.

## Appendix B. convergence and diagnostics

This section reports convergence and diagnostic tests for the 2014 GSS sampler without alternative questions (the results presented in section 3) and for the run with the 1998 GSS and NPAT data jointly scaled (results in section 4).

We also inspected the diagnostic suggested by Geweke (1991). Again, the large number of parameters prevents us from manually inspecting each parameter. A density plot of the $z$ statistics shows that they look very similar to a standard normal distribution, which is what we would expect if the sampler had converged to its stationary distribution. Furthermore, we reject roughly five percent of the Geweke tests across all parameters—what we should expect if the null hypothesis of convergence is actually true. Similarly, the convergence test statistic proposed by Heidelberger and Welch (1983) also suggests convergence. All but 13 of the 2592 parameters estimated in the 2014 analysis pass this convergence test. Given the large number of parameters, this tiny fraction of rejections (around one half of one percent of parameters) does not seem problematic. The results for the Heidelberger and Welch test for the 1998 analyses is similar, with only 25 out of 3360 parameters failing the convergence test (less than one percent of parameters).

Finally, we also ran five separate chains of the sampler for both the 2014 and 1998 datasets starting from over-dispersed initial values and calculated the convergence diagnostic proposed by Gelman and Rubin (1992). For both datasets, the test statistic was well below the conventionally used cutoff of 1.1 (the largest value across all parameters and both datasets was 1.03) after a short number of iterations, implying that our burn-in period of 100,000 iterations was more than enough.

Effective sample sizes are large for all parameters in both model runs. In the 2014 run containing the GSS data, the minimum effective sample size we find is 6,054. Similarly, the 1998 run with both the GSS and NPAT data yields a minimum effective sample size of 3,662.

## References


Abramowitz, A., Saunders, K., 2008. Is polarization a myth? J. Polit. 70 (2), 542–555.

Ansolabehere, S., Meredith, M., Snowberg, E., 2012. Asking about numbers: why and how. Polit. Anal. 21 (1), 48–69.

Ansolabehere, S., Rodden, J., Snyder, J.M., 2008. The strength of issues: using multiple measures to gauge preference stability, ideological constraint, and issue voting. Am. Political Sci. Rev. 102 (02), 215–232.

Ansolabehere, S., Snyder, J., Stewart, C., 2001. The effects of party and preferences on congressional roll-call voting. Legis. Stud. Q. 533–572.

Bafumi, J., Herron, M., 2010. Leapfrog representation and extremism: a study of american voters and their members in congress. Am. Political Sci. Rev. 104 (3), 519–542.

Baldassarri, D., Gelman, A., 2008. Partisans without constraint: political polarization and trends in american public opinion. Am. J. Sociol. 114 (2), 408–446.

Bonica, A., 2013a. Database on Ideology, Money in Politics, and Elections. Computer File, public version 1.0.

Bonica, A., 2013b. Mapping the ideological marketplace. Am. J. Political Sci. 58 (2), 367–386.

Bonica, A., 2015. Measuring public spending preferences using an interactive budgeting questionnaire. Res. Polit. 2 (2).

Brandolini, A., Smeeding, T.M., 2006. Patterns of economic inequality in western democracies: some facts on levels and trends. PS Political Sci. Polit. 39 (01) nil.

Branham, J.A., Soroka, S., Wlezien, C., 2017. When do the rich win? Political Sci. Q. 132 (1), 43–62.

Citrin, J., 1979. Do people want something for nothing: public opinion on taxes and government spending. Natl. Tax J. 32 (2), 113–129.

Clinton, J., Jackman, S., Rivers, D., 2004. The statistical analysis of roll call data. Am. Political Sci. Rev. 98 (02), 355–370.

Conover, P.J., Feldman, S., Knight, K., 1986. Judging inflation and unemployment: the origins of retrospective evaluations. J. Polit. 48 (3), 565–588.

Downs, A., 1960. Why the government budget is too small in a democracy. World Polit. 12 (04), 541–563.

Ellis, C., Stimson, J., 2012. Ideology in America. Cambridge University Press.

Enns, P., Kellstedt, P., 2008. Policy mood and political sophistication: why everybody moves mood. Br. J. Political Sci. 38 (03), 433–454.

Erikson, R., MacKuen, M., Stimson, J., 2002. The Macro Polity. Cambridge University Press, New York.

Fiorina, M.P., Abrams, S.J., 2008. Political polarization in the american public. Annu. Rev. Political Sci. 11, 563–588.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 457–472.

Geweke, J., 1991. Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments, vol 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN, USA.

Gilens, M., 2001. Political ignorance and collective policy preferences. Am. Political Sci. Rev. 95, 379–396.

Gill, J., 2008. Is partial-dimension convergence a problem for inferences from mcmc algorithms? Polit. Anal. 16 (2), 153–178.

Hansen, J.M., 1998. Individuals, institutions, and public preferences over public finance. Am. Political Sci. Rev. 92 (3), 513–531.

Heckman, J.J., Snyder Jr., J.M., 1996. Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators. Tech. rep.. National Bureau of Economic Research.

Heidelberger, P., Welch, P., 1983. Simulation run length control in the presence of an initial transient. Operations Res. 31 (6), 1109–1144.

Holbrook, T., Garand, J.C., 1996. Homo economus? economic information and economic voting. Political Res. Q. 49 (2), 351–375.

Jacoby, W., 1994. Public attitudes toward government spending. Am. J. Political Sci. 38 (2), 336–361.

Jacoby, W., 2000. Issue framing and public opinion on government spending. Am. J. Political Sci. 44 (4), 750–767.

Jacoby, W., 2008. Comment: the dimensionality of public attitudes toward government spending. Political Res. Q. 61 (1), 158–161.

Jessee, S., 2009. Spatial voting in the 2004 presidential election. Am. Political Sci. Rev. 103 (01), 59–81.

Jessee, S., 2016. (how) can we estimate the ideology of citizens and political elites on the same scale? Am. J. Political Sci. 60, 1108–1124.

Keith, B., Magleby, D., Nelson, C., Orr, E., Westlye, M., 1992. The Myth of the Independent Voter. University of California Press.

Krimmel, K., Rader, K., 2017. The Federal Spending Paradox: Economic Self-interest and Symbolic Racism in Contemporary Fiscal Politics. American Politics Research, 1532673X1770122.

Lauderdale, B., 2010. Unpredictable voters in ideal point estimation. Polit. Anal. 18 (2), 151–171.

Lax, J.R., Phillips, J.H., 2012. The democratic deficit in the states. Am. J. Political Sci. 56 (1), 148–166.

Lelkes, Y., 2016. Mass polarization: manifestations and measurements. Public Opin. Q. 80 (S1), 392–410.

Levendusky, M., 2009. The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans. University of Chicago Press.

McCright, A., Dunlap, R., 2011. The politicization of climate change and polarization in the american public's views of global warming, 2001-2010. Sociol. Q. 52 (2), 155–194.

Orriols, L., 2010. Social spending, partisan bias and vote choice. evidence from the u.s. Elect. Stud. 29 (1), 54–65.

Pacheco, J., 2013. The thermostatic model of responsiveness in the american states. State Polit. Policy Q. 13 (3), 306–332.

Plummer, M., 2003. JAGS: a program for analysis of bayesian graphical models using gibbs sampling. In: Proc. 3rd Int. Workshop Distributed Stat. Comput.. Vol. 124. Vienna, p. 125.

Plummer, M., 2015. Rjags: Bayesian Graphical Models Using MCMC.

Poole, K., Rosenthal, H., 2000. Congress: a Political-economic History of Roll Call Voting. Oxford University Press.

Poole, K., Rosenthal, H., 2011. Ideology and Congress. Transaction Publishers.

Poole, K.T., Rosenthal, H., 1991. Patterns of congressional voting. Am. J. Political Sci. 35 (1), 228.

Rasinski, K.A., 1989. The effect of question wording on public support for government spending. Public Opin. Q. 53 (3), 388–394.

Richman, J., 2011. Parties, pivots, and policy: the status quo test. Am. Political Sci. Rev. 105 (01), 151–165.

Schneider, S., Jacoby, W., 2005. Elite discourse and american public opinion: the case of welfare spending. Political Res. Q. 58 (3), 367.

Schuman, H., Presser, S., 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Sage.

Shor, B., McCarty, N., 2011. The ideological mapping of american legislatures. Am. Political Sci. Rev. 105 (03), 530–551.

Soroka, S., Wlezien, C., 2008. On the limits to inequality in representation. PS Political Sci. Polit. 41 (02), 319–327.

Soroka, S., Wlezien, C., 2010. Degrees of Democracy. Cambridge University Press.

Stimson, J., 1991. Public Opinion in America: Moods, Cycles, and Swings, vol 14. Westview Press.

Theriault, S., 2006. Party polarization in the us congress: member replacement and member adaptation. Party Polit. 12 (4), 483–503.

Theriault, S., 2013. The Gingrich Senators. Oxford University Press.

Ura, J.D., Ellis, C.R., 2008. Income, preferences, and the dynamics of policy responsiveness. PS Political Sci. Polit. 41 (04), 785–794.

Ura, J.D., Ellis, C.R., 2012. Partisan moods: polarization and the dynamics of mass party preferences. J. Polit. 74 (01), 277–291.

Webster, S., Abramowitz, A., 2017. The Ideological Foundations of Affective Polarization in the u.S. Electorate. American Politics Research.

Welch, S., 1985. The “more for less” paradox: public attitudes on taxing and spending. Public Opin. Q. 49 (3), 310.

Williams, L., 2015. Estimating the defense spending vote. Elect. Stud. 39, 243–255.

Wlezien, C., 1995. The public as thermostat: dynamics of preferences for spending. Am. J. Political Sci. 39, 981–1000.

Wlezien, C., Soroka, S., 2012. Political institutions and the opinion–policy link. West Eur. Polit. 35 (6), 1407–1432.