Spatial Voting in US Presidential Election

Quantitative Methods 2020: Final Data Essay

Omid Soroush

22 Dezember, 2020

1 Introduction

Three main theories try to explain the voting behavior of citizens based on different factors: first, the sociological model of voting behavior thinks of voting behavior as group behavior. For the proponents of this theory, long-term effects of voting such as group-membership and identity are very important (Paul F. Lazarsfeld, und Lipset & Rokkan). Second, the Rational-Choice-Theory assumes that voters are utility maximizers, and vote for the party or candidate which brings them the maximum utility (Anthony Downs 1957). Short-term effects are considered to be decisive. Finally, the psychological theory of voting combines both long-term and short-term effects, and states that voters behavior is affected by long-term effects such party identification and short-term effects such as candidates evaluation and issue-orientation. Current study builds up on the assumptions of the last theory.

This paper focuses on proximity voting which is one of the most famous theories of voting. Spatial or proximity voting is the idea that candidates and voters have positions in a policy space and that these positions determine the voter's preferences. This means that voters decide based the the principle of smallest distance between their preferred position regarding a policy and the position of the existing parties in an election. In other words a voter compares his position with the position of all the parties regarding the important issues, and at the end votes for the party which is nearest to him. The current paper seeks to empirically analyze the following two hypotheses derived from the spatial voting assumption:

H1: *party identification moderates the effect of ideological self-placement: the expected relationship that the more Republican (more liberal) voters' ideological views, the higher their probability of voting for the Republican (Democrat) candidate is stronger for independents than it is for either Democrats or Republicans.*

H2: *the effect of ideological self-placement has a "pure" spatial effect among the highly informed independents: on average, members of this group are indifferent between the two presidential candidates when positioned exactly at the midpoint of the candidates' average perceived positions.*

2 Data and description

Data from the 2008 US presidential election are used to test these arguments. The data come from the post-election wave of the *American National Study 2008* (ANES 2008). The dependent variable is vote choice coded 1 for Obama, and 0 for McCain. To test the first hypothesis, Voters' ideological self-placement is used as the main independent variable. It runs from 1 (liberal) to 7 (Republican). Party identification is used as the moderator of the effect of ideology on the vote choice. It distinguishes between Democrats, Independents, and Republicans. The model for the first hypothesis controls for respondents' race and income. Race distinguishes between White, Black/African-American, and Mixed. Income is coded 1 if a respondent's family income is above the national median income; and coded 0 otherwise. To test the Second hypothesis, in addition to voter's ideological self-placement as main independent variable, I use respondents' political knowledge as the moderator of the effect of ideology on vote choice. Political knowledge is coded 1 if a respondent correctly identified the majority parties in US House of Representatives and Senate prior to the election; coded 0 otherwise. Next, summary statistics of the above mentioned variables and some more characteristics are presented to get a better idea of the distribution of respondents in this sample.

Table 1: Summary Statistics

	Mean	Median	St.dev	Min	Max	CI_95
Self-placement	4.15	4	1.63	1	7	4.07 - 4.23
Obama's placement	3.15	3	1.7	1	7	3.07 - 3.24
McCain's Placement	5.08	6	1.71	1	7	5 - 5.17
Like Obama	73.07	85	28.07	0	100	71.67 - 74.47
Like McCain	37.46	40	28.33	0	100	36.05 - 38.88
Age	48.23	48	16.81	17	90	47.39 - 49.07

Table 1 summarizes the distribution of the continuous variables in the sample. Respondents' age ranges from 17 to 90 years old. The average ideological self-placement is 4.15. While Obama is on average positioned at 3.1 on the left-right scale, McCain is positioned at 5.1 on this scale. It is worth noting that Obama is much more popular among the respondents than McCain.

Figure 1 shows the distribution of ideology for respondents as well as for both candidates. Respondents have a more evenly distributed ideology than the two candidates. While Obama is commonly placed at 2 by a large number of respondents, McCain is positioned at 6 on the scale from 1 to 7.

Next, Figure 2 shows the distribution of categorical features. Obama has a greater vote share in this sample than McCain. Females are more than males and respondents without college degree are more than those with college degree.

Figure 3 shows the distribution of vote over gender, political knowledge, race and party identification. Plot

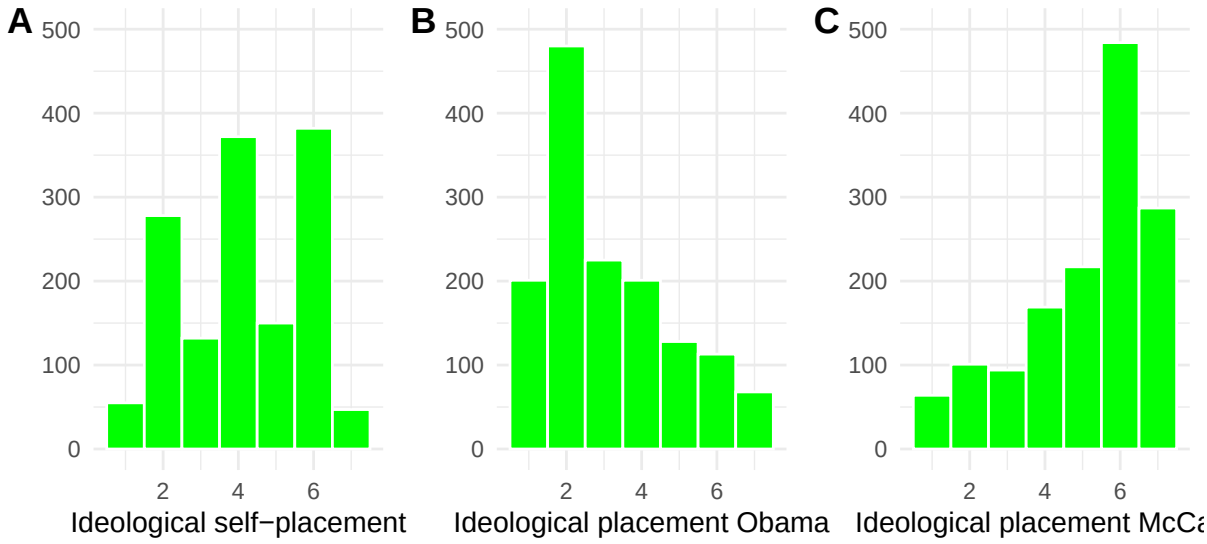

Figure 1: Distribution of ideology

A shows that females voted more for Obama and males for McCain. Similarly, politically knowledgeable people cast their ballots more in favor of Obama than McCain. As expected, Democrats supported Obama; and Republicans McCain. Independents seems to have supported both equally. Distribution of votes over race (Plot D) is interesting. Whites tend to disproportionately support McCain, while Blacks favored Obama.

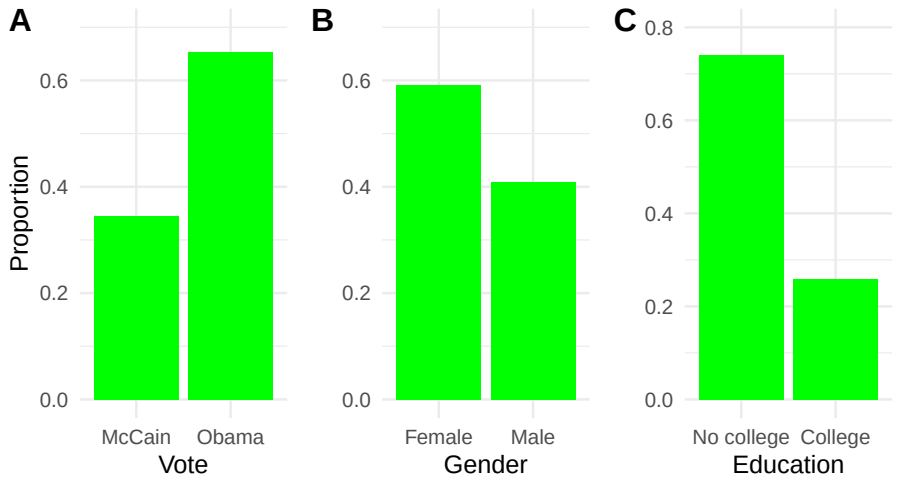

Figure 2: Distribution of categorical features

Figure 4 illustrates the distribution of votes over income and age. Plot A shows that people with a household income above the national median income favored Obama, while those under the median are in favor of McCain. Finally, respondents who supported Obama have a lower median age than those who favored McCain. Next section discusses the methodology.

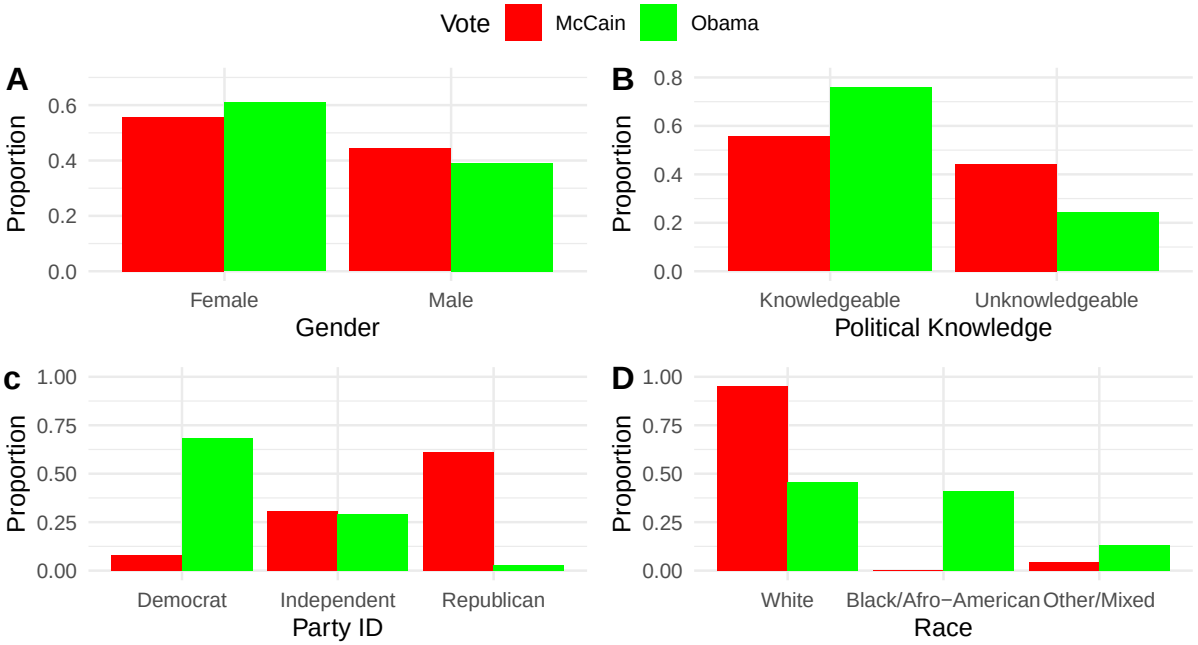

Figure 3: Distribution of vote over gender, education, race, and party ID

3 Methodology

Generalized Linear Model (GLM) with the logistic function is used to estimate the relationship between the outcome and predictor variables. To test the first hypothesis, I define three different models. The first one is the base model with no control variables. The following shows the specification for this model:

$$Log(\frac{p}{1-p}) = \beta_0 + \beta_1 Ideology + \beta_2 Ind + \beta_3 Rep + \beta_4 Ideology * Ind + \beta_5 Ideology * Rep$$

The Second model *(Income)* adds respondents' income, and the third model *(Race)* adds respondents' race as controls. While exploring the data, I found out that these two variables are potentially relevant for this analysis. Furthermore, the voting behavior literature traditionally take these two explanations into account when analyzing the vote choice.

The Second hypothesis is tested using ideology as main predictor and respondents' political knowledge as moderator of this effect. Again, Three different models are conducted to test the Second hypothesis. The base model has the following specification with the Democrats as the reference category:

$$Log(\frac{p}{1-p}) = \beta_0 + \beta_1 Ideology + \beta_2 Ind + \beta_3 Rep + \beta_4 KnowledgeableResp. + \beta_5 Ideology * KnowledgeableResp.$$

```
## Don't know how to automatically pick scale for object of type haven_labelled/vctrs_vctr/double. D
```

Since we are interested in the vote choice of knowledgeable independent voters, the Second model only includes these respondents. Finally, the last model for hypothesis 2 takes the following specification, which controls for income. I do so, because independents of different household income seem to have

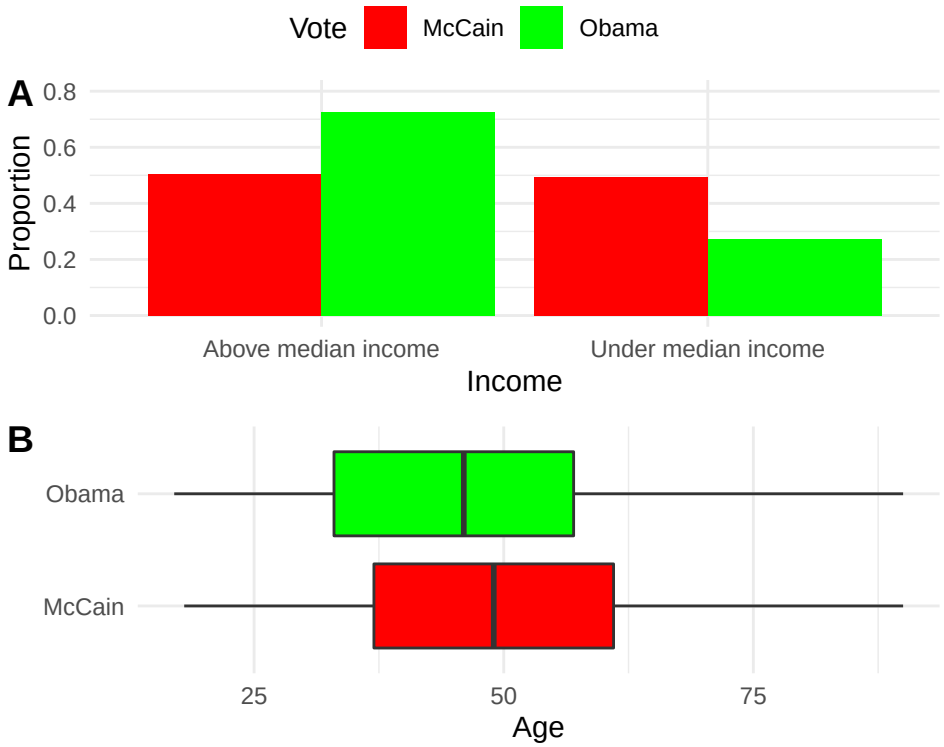

Figure 4: Distribution of vote over income and age

different probability of voting for Obama, when positioned at the mid-point of candidates average:

$$Log(\frac{p}{1-p}) = \beta_0 + \beta_1 Ideology + \beta_2 Ind + \beta_3 Rep + \beta_4 Income + \beta_5 KnowledgeableResp. + \beta_6 Ideology * KnowledgeableResp.$$

Next section presents the results.

4 Results

4.1 Results for the first hypothesis

Table 2 summarizes the result of the GLM models for the hypothesis 1.

The Race model shows that a one unit increase in the self-placement scale yields a 0.33 change in log odds of voting for Obama, given that the respondent is democrat and all other variables are held at a certain value. Since the coefficient is negative, we can say that the probability of voting for Obama decreases in this case. If a respondent is independent, then this change in log odds of voting for Obama is even greater *((-0.334) + (-0.265) = -0.599)*; and for a Republican respondent, this effect is the largest *((-0.334) + (-0.271) = -0.605)*. Since the interpretation of log of adds isn't straight forward, Figure 5 shows the predicted probabilities of voting for Obama from the base model. However, the coefficient of self-placement in the base model is not statistically significant.

We observe that voters with different party identification and different positions on the self-placement

Table 2: Regression Results for Hypothesis 1

	Vote for Obama		
	Base	Income	Race
Ideological self-placement	−0.159	−0.160	−0.334***
	(−0.356, 0.038)	(−0.365, 0.046)	(−0.552, −0.116)
Above median income		−0.527***	
		(−0.866, −0.188)	
Black			4.727***
			(3.274, 6.180)
Mixed			1.441***
			(0.857, 2.024)
Independent	−0.693	−0.679	−0.549
	(−1.785, 0.399)	(−1.806, 0.449)	(−1.752, 0.654)
Republicans	−2.782***	−2.744***	−2.763***
	(−4.423, −1.142)	(−4.519, −0.968)	(−4.550, −0.976)
Self-placement*Ind	−0.337***	−0.329**	−0.265*
	(−0.585, −0.088)	(−0.587, −0.071)	(−0.546, 0.015)
Self-Placement*Rep	−0.425**	−0.407**	−0.271
	(−0.771, −0.078)	(−0.779, −0.036)	(−0.653, 0.111)
Constant	3.384***	3.554***	3.159***
	(2.544, 4.224)	(2.671, 4.437)	(2.252, 4.066)
Observations	1,477	1,391	1,470

Note: *p<0.1; **p<0.05; ***p<0.01

scale have different probabilities of voting for Obama. However, the range of probability for Independents is larger than it is for Democrats or Republicans. This implies that these voters are more variable in their vote choice given their position on self-placement scale. In other words, self-placement of Democrats and Republicans seem to be moderated by their party identification.

Next, Figure 6 shows the predicted probabilities of the model with income as control. We can observe that when considering income, it doesn't make much of a difference for Democrats. However, Independents and Republicans of different income have slightly different probabilities of voting for Obama. The moderation effect of party for Republicans of low income doesn't seem to be very strong because their range of the predicted probability (which is 0.52) is somewhat similar to that of Independents (which is 0.55).

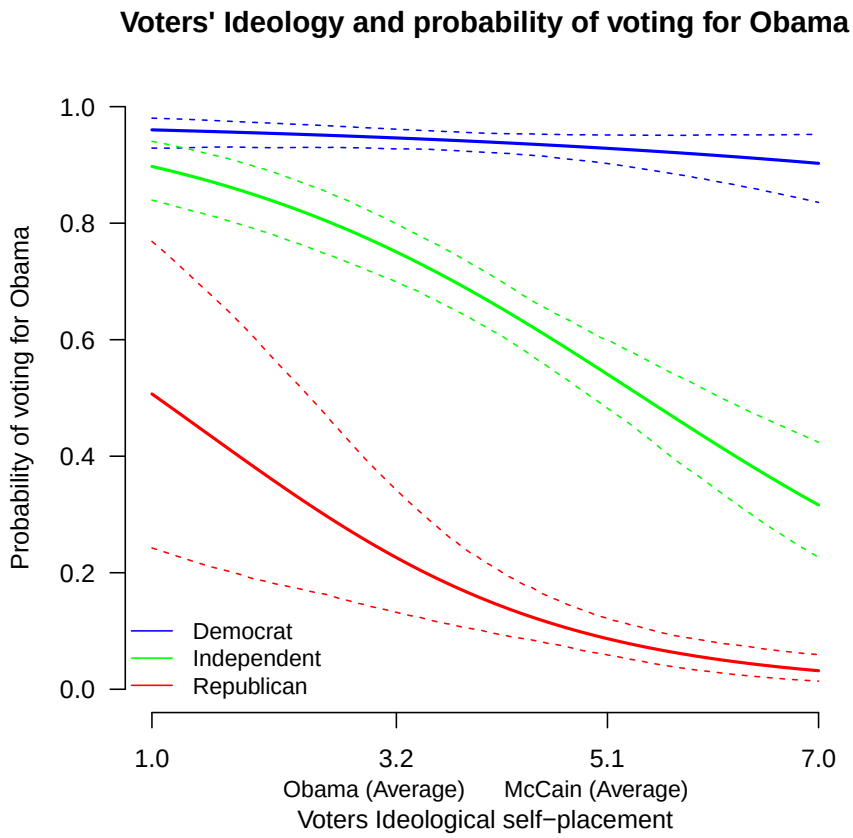

Figure 5: Predicted Probabilities for the base model

Things get even more interesting when we control for race. Figure 7 shows the predicted probabilities for the model with race added as control. Again, we can observe that White Independent respondents are more variable than white Democrats and white Republicans. Confirming the first hypothesis that the effect of ideology on vote choice is greater for Independents. However, when talking about Black and Mixed respondents, the Republicans shows the largest variability in predicted probabilities. This implies that if someone is Black or Mixed and Republican, his ideological self-placement is not very constrained by his party identification. Black and Mixed respondents, in general, shows larger probabilities of voting for Obama than Whites, holding everything else constant (larger intercepts on the y axis).

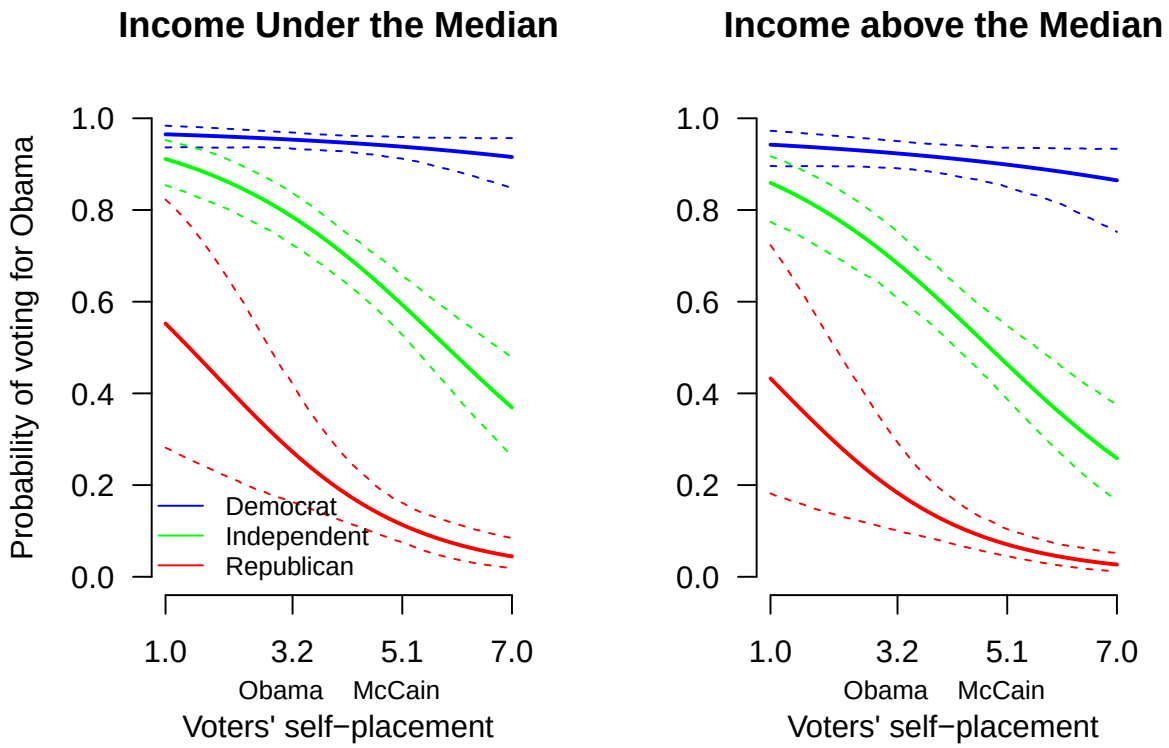

Figure 6: Predicted Probabilities for the model with income added

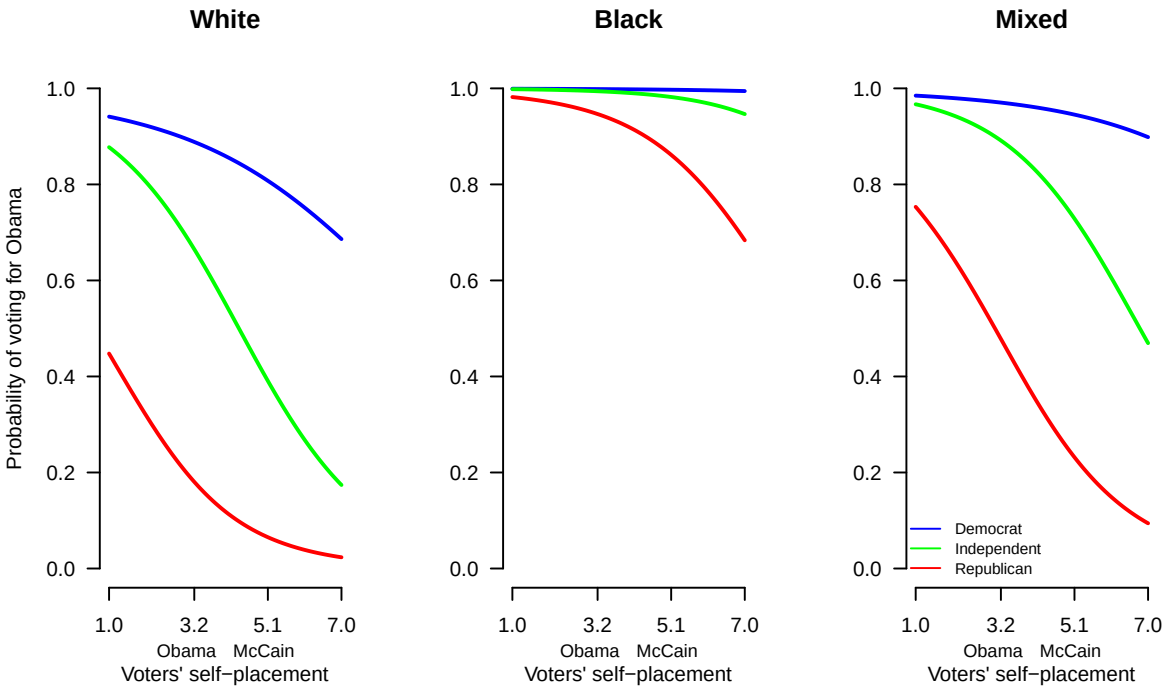

Figure 7: Predicted Probabilities for the model with race added

Table 3: Predicted probabilities (Base model)

Self-placement	Ind	Rep	Pred. Prob	SE	Lower	Upper
1	0	0	0.962	0.012	0.937	0.986
7	0	0	0.906	0.029	0.849	0.964
1	1	0	0.900	0.026	0.850	0.950
7	1	0	0.314	0.049	0.218	0.411
1	0	1	0.504	0.145	0.220	0.789
7	0	1	0.030	0.011	0.008	0.051

Table 3 presents the predicted probabilities of the base model for some interesting scenarios. The difference in probability of voting for Obama between two Democrats, one very liberal and the other very conservative, is only *0.962 - 0.906 = 0.056* (not even 1%). However, this difference for two Independents is *0.900 - 0.314 = 0.586* (almost 60%); and for two Republicans on the two different ends of the self-placement scale, it is *0.504 - 0.030 = 0.474* (47%).

Table 4: Predicted probabilities (Race model)

Self-placement	Black	Mixed	Ind	Rep	Pred. Prob	SE	Lower	Upper
1	0	0	0	0	0.944	0.019	0.907	0.981
7	0	0	0	0	0.695	0.082	0.533	0.856
1	1	0	0	0	0.999	0.000	0.999	1.000
7	1	0	0	0	0.996	0.003	0.990	1.002
1	0	0	1	0	0.882	0.033	0.817	0.947
7	0	0	1	0	0.170	0.039	0.094	0.247
1	1	0	1	0	0.999	0.001	0.997	1.001
7	1	0	1	0	0.959	0.030	0.899	1.018
1	0	0	0	1	0.448	0.157	0.141	0.755
7	0	0	0	1	0.021	0.009	0.004	0.038
1	1	0	0	1	0.989	0.011	0.968	1.010
7	1	0	0	1	0.709	0.168	0.379	1.039

Table 4 shows some interesting scenarios for the Race model. Two White Republicans (one very liberal, the other very conservative) have a difference of almost 23%. The same two White Independents have a difference of 71%. However, it is different if we consider Black respondents. The difference between two Black Republicans (one liberal, the other conservative) is 28%, where is the difference for the same Black Independents is less than 1%. Thus, we can again conclude that party identification moderates the effect of self-placement on vote choice to a lesser degree for Black people than for White people.

4.1.1 Points of indifference for different voters

I calculate the points of indifference for the three groups based on the base model. This is how the model is specified.

$$Log(\frac{p}{1-p}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3$$

The point of indifference is *0.5*. Thus, we add it as the value of p in the specification:

$$Log(\frac{0.5}{1-0.5}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3$$

After evaluating the expression (solving for X_1), we get:

$$X_1 = \frac{\beta_0 + \beta_2 X_2 + \beta_3 X_3}{-\beta_1 - \beta_4 X_2 - \beta_5 X_3}$$

A Democrat should have a position of 21 on the ideological self-placement to be undecided between Obama and McCain. However, the point of indifference for an Independent is 5.4, and for a Republican it is 1.

4.1.2 First differences

Next, the first differences in the predicted probabilities of voting for Obama are calculated for each category of party identification. These first differences are resulted from changing voter's ideal point from the average perceived position of McCain to average perceived position of Obama. Again, I use the base model to calculate the first differences. Figure 8 shows the plotted first differences which is 0.018 for a Democrat, 0.21 for an Independent, and 0.14 for a Republican. The effect is, as expected, strongest for an Independent. This implies that if an Independent changes his ideological position from the mean position of McCain to the mean position of Obama on the left-right scale, he will be 21% more likely to vote for Obama than before.

4.2 Results for hypothesis 2

Table 5 reports the results of the GLM models. The base model shows that if a respondent is unknowledgeable, we expect to see *(exp(-0.260) = 0.77)* 13% decrease in the odds of voting for Obama, for one unit increase on the self-placement scale, holding everything else constant. If the respondent is knowledgeable, we expect to see *(exp(-0.765) = 46)* 54% decrease in odds of voting for Obama, by one unit increase on left-right scale, all else being equal.

Figure 9 plots the predicted probabilities of the three models only for knowledgeable Independents. I stipulate the following from the second hypothesis to test based the the predicted probabilities from the three models above:

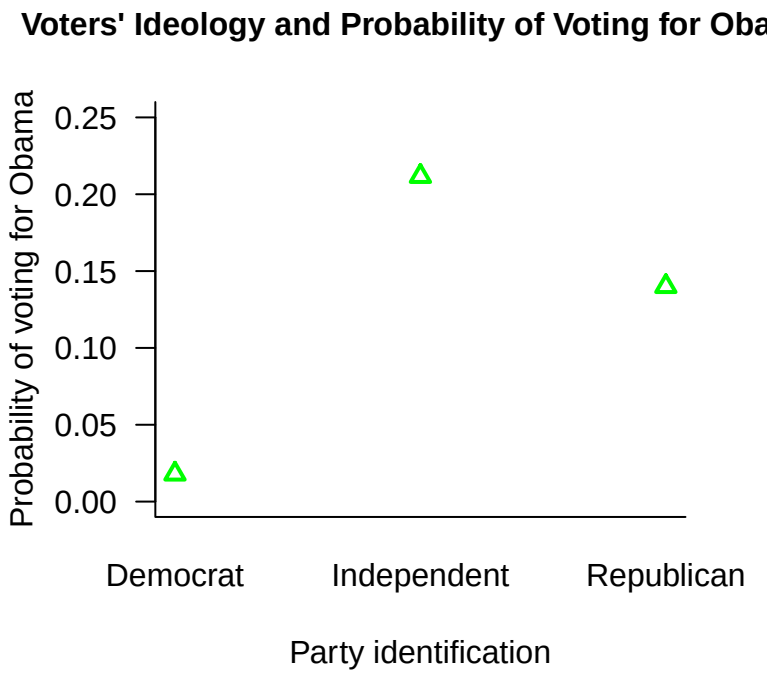

Figure 8: First differences

Table 5: Regression Results for Hypothesis 2

	Base	Vote for Obama Only Ind.	Income
Ideological self-placement	−0.260***	−0.257***	−0.253***
	(−0.393, −0.127)	(−0.437, −0.078)	(−0.391, −0.115)
Independent	−2.092***		−2.037***
	(−2.494, −1.689)		(−2.455, −1.618)
Republican	−4.745***		−4.658***
	(−5.298, −4.193)		(−5.230, −4.086)
Above median income			−0.522***
			(−0.886, −0.157)
Pol. knowledge	1.817***	3.122***	1.908***
	(0.457, 3.177)	(1.005, 5.239)	(0.475, 3.340)
Self-placement*Pol. knowledge	−0.505***	−0.869***	−0.505***
	(−0.792, −0.219)	(−1.335, −0.403)	(−0.805, −0.205)
Constant	3.899***	1.882***	4.005***
	(3.224, 4.575)	(1.090, 2.673)	(3.294, 4.716)
Observations	1,347	396	1,267

Note: *p<0.1; **p<0.05; ***p<0.01

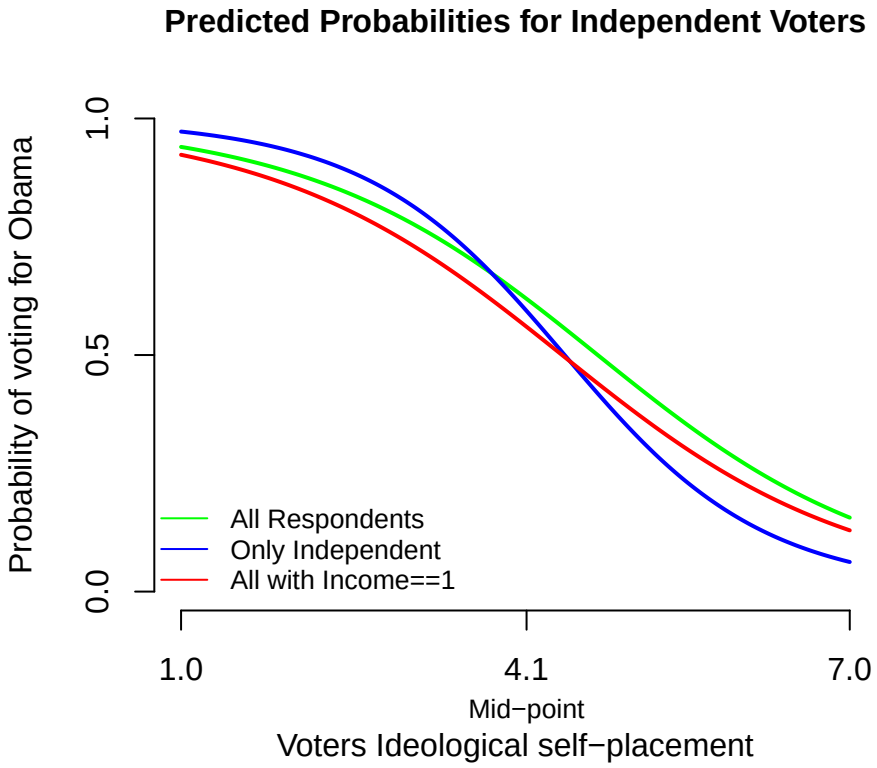

Figure 9: Predicted probabilites from the three models for hypothesis 2

Table 6: Predicted probabilities (Income model H2)

Self-placement	Ind	Rep	Income	Pol. know	Pred. Prob	SE	Lower	Upper
4.1	1	0	0	1	0.683	0.048	0.590	0.776
4.1	1	0	1	1	0.561	0.053	0.456	0.666

Stipulated hypothesis: knowledgeable Independents should have a probability of *0.5*, when positioned at 4.1 on the self-placement scale (candidates mid-point).

Figure 10 shows the predicted point probability and uncertainty for all three models, when the voter is positioned exactly at *4.1*. Starting from the left, the base model (includes all respondents) yields a predicted point probability of *0.61* with uncertainty of *(0.53 - 0.7)* based on 95% CI. This excludes the indifference point of *0.5*. Thus we can not accept the pure spatial assumption regarding the knowledgeable Independents with an Ideology score of 4.1 based on this model. Including only Independents in the model (the blue line), we get a point probability *0.59* with uncertainty *(0.48 - 0.71)*. Based on this model, we can accept the pure spatial assumption (even though the 95% confidence interval is very wide because we have fewer respondents in this model).

Lastly, the red point (calculated only for knowledgeable Independents of above the median income) yields a point estimate of *0.56* and uncertainty range of *(0.45 - 0.66)*. It includes the indifference point of *0.5*. Thus, we cannot reject the pure spatial assumption about the knowledgeable Independents with an income above the median. However, as Table 6 shows knowledgeable Independents with Ideology score of 4.1

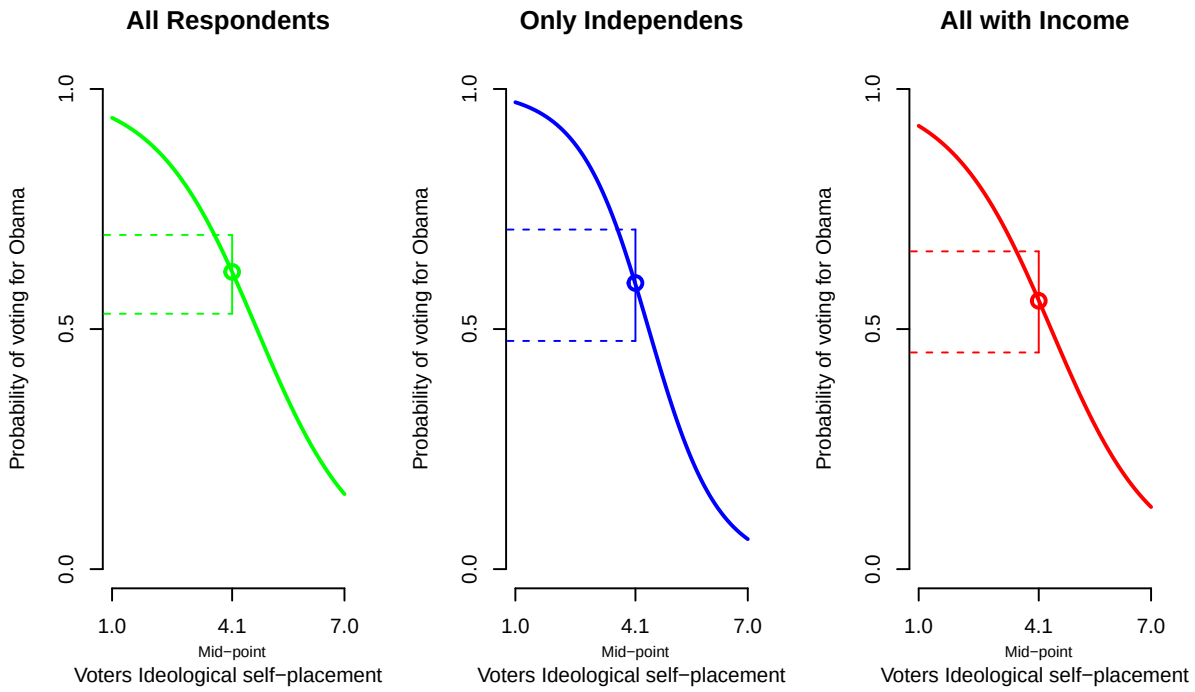

Figure 10: Predicted point probabilities

)

Table 7: Predicted probabilities (Race model H2)

Self-placement	Ind	Rep	Black	Mixed	Pol. know	Pred. Prob	SE	Lower	Upper
4.1	1	0	0	0	1	0.528	0.052	0.426	0.630
4.1	1	0	1	0	1	0.996	0.004	0.989	1.004
4.1	1	0	0	1	1	0.832	0.051	0.732	0.931
4.1	1	0	1	1	1	0.999	0.001	0.997	1.001

and an income under the median are not indifferent between the two candidates. Furthermore, when we control for race, we can observe that only White knowledgeable Independents are indifferent between the two candidates when positioned at the mid-point. Table 7 reports these probabilities for the three different races.

5 Robustness check

This section provide robustness checks for both models conducted for the two hypothesis. In these robustness checks, the same specifications are used as before, though, with different variables. I let the dependent variable to change for the race model of the first hypothesis. Vote choice in the 2008 election to the US House of Representatives is used instead of vote choice for Obama/McCain.

Figure 11 plots the predicted probabilities of voting for a Democrat candidate in the House of Representatives. The points of indifference for White Democrats is 12; for White Independents it is 4.3; and for White Republicans it is 0.23. Thus, these results are robust and support the first hypothesis. In contrast

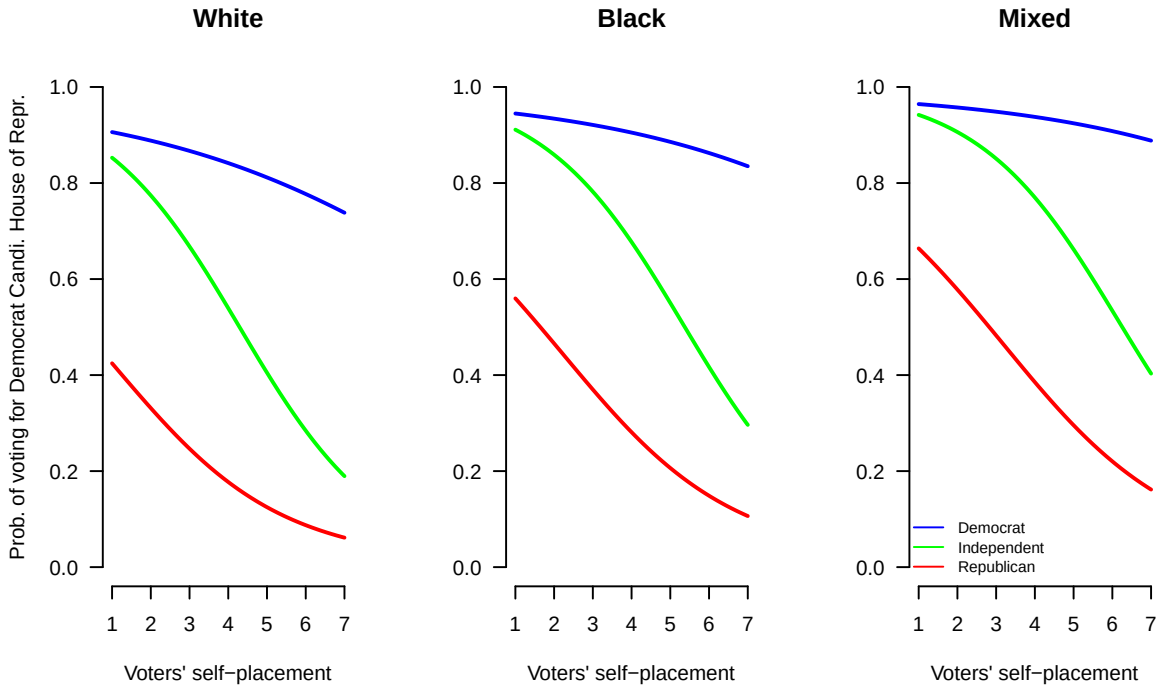

Figure 11: Predicted Probabilities Race model H1 (Robustness Check)

to the results for the presidential election, we observe that Black Independents are also, like White Independents, driven to a greater degree by Ideology. The point of indifference for Black Independents is also closer to the mid-point of the self-placement scale than is that of Black Democrats and Republicans.

To check the robustness of the second hypothesis, I change the political knowledge variable to college degree in the two models (Only Independent and the model with income as control). The college degree is coded 1 if a respondent has a four year college degree; and coded 0 otherwise. Figure 12 plots the predicted probability lines and point estimates for the robustness estimation. We can observe that both of the models don't hold anymore within the 95% confidence interval.

6 Conclusion

First, the current paper examines whether the effect of ideological self-placement is stronger for Independent voters than for Democrats or Republicans (whether party ID moderates the effect of ideology). I find evidence that this assumption is, in particular, true about the White Independents. Black and Mixed Independents are not more indifferent between two presidential candidates (one Democrat, the other Republican) than the Black and Mixed Republicans. This implies that party identification only moderates the effect of ideology on the vote choice for white Republicans and all Democrats based on the estimation of this paper. However, ideology has a greater effect for all Independents in the election of the US House of Representatives. I find evidence that this group is more indifferent in their vote choice between two Democrat and Republican Candidates than are Democrats or Republicans.

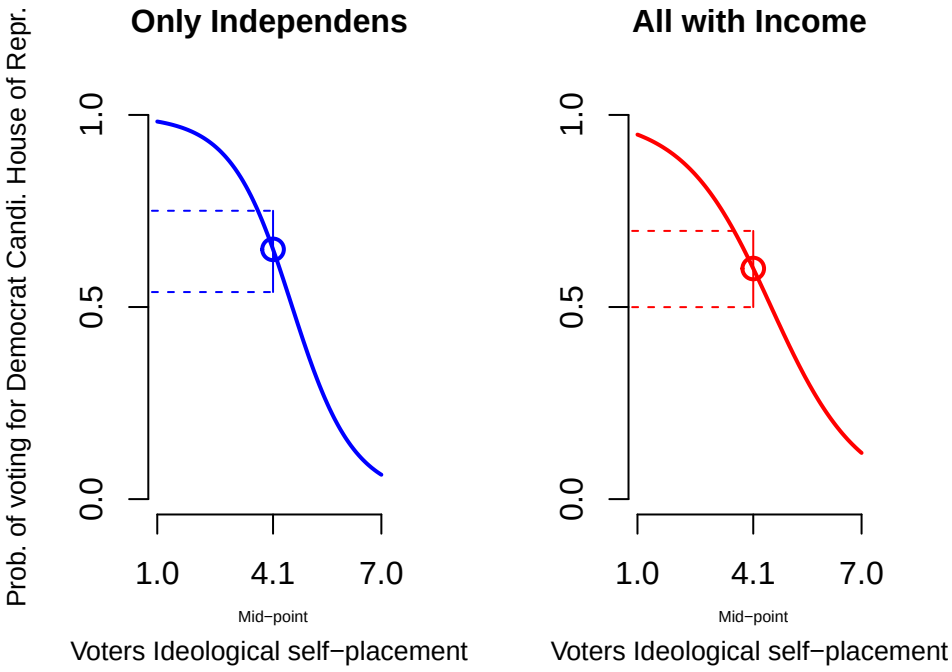

Figure 12: Predicted Probabilities H2 (Robustness check)

Second, I investigate whether ideology has a pure spatial effect among knowledgeable Independents. I find evidence supporting this assumption, in particular, for White knowledgeable Independents and knowledgeable Independents above the national median income in the presidential elections. However, the pure spatial effect of ideology vanishes when we look at the elections for the House of Representatives.

The moderating effect of party identification can further be studied, particularly in other parts of the globe than in the US. I believe that USA are a very polarized country in terms of party belonging, and this effect may be overemphasized. Also, multi-party systems need to be taken into account.