UMR 8184

THéorie Économique, Modélisation et Applications

THEMA Working Paper n°2019-10
Université de Cergy-Pontoise, France

Empirical foundation of valence using Aldrich-McKelvey scaling

Fabian Gouret

September 2019

Empirical foundation of valence using Aldrich-McKelvey scaling *

Fabian Gouret†

September 5, 2019

Abstract

This paper uses data from the 2004 pre-election survey of the American National Election Study to test empirically different ways of incorporating a valence parameter into a Downsian utility function. We call particular attention to the problem of interpersonal incomparability of responses to the liberal-conservative scale, and use the Aldrich-McKelvey's pathbreaking method to obtain accurate distances between respondents and candidates, the key regressors. We find that the utility function the most supported by the empirical evidence, the intensity valence utility function, is the one which permits to make the better predictions for the 2004 presidential election. We also consider counterfactual analyses wherein we test if Bush, the candidate with the highest intensity valence, has dominant strategies which would have insured him to obtain a majority of the popular vote. According to the theory, it is known that the candidate with the highest intensity valence does not have such dominant strategies if the distribution of voters in the policy space is too heterogenous. Nevertheless, we show the distribution of voters in 2004 is sufficiently homogenous for Bush to have dominant strategies.
JEL Classification: D72, C81

Keywords: spatial models of voting, valence, survey, Aldrich-McKelvey scaling.

---

*I am very grateful to William Kengne and Alfonso Valdesogo for discussions on bootstrap, as well as Vladimir Apostolski, François Belot, Thomas Brodaty, Mathieu Martin, Agustín Pérez-Barahona and Julien Vauday. All remaining errors are mine, and I acknowledge the financial support of Labex MME-DII (ANR-11-LBX-0023-01) and a "Chaire d'Excellence CNRS".

†Théma, Université de Cergy-Pontoise, 33 Bvd du Port, 95011 Cergy-Pontoise Cedex, France (Email: fabiangouret@gmail.com). Homepage: https://sites.google.com/site/fabgouret/

1 Introduction

Since Downs (1957), the spatial theory of voting has predominated in formal political science. Such a theory fails to take into account that candidates often possess valence characteristics, i.e., characteristics unrelated to policy selection and unanimously evaluated by voters (e.g., charisma, office-holding experience). Thus, to add realism into the spatial model, various authors have included an additive valence into the Downsian utility function and explored its implications (e.g., Groseclose, 2001, Hummel, 2010). But alternative ways of introducing a valence parameter exist (e.g., Gouret and Rossignol, 2019, Kartik and McAfee, 2007, Section IV.C.). Ideally, one would like to know if one of these theoretical utility functions is supported by the empirical evidence using pre-election surveys. It is appealing for at least two reasons. First, if a theoretical utility function is empirically founded, one can use it to make a priori better predictions for the election considered. Second, one can use it to make counterfactual analyses. For instance, one can test statistically if a candidate may propose an alternative policy which would have insured him to win for sure. However, few papers have tested statistically the best way to model valence, i.e., how to introduce the valence parameter into a utility function. One exception is Gouret *et al.* (2011) who use a pre-election survey in the French presidential election of 2007; but they do not make any prediction nor any counterfactual analysis.

Furthermore, a key problem to test these utility functions and make predictions or counterfactual analyses is the problem of interpersonal incomparability of responses to issue scales. Indeed, to compute the distances between respondents and candidates (the crucial regressors), it is natural to use survey items which ask respondents to place themselves and candidates on issue scales –typically a liberal-conservative scale. The prevalent practice in most empirical studies is to take these responses at face value. However, and for example, a conservative voter may place a liberal candidate more on the left than do liberal voters to exaggerate the distance between him and this candidate he views unfavorably. If so, taking these responses at face value might bias the computed distances and the final results.

This paper takes this problem of interpersonal incomparability seriously, and uses pre-election data from the American National Election Study (ANES) to test empirically the different ways of incorporating a valence parameter into the Downsian utility function. Then, we show that the utility function the most supported by the empirical evidence permits to make better predictions for the election considered. We also consider counterfactual analyses wherein we test if one candidate has dominant strategies which would have insured him to obtain a majority of the popular vote.

We exploit the Aldrich and McKelvey's (1977) seminal contribution to solve the problem of interpersonal incomparability of responses. Their idea to recover the underlying locations of the candidates and respondents in a common policy space is to treat the reported positions of candidates by a respondent as linear distortions of the true locations of the candidates. The solution is equivalent to a principal components solution for the true locations of candidates, together with least squares estimate of each respondent's distortion parameters. Then, the true locations of the candidates and the bliss points of the respondents are used to compute accurate distances, test different utility functions, make predictions, and conduct counterfactual analyses.

We test three theoretical utility functions: (1) the basic Downsian utility function; (2) the additive valence utility function which has been widely studied in the literature; (3) lastly, the intensity valence utility function studied formally by Gouret and Rossignol (2019) and proposed initially by Gouret *et al.* (2011); see also Kartik and McAfee (2007, Section IV.C.) for a close idea. The intensity valence assumes that all voters agree that one candidate has more ability to implement his announced policy than his opponents. However, and in contrast with the (2) additive valence, the intensity valence has a different impact on the utility of voters according to their position in the policy space. If a candidate is more efficient at implementing his announced policy, it will increase the utility of voters whose bliss points are near this policy, but decrease the utility of those who are too far. For instance, if a conservative policy is implemented, a liberal voter does not want this policy to be implemented intensively.

To test these three theoretical utility functions, we formulate a statistical model which is a system of regression equations. The number of equations in the system is the number of candidates, given that each equation represents the utility if a specific candidate is elected. The different theoretical utility functions imply different testable cross-equation parameter restrictions on the statistical model. Subsection 2.2 presents the method and shows that the Downsian and the additive valence utility functions have testable implications if there are at least two candidates. Nevertheless, this is not the case for the intensity valence utility function: with two candidates, the statistical model would exactly identify the unknown parameters of the intensity valence utility function. To overidentify and thus place testable restrictions on the statistical model, it is necessary to have at least three candidates.

This is the reason why we use data drawn from the 2004 pre-election survey of the ANES. The 2004 survey is the last survey wherein respondents were asked to rate on a 100-point scale their affect toward the three main Presidential candidates: the Democratic Presidential candidate John Kerry, the Republican Presidential candidate George W. Bush and the Independent Presidential candidate Ralph Nader.[1] Thus, the regressands of the econometric specifications are feeling thermometers rather than stated choices. Feeling thermometers are more likely to produce unbiased estimates of the parameters of the utility functions, given that strategic voting might

[1] Since the presidential election of 2008, the ANES has asked respondents to rate only the two main Presidential candidates, i.e., the Democrat and the Republican. The ANES provides on its website a questionnaire utility which lists the years in which a question has appeared. The information concerning the thermometer questions for the candidates are lines 80, 81 and 82 (Part II "Candidate and incumbent evaluations"); see http://isr-anesweb.isr.umich.edu/_ANESweb/utilities/questutility/all.htm.

occur with three candidates.[2]

We find that the intensity valence utility function is the sole utility function which is supported by the empirical evidence. Then, using the estimated parameters of this utility function, the location of the candidates and the distribution of voters in the policy space, we study the source of support for each candidate according to this model. Under the assumption that all the voters who prefer Nader vote strategically, as well as under the assumption that all the voters who prefer Nader vote sincerely, the relative frequencies of vote obtained via the intensity valence model for Bush (51.8 and 50.9 percent, respectively) fit well with reality, given that Bush won the popular vote with 50.73 percent; these results outperform those obtained with the Downsian and the additive valence models. We also study if one candidate has dominant strategies which would have insured him to obtain the majority of the popular vote. Bush is the candidate with the highest intensity valence, but nothing insures theoretically that he has such dominant strategies. Indeed, Gouret and Rossignol (2019) have shown that a low intensity valence candidate may have dominant strategies which insure him to obtain the majority if the distribution of voters in the policy space is too heterogenous. Nevertheless, the distribution of voters in 2004 is sufficiently homogenous for Bush to have dominant strategies which are statistically significant, at least under the assumption that all the voters who would have preferred Nader vote strategically.

Various authors have the same objective as ours: they try to obtain a realistic extended Downsian utility function to make prediction, and that this utility function remains parsimonious enough to understand its theoretical implications (e.g., Adams *et al.*, 2005, Degan, 2007, Schofield *et al.*, 2011). But they consider an additive valence, and do not test statistically if there is a better way to introduce a valence parameter in a utility function.[3] Furthermore, although

[2] A voter votes sincerely when he votes for his most preferred candidate. Conversely, he votes strategically when he decides to vote for his second most preferred candidate because his most preferred one is unlikely to win.

[3] For instance, using stated voting behaviors, Schofield *et al.* (2011, pp.485-486) estimate mixed logit wherein the intercept term associated to one candidate measures his additive valence. Interestingly, they also add party-leader

the problem of interpersonal incomparability of responses to issue scales has been widely recognized since Aldrich and McKelvey (1977), and that their solution is used (e.g., Hollibaugh *et al.*, 2013, Armstrong II *et al.*, 2014), empirical work on valence usually does not deal with this problem.[4] For Adams *et al.* (2005, p.27), one question is "whether to use voter-specific placements of candidates or mean placements when computing the distance". But Armstrong II *et al.* (2014, p.43) note that using the mean placement of each candidate is not a solution either; it is prone to failure since errors are unlikely to cancel out in case of heteroskedasticity. On the contrary, Palfrey and Poole (1987) show via Monte Carlo simulations that the Aldrich-McKelvey method permits to recover an accurate location of the candidates, even with high heteroskedasticity.[5]

We arrange our presentation in the following way. Section 2 recalls the theoretical utility functions that will be tested and explains how to estimate and test them. Section 3 describes the 2004 pre-election survey of the ANES, shows that respondents answer the liberal-conservative scale differently, and uses an Aldrich-McKelvey correction to solve this problem. Section 4 tests the different utility functions considered. Section 5 uses the results to study the support for each candidate according to the intensity valence model, compares the results with the other models, and tests if Bush has dominant strategies which would have insured him to obtain a majority of the popular vote. Following all of this, Section 6 concludes. Some additional results are relegated to various appendixes.

trait indices which reflect valence factors in some specifications.

[4]Degan (2007) is a notable exception. Analyzing the 1968 and 1972 U.S. Presidential elections, she uses the first dimension of the DW-NOMINATE scores in the Senate as an accurate measure of candidates' positions on a liberal-conservative scale. Concerning voters' positions, she estimates a parametric distribution of their positions (rather than point estimates of individual positions) using stated choices and voters' characteristics.

[5]There are also some empirical studies which do not estimate the parameters of the utility function of voters to make predictions and/or counterfactual analyses as we do. They propose some estimations wherein divergence between candidates or from the median voter is explained by some proxies for valence advantage. Incumbency has been a standard (e.g., Burden, 2004, Ansolabehere *et al.*, 2001). More recently, Adams *et al.* (2011) and Stone and Simas (2010) use district expert informants in the 2006 House elections to distinguish between valence which reflects campaign skills or fundraising ability and valence that voters value for their own sake (competence, integrity). These papers do not consider the problem of interpersonal incomparability of responses. One exception is Zakharova and Warwick (2014). Using data from the Comparative Study of Electoral Systems, they show in particular that individuals' valence judgements depend negatively on the distance between them and the parties.

2 Existing utility functions and econometric models

Subsection 2.1 describes the utility functions that will be tested and recalls additional literature on valence. Subsection 2.2 explains how to estimate and test them.

2.1 Theoretical utility functions

We consider an election between M candidates indexed by $j = 1, ..., M$. Each candidate j proposes a policy platform x_j in the unidimensional policy space $\mathbb{R}$. Each voter i has a bliss point $a_i \in \mathbb{R}$ in this policy space. In this basic setting, we are interested in three theoretical utility functions.

The Downsian utility function, the simplest one, considers that the utility of voter i if candidate j is elected is given by:

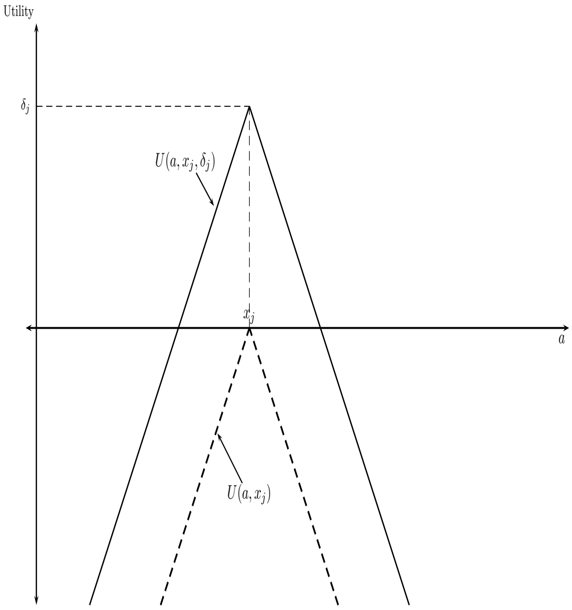
$$U(a_i, x_j) = -|x_j - a_i| \tag{1}$$

That is, the utility of voter i is a decreasing function of the distance between x_j and a_i. Note that Equation (1) assumes that the utility is linear in distance. Various authors consider a quadratic utility, specified as the negative of the squared distance between x_j and a_i (i.e., $U(a_i, x_j) = -(x_j - a_i)^2$). In Subsection 4.3, we will try to distinguish empirically whether a linear distance, a squared distance, or another power function of distance, better represents voters' relative evaluations of candidates. Considering for the moment a utility linear in distance is enough to understand the difference between the different utility functions considered.

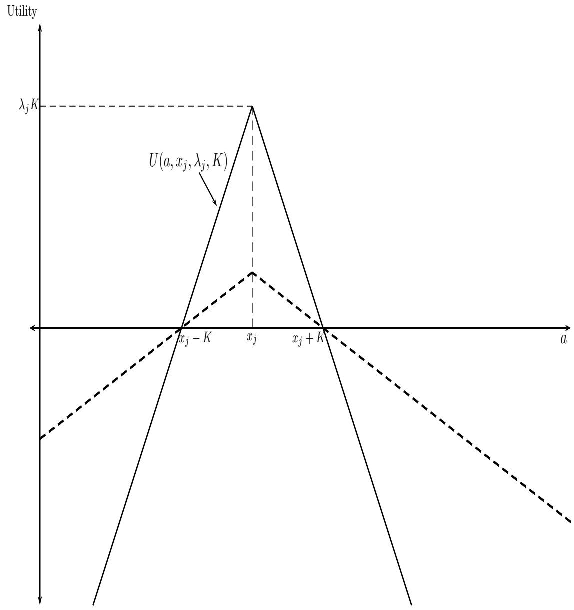
The additive valence utility function adds a candidate-specific parameter $\delta_j \in \mathbb{R}$ to the Downsian utility function:

$$U(a_i, x_j, \delta_j) = \delta_j - |x_j - a_i| \tag{2}$$

If $\delta_j > \delta_{j'}$, candidate j has an additive-valence advantage over candidate j'. As shown in Panel (A) of Figure 1, a higher additive valence δ_j adds the same amount of utility to all voters whatever their bliss point in the policy space. The voter whose bliss point is $a_i = x_j$ obtains the highest level of utility. Various authors have tried to understand the consequences of this simple extension of the Downsian utility function (Ansolabehere and Snyder, 2000, Groseclose, 2001, Dix and Santore, 2002, Aragones and Palfrey, 2002, Evrenk, 2009, Hummel, 2010, Xefteris, 2012, Aragonès and Xefteris, 2012, Xefteris, 2014); see Evrenk (2019) for a review of the literature.[6]



(A) Additive valence



(B) Intensity valence

Figure 1: Additive and intensity valence

The intensity valence utility function has been analyzed formally by Gouret and Rossignol (2019) and proposed initially by Gouret *et al.* (2011) based on a French pre-election survey. The intensity valence supposes that the valence represents the ability of a candidate for implementing a policy. All voters may agree that one candidate will implement more intensively a policy than an opponent, but may be affected differently. A candidate who is more efficient at implementing

[6]Note that several papers also consider an additive valence, but this valence parameter is not fixed. They assume that it depends on campaign spending or a costly effort from the part of a candidate (see, e.g., Ashworth and Bueno de Mesquita, 2009, Carrillo and Castanheira, 2008).

a policy will increase the utility of voters whose bliss points are near this policy, but decrease the utility of those who are too far. More formally, if the distance between voter i's bliss point a_i and candidate j's platform x_j is less than $K \in \mathbb{R}_+^*$, i.e., $|x_j - a_i| < K$, and candidate j is elected, then the higher the intensity valence parameter $\lambda_j \in \mathbb{R}_+^*$ of candidate j, the higher the utility of voter i. However, if the distance between a_i and x_j is higher than K, i.e., $|x_j - a_i| > K$, then the higher the intensity valence parameter λ_j, the lower the utility of voter i, as shown in Panel (B) of Figure 1. The intensity valence utility function is then:

$$U(a_i, x_j, \lambda_j, K) = \lambda_j(K - |x_j - a_i|) \tag{3}$$

Contrary to the additive valence utility function, the intensity valence utility function is not additively separable in distance and valence. A close idea has been proposed by Kartik and McAfee (2007) who investigate the effect of "character". Such a character is similar to an (ex-ante uncertain) additive valence in most of their article, but they highlight at the end of it (Subsection IV.C., p.863) that a preference weight on character may depend on both the platform and a voter bliss point: "a voter with ideal point $[a_i = 1]$ may prefer a candidate with platform $[x_j = 0]$ *not* to have character [...] [T]he same voter may prefer a candidate with $[x_j = 1]$ to in fact have character." More generally, the intensity valence utility function is linked to what Krasa and Polborn (2012) call non-uniform candidate ranking (non-UCR) preferences.[7] Krasa and Polborn (2010), Soubeyran (2009) and Câmara (2012) are some of the few papers wherein competence differentials between candidates give rise to non-UCR preferences, like the intensity valence utility function. However, they assume that their utility functions satisfy the well-known single-crossing property,

[7]If both candidates propose the same policy x, but voter i prefers one candidate because of his characteristics (which may include valence), then voter i has UCR preferences if he also prefers the same candidate when both propose the alternative policy x'. Contrary to the additive valence, the intensity valence violates the UCR-property, as can be shown in Figure 1.

while the intensity valence utility function does not. Indeed, with the intensity valence, the set of voters who prefer the candidate with the lowest intensity valence is always a non-convex set: the candidate with the lowest intensity valence is supported by voters whose ideal points are on both sides of the policy space, as shown in Panel (B) of Figure 1.[8]

2.2 Econometric models

To test the hypotheses of the three utility functions presented in Subsection 2.1, we formulate a seemingly unrelated regressions (SUR) model that contains these hypotheses as restrictions on its parameters.

Consider a survey of N voters, $i = 1, ..., N$, representative of the electorate of an election. Recall that M candidates, $j = 1, ..., M$, compete at this election. In the survey, each voter i is asked to rate his felicity $U_{i,j}$ toward each Presidential candidate j. As we will show in Section 3, the survey also permits to obtain each voter i's bliss point a_i, as well as the actual location x_j of each candidate j. Stacking all M utilities for the ith voter, we get:

$$\begin{bmatrix} U_{i,1} \\ U_{i,2} \\ \vdots \\ U_{i,M} \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} - \begin{bmatrix} d_{i,1} & 0 & \cdots & 0 \\ 0 & d_{i,2} & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & d_{i,M} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{bmatrix} + \begin{bmatrix} \varepsilon_{i,1} \\ \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,M} \end{bmatrix} \tag{4}$$

where $d_{i,j} = |x_j - a_i|$ and $\varepsilon_{i,j}$ is an error term. The M pairs of parameters $\{(\delta_1, \lambda_1), \ldots, (\delta_M, \lambda_M)\}$ could be estimated separately by ordinary least squares (OLS) using the N observations. However, the three utility functions that we will test impose some cross-equation restrictions on the system

[8] Note that the intensity valence model is not the sole model which predicts that the set of voters who support a candidate with less ability is a non-convex set. For example, Miller (2011) proposes a model which combines an additive valence and the candidate's likelihood of changing policy from an exogenous status quo. If the candidate the less able to change the status quo is additive-valence-advantaged, then the set of voters who prefer this candidate is a non-convex set.

of equations (4). Imposing cross-equation constraints is not possible using equation-by-equation OLS, but it is possible using SUR estimation.

As it is standard in SUR models, for a given voter i, the errors may be correlated across equations; that is, $\mathbb{E}\left[\varepsilon_{i,j}\varepsilon_{i',j'}\right] = \sigma_{jj'}$ if $i = i'$ and 0 otherwise. It makes sense here: for example, the utility that a liberal voter will obtain with a liberal candidate give some information about the utility that he will obtain with a conservative candidate.

The three theoretical utility functions that we will test impose some cross-equation restrictions on the SUR model (4). One should have in mind that according to economic theory, the different utility functions considered are unique up to positive affine transformations. The Downsian utility function $U(a_i, x_j) = -|x_j - a_i|$ is equivalent to $U(a_i, x_j) = \delta - \lambda |x_j - a_i|$ for some scalar δ and some scalar $\lambda > 0$ independent of i and j.[9] The values of these scaling parameters δ and λ will depend on the scales used in the questions of the survey; for instance, the regressands measure the felicities toward various candidates on a 100-point scale. Thus, the Downsian utility function implies the following testable restrictions on the unconstrained model (4):

$$\mathrm{H}_0: \ \lambda_j = \lambda \ , \forall j, \text{ and } \delta_j = \delta \ , \forall j \tag{5}$$

Concerning the additive valence utility function, and following a similar argument, the additive parameter δ_j is candidate-specific, while the slope coefficient λ is not. Thus, the additive valence implies the following testable restrictions on the unconstrained model (4):

$$\mathrm{H}_0: \ \lambda_j = \lambda \ , \forall j \tag{6}$$

[9] Indeed, and without loss of generality, consider that there are two candidates, $j = 1, 2$. If the utility function of voter i if candidate j is elected is the Downsian utility function $U(a_i, x_j) = -|x_j - a_i|$, then voter i strictly prefers candidate 1 if $-|x_1 - a_i| > -|x_2 - a_i| \Leftrightarrow |x_1 - a_i| < |x_2 - a_i|$. Now if the utility function of voter i if candidate j is elected is $U(a_i, x_j) = \delta - \lambda |x_j - a_i|$, with $\lambda > 0$, then voter i strictly prefers candidate 1 if $\delta - \lambda |x_1 - a_i| > \delta - \lambda |x_2 - a_i| \Leftrightarrow |x_1 - a_i| < |x_2 - a_i|$.

Lastly, the intensity valence utility function $U(a_i, x_j, \lambda_j, K) = \lambda_j(K - |x_j - a_i|)$ is equivalent to $U(a_i, x_j, \lambda_j, K) = c + \lambda_j(K - |x_j - a_i|)$ for some scalar c independent of i and j; again, c is a scaling parameter which depends on the scales used in the questions of the survey. The intensity valence implies the following testable restriction on the unconstrained model (4):

$$\mathrm{H}_0 : \ \delta_j = \lambda_j K + c \ , \forall j \tag{7}$$

It is prudent to emphasize that there is a minimal number of candidates (i.e., a minimal number of equations) to have testable implications on the unconstrained model (4). It is obvious that the Downsian and the additive models imply testable restrictions on the unconstrained model if there are at least two candidates ($M \geq 2$). However, the intensity valence model has no testable implication if $M = 2$. The reason is that if $M = 2$, the parameters of the unconstrained model would exactly identify the parameters of the intensity valence model. To see that, consider that the parameters of the unconstrained model $\{(\delta_1, \lambda_1), (\delta_2, \lambda_2)\}$ are known. Then, we can recover the underlying parameters of the intensity valence model if $\lambda_1 \neq \lambda_2$.[10] The parameters λ_1 and λ_2 of the unconstrained model give the intensity valence indices. And, according to the restriction (7), $\delta_1 = \lambda_1 K + c$ and $\delta_2 = \lambda_2 K + c$, so $K = \frac{\delta_1 - \delta_2}{\lambda_1 - \lambda_2}$ and $c = \delta_1 - \frac{\lambda_1(\delta_1 - \delta_2)}{\lambda_1 - \lambda_2} = \delta_2 - \frac{\lambda_2(\delta_1 - \delta_2)}{\lambda_1 - \lambda_2}$.[11] Thus, the underlying parameters of the intensity valence model are exactly identified in terms of the parameters of the unconstrained model. Hence, the intensity valence model implies no testable restriction on the unconstrained model when $M = 2$.

Now, if $M = 3$, the restriction (7) on the unconstrained model can be rewritten in terms of the

[10]If $\lambda_1 = \lambda_2$, the intensity valence utility function has no candidate-specific parameter, so it becomes equivalent to the Downsian utility function, as one can see in Equations (1) and (3); hence, it makes no sense to try to recover the other intensity valence parameters K and c when $\lambda_1 = \lambda_2$. Furthermore, if $\lambda_1 = \lambda_2$, it is in fact impossible to recover K and c in terms of the parameters of the unconstrained model as one can easily see below.

[11]The second equality $\delta_1 - \frac{\lambda_1(\delta_1 - \delta_2)}{\lambda_1 - \lambda_2} = \delta_2 - \frac{\lambda_2(\delta_1 - \delta_2)}{\lambda_1 - \lambda_2}$ is easy to verify: by multiplying each side by $(\lambda_1 - \lambda_2)$, we obtain $-\delta_1\lambda_2 + \delta_2\lambda_1 = \delta_2\lambda_1 - \delta_1\lambda_2$.

parameters of the unconstrained model as $\mathrm{H}_0: \ \delta_1 = \frac{\delta_2(\lambda_1-\lambda_3)+\delta_3(\lambda_2-\lambda_1)}{\lambda_2-\lambda_3}$. So the intensity valence implies one testable restriction on the unconstrained model. To see why, note that according to the restriction (7), $\delta_1 = \lambda_1 K + c$, $\delta_2 = \lambda_2 K + c$ and $\delta_3 = \lambda_3 K + c$. Hence, we should have $K = \frac{\delta_1-\delta_2}{\lambda_1-\lambda_2}$, as well as $K = \frac{\delta_1-\delta_3}{\lambda_1-\lambda_3}$ and $K = \frac{\delta_2-\delta_3}{\lambda_2-\lambda_3}$. We thus have a priori three testable restrictions: (i.) $\frac{\delta_1-\delta_2}{\lambda_1-\lambda_2} = \frac{\delta_1-\delta_3}{\lambda_1-\lambda_3}$; (ii.) $\frac{\delta_1-\delta_2}{\lambda_1-\lambda_2} = \frac{\delta_2-\delta_3}{\lambda_2-\lambda_3}$; (iii.) $\frac{\delta_2-\delta_3}{\lambda_2-\lambda_3} = \frac{\delta_1-\delta_3}{\lambda_1-\lambda_3}$. Note that (i.) $\frac{\delta_1-\delta_2}{\lambda_1-\lambda_2} = \frac{\delta_1-\delta_3}{\lambda_1-\lambda_3} \Leftrightarrow \delta_1 = \frac{\delta_2(\lambda_1-\lambda_3)+\delta_3(\lambda_2-\lambda_1)}{\lambda_2-\lambda_3}$; this is how H_0 has been stated above. It is easy to show that this equivalence is also true for (ii.) and (iii.), so the restrictions (ii.) and (iii.) are redundant.

3 The data

The data used in this paper are drawn from the 2004 pre-election ANES. This survey was conducted by the Survey Research Center of the University of Michigan's Institute for Social Research. It began on September 7, 2004 and ended November 1, 2004. No interviewing was conducted on Presidential Election Day, November 2, 2004. The sample was structured to be representative of the electorate. 1212 respondents were interviewed. Each respondent i was asked three key questions for our analysis. First, each respondent i was asked to rate his affect toward Kerry ($U_{i,k}$), Nader ($U_{i,n}$) and Bush ($U_{i,b}$).[12] These feeling thermometers are sometimes considered as

[12]The interviewer first said:

> I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one.

At the same time, the survey also made use of a respondent booklet and showed a 0-100 degree scale indicating in addition the meaning of 15, 30, 40, 60, 70 and 85 degrees. Then, the interviewer asked respondent i to rate his affect toward the three main Presidential candidates. For instance, for John Kerry, the question was:

> How would you rate JOHN KERRY?

the best available measures of voters' utility from political alternatives (e.g., Armstrong II *et al.*, 2014, p.147). They will be the regressands of the systems of equations described in Section 2.2. Second, each respondent i was asked his placement $\widetilde{a}_i$ on a 7-point scale wherein the political views were arranged from extremely liberal (1) to extremely conservative (7).[13] Finally, each respondent i was also asked to place Kerry ($\widetilde{x}_{i,r}$), Nader ($\widetilde{x}_{i,n}$) and Bush ($\widetilde{x}_{i,b}$) on this 7-point liberal-conservative scale.[14] The Original sample in Table 1 provides descriptive statistics of the responses to these different questions. Some observations are missing because some respondents refused to answer some questions, or they provided unsuitable answers (e.g., they said that they "Haven't thought much" or provided a "Don't know" to the self-placement question). In order to generate the estimation sample (called the Final sample in Table 1), observations where either one of the thermometer scores ($U_{i,k}$, $U_{i,n}$ or $U_{i,b}$), the self-placement ($\widetilde{a}_i$), or the perceived location of one of the candidate ($\widetilde{x}_{i,r}$, $\widetilde{x}_{i,n}$ or $\widetilde{x}_{i,b}$) are missing are dropped. It reduces the sample size to 607. Furthermore, 5 additional observations are excluded. Anticipating on Subsection 3.2, the reason is that these 5 respondents locate Kerry, Nader and Bush at the same place (i.e., $\widetilde{x}_{i,r} = \widetilde{x}_{i,n} = \widetilde{x}_{i,b}$). This absence of variability in the perceived location of the three candidates makes it impossible to estimate the Aldrich-McKelvey distortion parameters for these respondents (denoted c_i and w_i in Subsection 3.2); without these distortion parameters, it is impossible to obtain these respondents' bliss points in the same policy space as the actual locations of the candidates, and then compute

[13] The wording of the question was as follows:

> We hear a lot of talk these days about liberals and conservatives. Here is a seven-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place YOURSELF on this scale, or haven't you thought much about this? [1] Extremely liberal, [2] Liberal, [3] Slightly liberal, [4] Moderate/middle of the road, [5] Slightly conservative, [6] Conservative, [7] Extremely conservative, [80] Haven't thought much, [88] Don't know, [89] Refused.

[14] The questions concerning the locations of the candidates followed the self-placement question. As an example, the wording for Bush was as follows:

> Where would you place GEORGE W. BUSH on this scale?

accurate distances. Thus, the Final sample is composed of 602 respondents.

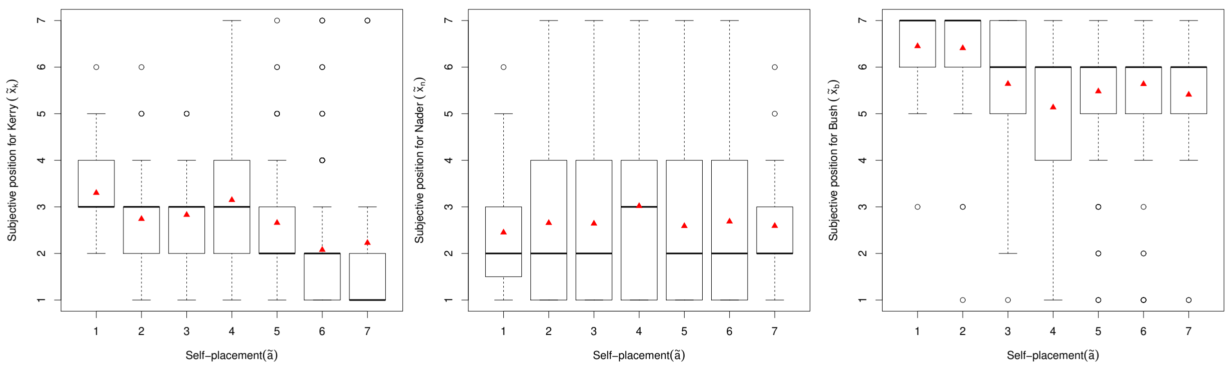
Table 1: Descriptive statistics

Variable	Original sample Obs.	Mean	Std.Dev.	Median	Min	Max	Final sample Obs.	Mean	Std.Dev.	Median	Min	Max
$U_{i,k}$	1191	53.019	26.360	60	0	100	602	51.612	26.729	60	0	100
$U_{i,n}$	980	42.814	22.605	50	0	100	602	42.647	22.928	50	0	100
$U_{i,b}$	1207	54.941	33.547	60	0	100	602	55.453	35.157	70	0	100
$\widetilde{a}_i$	920	4.269	1.475	4	1	7	602	4.279	1.522	4	1	7
$\widetilde{x}_{i,k}$	1088	2.987	1.488	3	1	7	602	2.692	1.340	2	1	7
$\widetilde{x}_{i,n}$	784	2.933	1.652	3	1	7	602	2.745	1.599	2	1	7
$\widetilde{x}_{i,b}$	1084	5.183	1.728	6	1	7	602	5.591	1.602	6	1	7
Candidates locations and respondents' bliss points according to the Aldrich-McKelvey method												
a_i							602	0.153	0.819	0.113	-3.704	4.233
x_k							1	-0.422				
x_n							1	-0.394				
x_b							1	0.816				

Note that it would have been possible to use the self-placement $\widetilde{a}_i$ and the perceived location $\widetilde{x}_{i,j}$ of candidate j, $j = k, n, b$, to obtain the distance $d_{i,j} = |\widetilde{x}_{i,j} - \widetilde{a}_i|$. However, the perceived location $\widetilde{x}_{i,j}$ is respondent-specific, while x_j in the utility functions (1), (2) and (3) is not. Thus, using $\widetilde{x}_{i,j}$ and $\widetilde{a}_i$ to compute the distance is problematic if respondents do not interpret the scale in the same way, i.e., if there is a problem of interpersonal incomparability of responses. Subsection 3.1 shows that this problem is prevalent. Thus, Subsection 3.2 uses the Aldrich-McKelvey solution to obtain accurate locations of the candidates and the bliss point of each respondent in a common policy space.

3.1 A problem of interpersonal incomparability of responses

Panel (A) in Figure 2 suggests that, on average, voters who consider themselves as conservative ($\widetilde{a} = 6$) or extremely conservative ($\widetilde{a} = 7$), place Kerry more on the left than those who consider themselves as liberal ($\widetilde{a} = 2$) or extremely liberal ($\widetilde{a} = 1$). Indeed, Mean($\widetilde{x}_k|\widetilde{a} = 6$) = 2.07 and Mean($\widetilde{x}_k|\widetilde{a} = 7$) = 2.22 while Mean($\widetilde{x}_k|\widetilde{a} = 1$) = 3.3 and Mean($\widetilde{x}_k|\widetilde{a} = 2$) = 2.74. Note that the conditional means are depicted by (red) solid triangles in the box-and-whisker diagrams of Figure 2; it is then also easy to see in Panel (C) that, on average, voters who consider themselves as liberal or extremely liberal place Bush more on the right than those who consider themselves as conservative or extremely conservative. The conditional $(0.25, 0.50, 0.75)$ quantiles of the box-and-whisker diagrams provide similar stories, so these trends seem to be robust to outliers.



(A) Kerry (B) Nader (C) Bush

Figure 2: Perceived location of Kerry ($\widetilde{x}_k$), Nader ($\widetilde{x}_n$) and Bush ($\widetilde{x}_b$) conditional on self-placement ($\widetilde{a}$) (Final sample)
Notes: The three figures represent box-and-whisker diagrams. The bottom and the top of a box are the first and third quartile. The ends of the whiskers are the lowest datum still within 1.5 times the interquartile range from the first quartile and the highest datum still within 1.5 times the interquartile range from the third quartile. If there are any data beyond that distance (i.e., outliers), they are represented as circles. The conditional median is represented by a line (inside the box). The graphics also provide the conditional mean, represented by a (red) solid triangle.

Table 2 provides more formal tests of this problem of interpersonal incomparability. We first linearly regress (via OLS) $\widetilde{x}_{i,j}$ on $\widetilde{a}_i$ for each candidate j, i.e., $\widetilde{x}_{i,j} = \beta_{0,j} + \beta_{1,j}\widetilde{a}_i + \varepsilon_{i,j}$. If there is interpersonal incomparability of responses for candidate j, then the null hypothesis $\mathrm{H}_0 : \ \beta_{1,j} = 0$ should be rejected. Part [A] of Table 2 shows that $\widehat{\beta}_{1,j}$ is negative and significantly different from zero for Kerry ($j = k$) and Bush ($j = b$); that is, conservative respondents place Kerry significantly more on the left than do liberal respondents and liberal respondents place Bush significantly more on the right than do conservative respondents. Given the ordinal nature of the data, a Spearman's rank correlation ρ between $\widetilde{x}_j$ and $\widetilde{a}$ has also been considered to test the null of interpersonal comparability $\mathrm{H}_0 : \ \rho(\widetilde{a}, \widetilde{x}_j) = 0$. Part [B] of Table 2 shows that the null is again rejected for Kerry and Bush.

Table 2: Statistical tests of interpersonal comparability of responses

	[1] Kerry ($\widetilde{x}_k$)	[2] Nader ($\widetilde{x}_n$)	[3] Bush ($\widetilde{x}_b$)
[A] OLS estimates			
$\widehat{\beta}_{1,j}$	-0.189***	-0.002	-0.141***
	(0.035)	(0.041)	(0.034)
	[0.000]	[0.948]	[0.000]
$\widehat{\beta}_{0,j}$	3.501***	2.754***	6.197***
	(0.145)	(0.194)	(0.156)
R^2	0.046	0.000	0.020
[B] Spearman's rank correlations			
$\widehat{\rho}(\widetilde{a}, \widetilde{x}_j)$	-0.316***	-0.000	-0.214***
	[0.000]	[0.986]	[0.000]
N	602	602	602

Notes: (i.) *, ** and *** represent statistical significance at the 10, 5 and 1% levels, respectively.
(ii.) Heteroskedasticity-robust standard errors are in parentheses.
(iii.) Achieved significance levels (or p-values) of interest are in brackets [·].

3.2 Recovering the underlying distances between the candidates and the respondents

Given that the problem of interpersonal incomparability of responses is obvious, using $\widetilde{a}_i$ and $\widetilde{x}_{i,j}$ to obtain the right-hand side variables $d_{i,j}$ can bias the results. Thus, we follow Aldrich and McKelvey (1977) to recover the underlying locations of the candidates and the respondents on a common dimension, the real line $\mathbb{R}$. Then, these values are used to compute the actual distances. We briefly explain the method below.

The general idea of Aldrich and McKelvey is that the responses $\widetilde{x}_{i,j}$, $j = k, n, b$, provided by respondent i follow a two-step process.[15] In a first step, respondent i retrieves relevant information

[15]The assumption that the response process is multi-steps is in line with the current literature on the psychology of survey responses. For instance, the state-of-the-art book of Tourangeau *et al.* (2000) considers that a

on the actual locations x_j, $j = k, n, b$, of the candidates on $\mathbb{R}$. However, because respondent i has only a finite amount of time to process information and provide answers, it is assumed that his perception $x_{i,j}$ of the candidate j is subject to an error term $\varepsilon_{i,j}$, such that $x_{i,j} = x_j + \varepsilon_{i,j}$. Note that $\varepsilon_{i,j}$ satisfies the traditional Gauss-Markov assumptions. In a second step, respondent i reports an answer $\widetilde{x}_{i,j}$ for each candidate j to the interviewer. This answer is assumed to be a (linear) distortion of his perception $x_{i,j}$ since there is not a common metric for placing the candidates. If so, there are distortion parameters c_i and w_i for each respondent i such that $x_{i,j} = x_j + \varepsilon_{i,j} = c_i + w_i\widetilde{x}_{i,j}$ (Aldrich and McKelvey, 1977, Equation (3), p.114).

Consider the following matrix notation:

$$X = \begin{bmatrix} x_k \\ x_n \\ x_b \end{bmatrix} \quad \widetilde{X}_i = \begin{bmatrix} 1 & \widetilde{x}_{i,k} \\ 1 & \widetilde{x}_{i,n} \\ 1 & \widetilde{x}_{i,b} \end{bmatrix} \quad \beta_i = \begin{bmatrix} c_i \\ w_i \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

If the vector X of actual locations were known, then the best linear unbiased estimator $\widehat{\beta}_i$ of the distortion parameters for respondent i would be $\widehat{\beta}_i = (\widetilde{X}_i'\widetilde{X}_i)^{-1}\widetilde{X}_i'X$, and the sum of squared residuals for this respondent $(X - \widetilde{X}_i\widehat{\beta}_i)'(X - \widetilde{X}_i\widehat{\beta}_i)$. To obtain the vector X of actual locations, it is assumed that the scale is standardized, i.e., $\sum_{j\in\{k,n,b\}} x_j = X'F = 0$ and $\sum_{j\in\{k,n,b\}} x_j^2 = X'X = 1$. Then the total sum of squared residuals of all the respondents is minimized subject to this standardized scale constraint. That is, a Lagrangian multiplier problem is set up as follows: $\mathcal{L}(\widehat{\beta}_i, X, \alpha_1, \alpha_2) = \sum_{i=1}^{N}(X - \widetilde{X}_i\widehat{\beta}_i)'(X - \widetilde{X}_i\widehat{\beta}_i) + 2\alpha_1 X'F + \alpha_2(X'X - 1)$, where α_1 and α_2 are Lagrangian multipliers. Setting $A = \sum_{i=1}^{N} \widetilde{X}_i(\widetilde{X}_i'\widetilde{X}_i)^{-1}\widetilde{X}_i'$, the Lagrangian multiplier problem permits to obtain $[A - NI_3]X = \alpha_2 X$, where I_3 is the 3×3 identity matrix (Aldrich and McKelvey, 1977, Equation (24), p.115). By definition, α_2 is an eigenvalue of $[A - NI_3]$ and X an eigenvector

survey response process involves four steps: understanding the question, retrieving relevant information, using this information to make a judgment, and selecting and reporting of an answer.

of $[A - NI_3]$ (Fuente, 2000, p.146). It can then be shown that $-X'[A - NI_3]X = \sum_{i=1}^{N}(X - \widetilde{X}_i\widehat{\beta}_i)'(X - \widetilde{X}_i\widehat{\beta}_i) = -\alpha_2$ (Aldrich and McKelvey, 1977, Equation (26), p.116). In words, the solution X is the eigenvector of $[A - NI_3]$ with the highest (negative) nonzero eigenvalue.

Having obtained the candidate locations X, it is then possible to obtain each respondent's bliss point in the common space, the real line. Indeed, it is now possible to estimate the distortion parameters by computing $\widehat{\beta}_i = (\widetilde{X}_i'\widetilde{X}_i)^{-1}\widetilde{X}_i'X$. Then, one has to subject each respondent's bliss point to the same transformation that his perceptions of the candidates are subjected to. Given that $\widetilde{a}_i$ is respondent i's self placement, his bliss point a_i in the common policy space is $a_i = \widehat{c}_i + \widehat{w}_i\widetilde{a}_i$ (Aldrich and McKelvey, 1977, Equation (32), p.117).

The computations are carried out in the R environment and make use of the basicspace package (Poole *et al.*, 2016). We obtain that the location of Kerry is $x_k = -0.422$, the one of Nader is $x_n = -0.394$, and the one of Bush is $x_b = 0.816$. Concerning the respondents' bliss points a, Figure 3 presents the kernel density estimate, based on a Gaussian kernel and a bandwidth chosen according to likelihood cross-validation. Note this kernel density estimate will be helpful for the counterfactual analyses based on the intensity valence model. Indeed, with this model, the existence of dominant strategies depends on the whole distribution of voter preferences (and not only the median voter). To test if one candidate has dominant strategies which would have insured him to obtain a majority of the popular vote, it is necessary to understand where are his plausible dominant strategies. The shape of the kernel density estimate will be helpful for that; we will go back to this issue in Subsection 5.2.

4 Estimation results

Given that we have now an accurate location x_j for each candidate j and the bliss point a_i of each respondent i in a common policy space, it is possible to compute accurate distances

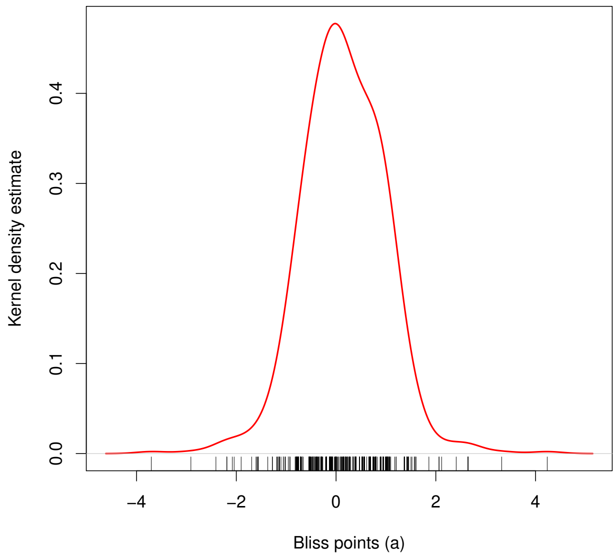


Figure 3: Kernel density estimate for bliss points (a) –Gaussian kernel with bandwidth of 0.3043–
Note: The selection of the bandwidth ($h = 0.3043$) is done by likelihood cross-validation (Li and Racine, 2007, p.18) making use of the R np package (Hayfield and Racine, 2008).

$d_{i,j} = |x_j - a_i|$, and estimate the different SUR models described in Subsection 2.2. This Section presents the maximum likelihood estimates[16], and tests the three constrained utility functions (i.e., the Downsian, the additive valence and the intensity valence). Subsection 4.1 provides the results but treats the distances between a_i and x_j as observed variables, i.e., ignoring any estimation error in these variables. It is prudent to stress that the locations of the candidates and the respondents' bliss points are estimates; the uncertainty in the estimates of these variables can influence the test statistics considered. Hence, Subsection 4.2 proposes a bootstrap method to solve this potential problem. Lastly, Subsection 4.3 relaxes the assumption that the utility functions are linear in distance, and compares the results with those obtained without an Aldrich-McKelvey correction.

[16]In practice, the maximum likelihood estimates of the different SUR models considered are computed using an iterated Zellner scheme (see, e.g., Ruud, 2000, p.706).

4.1 Preliminary results

Column [1] of Table 3 presents the maximum likelihood estimate of the unconstrained SUR model. Columns [2]-[4] give the maximum likelihood estimates of the three constrained models. Given that the estimation method rests on maximizing the likelihood function, a constrained model is rejected if its restrictions make a large difference to the maximized value of the log-likelihood function. As explained below, the likelihood ratio statistics indicate that the Downsian and the additive valence models are rejected, while the intensity valence model is not.[17]

Column [2] imposes the Downsian hypothesis, so there are only two parameters: one additive coefficient δ which is the same for the three candidates, and one slope coefficient λ which is also the same for the three candidates. Hence, the Downsian model imposes four restrictions on the unconstrained model. The likelihood ratio statistic is 146.964 ($\simeq 2 \times [-8312.973 - (-8386.455)]$). Given that the 1 percent critical value from the chi-squared distribution with 4 degrees of freedom is 13.28, the Downsian model is rejected.

Column [3] imposes the additive valence hypothesis, so there are four parameters: one slope coefficient λ which is the same for the three candidates, and the three additive coefficients δ_j, $j = k, n, b$. Thus, the additive valence model imposes two restrictions on the unconstrained model. The likelihood ratio statistic is 29.183 ($\simeq 2 \times [-8312.973 - (-8327.565)]$). The 1 percent critical value with 2 degrees of freedom is 9.21. So the additive valence model is also rejected.

Lastly, Column [4] imposes the intensity valence hypothesis. As stated in Subsection 2.2, and given that there are three candidates, the intensity valence model imposes one restriction on the unconstrained model. The likelihood ratio statistic, 0.133 ($\simeq 2 \times [-8312.973 - (-8313.04)]$), is

[17] A likelihood ratio statistic is twice the difference between the unconstrained maximum value of the log-likelihood function and the maximum subject to the restrictions: $2\left(\widehat{\ell}_u - \widehat{\ell}_c\right)$, where $\widehat{\ell}_u$ and $\widehat{\ell}_c$ denote, respectively, the unconstrained and constrained maximum log-likelihood values. It is asymptotically chi-square distributed with degrees of freedom equal to the number of restrictions imposed.

far less than 2.71, the 10 percent critical value with 1 degree of freedom. Hence, the intensity valence is the sole model which is not rejected.

Table 3: Maximum likelihood estimates of the 4 SUR models

	[1] Unconstrained	[2] Downs $\delta_j = \delta$, $\forall j$ and $\lambda_j = \lambda$, $\forall j$	[3] Additive valence $\lambda_j = \lambda$, $\forall j$	[4] Intensity valence $\delta_j = \lambda_j K + c$
U_k (Utility if Kerry is elected)				
$\widehat{\delta}_k$	60.769***	59.650***	60.086***	
	(1.445)	(0.910)	(1.209)	
$\widehat{\lambda}_k$	12.954***	11.007***	12.083***	13.284***
	(1.376)	(0.947)	(0.927)	(1.152)
U_n (Utility if Nader is elected)				
$\widehat{\delta}_n$	47.070***	59.650***	51.942***	
	(1.485)	(0.910)	(1.197)	
$\widehat{\lambda}_n$	5.748***	11.007***	12.083***	5.591***
	(1.500)	(0.947)	(0.927)	(1.470)
U_b (Utility if Bush is elected)				
$\widehat{\delta}_b$	71.988***	59.650***	65.589***	
	(1.866)	(0.910)	(1.504)	
$\widehat{\lambda}_b$	19.711***	11.007***	12.083***	19.530***
	(1.674)	(0.947)	(0.927)	(1.577)
$\widehat{K}$				1.797***
				(0.184)
$\widehat{c}$				36.966***
				(2.618)
N	602	602	602	602
Log-likelihood	-8312.973	-8386.455	-8327.565	-8313.04
Likelihood ratio test		146.964***	29.183***	0.133
Bootstrapped likelihood ratio test ($\widehat{ASL}$) (assuming C is the 1 percent critical value)		0***	0.011**	0.979
Bootstrapped likelihood ratio test ($\widehat{ASL}$) (assuming C is the 10 percent critical value)		0***	0.002***	0.867

Notes: i.*, ** and *** represent statistical significance at the 10, 5 and 1% levels, respectively.
ii. Standard errors are in parentheses.
iii. "Unconstrained" provides the maximum likelihood estimate of the unconstrained SUR model.
iv. Bootstrapped likelihood ratio test ($\widehat{ASL}$) provides the achieved significance level of the test, i.e., $\widehat{ASL} = \frac{\sharp\{\eta=1,\ldots,B\ ;\ LR^*(\eta)\leq C\}}{B}$ (we always consider $B = 999$ bootstrap samples). The first line assumes that C is the 1 percent critical value from the chi-squared distribution, i.e., $C = 13.28$ for the test of the Downsian model (given that it has 4 degrees of freedom), $C = 9.21$ for the test of the additive valence model (given that it has 2 degrees of freedom), and $C = 6.63$ for the test of the intensity valence model (given that it has 1 degree of freedom). The second bootstrapped likelihood ratio test assumes that C is the 10 percent critical value from the chi-squared distribution, i.e., $C = 7.78$ for the test of the Downsian model (given that it has 4 degrees of freedom), $C = 4.61$ for the test of the additive valence model (given that it has 2 degrees of freedom), and $C = 2.71$ for the test of the intensity valence model (given that it has 1 degree of freedom). The method and the results are discussed in Subsection 4.2.

4.2 Bootstrapped likelihood ratio tests

A critic which can be addressed against the likelihood ratio tests discussed above is that the log-likelihoods used to compute them depend in part on the first-step estimation of the locations of the candidates and the bliss points of the respondents obtained via the Aldrich-McKelvey method. This is a problem because we have inferred that a model is rejected or not ignoring any estimation error in the first-step estimation. To take into account this two-step estimation problem, this Subsection considers bootstrapped likelihood ratio tests; we explain below.

Let $\omega_i = (U_{ik}, U_{in}, U_{ib}, \widetilde{x}_{ik}, \widetilde{x}_{in}, \widetilde{x}_{ib}, \widetilde{a}_i)$ be the respondent i's answers to the set of questions which are needed for the estimations, and $\mathbb{W} = (\omega_1, \omega_2, \ldots, \omega_N)'$ the sample. A bootstrap sample $\mathbb{W}^*$ of size N is obtained by sampling from $\omega_1, \omega_2, \ldots, \omega_N$ with replacement. Recall that in our case the sample is of size $N = 602$. The steps to obtain a bootstrapped likelihood ratio test in order to test a model (e.g., the Downsian utility function) can then be summarized as:

(i.) Draw $B = 999$ bootstrap samples $\mathbb{W}^*$ of size $N = 602$.

(ii.) For each bootstrap sample $\mathbb{W}^*$: first, estimate the actual locations of the candidates, i.e., $\{x_k^*, x_n^*, x_b^*\}$ as well as the bliss point a_i^* of each respondent i using the Aldrich-McKelvey method; second, use the values obtained to compute $d_{i,j}^*$, and estimate the unconstrained SUR model as well as the constrained model under consideration (e.g., the Downsian utility function); third, use the log-likelihoods of these two models to obtain the likelihood ratio statistic (LR^*). Given that there are B bootstrap samples, this leads to B likelihood ratio statistics $LR^*(\eta)$, $\eta = 1, 2, \ldots, B$.

(iii.) Count the proportion of bootstrap samples for which $LR^* \leq C$, where C is a critical value. In the context of a bootstrap likelihood ratio test, the critical value C can be chosen from the chi-squared table, given that a likelihood ratio statistic is chi-square distributed. We will consider two versions of the test: the first one assuming that C is the 1 percent critical value, and the second one assuming that C is the 10 percent critical value (with degrees of freedom depending on the model considered; e.g., 4 in the case of the Downsian one). The achieved significance level (or p-value) of the test is then $\widehat{ASL} = \frac{\sharp\{\eta=1,\ldots,B\ ;\ LR^*(\eta)\leq C\}}{B}$.

(iv.) Fail to reject the null hypothesis whenever $\widehat{ASL}$ is larger than standard levels of significance.

The two last lines in Table 3 provide the achieved significance levels of the tests. In the case of the Downsian model, out of $B = 999$ bootstrap samples, we never observe $LR^* \leq C$, whether C is assumed to be 13.28 (i.e., the 1 percent critical value from the chi-squared distribution with 4 degrees of freedom) or 7.78 (i.e., the 10 percent critical value). Thus, it gives an achieved significance level of zero in both cases. So the Downsian model is always rejected. In the case of the additive valence model, we observe $LR^* \leq C$ as many as 11 times when C is assumed to be 9.21 (i.e., the 1 percent critical value from the chi-squared distribution with 2 degrees of freedom), giving an achieved significance level of 0.011. So this model is rejected at the 5 percent significance level. Finally, the intensity valence is again the sole model which is never rejected at any conventional level of significance.

4.3 Relaxing the assumption that the utility functions are linear in distance

The estimations in Table 3 assume that the utility functions are linear in distance. Some authors consider a squared distance (e.g., Adams *et al.*, 2005). Given that the objective is to find the utility function that better represents voters' evaluations of candidates, we now consider that the utility functions are power functions of the distance, e.g., $U_{i,j} = \delta_j - \lambda_j |x_j - a_i|^{\gamma} + \varepsilon_{i,j}$ for the unrestricted utility function, where the power γ has to be estimated.

Table 4 provides the results. The first result is that, again, the intensity valence is the sole model which is not rejected by the data, considering the simple likelihood ratio tests or the bootstrapped likelihood ratio tests (the procedure is similar to the one presented in Section 4.2).

The second important result concerns the estimated coefficient $\widehat{\gamma} = 0.389$ and its estimated standard error $\widehat{se}(\widehat{\gamma}) = 0.076$. The 99 percent confidence interval of γ is then $[0.192, 0.587]$. Thus, the null hypothesis $\mathrm{H}_0 : \gamma = 1$, the assumption made in Subsections 4.1 and 4.2, is rejected. The null hypothesis $\mathrm{H}_0 : \gamma = 2$, the assumption made in Adams *et al.* (2005), is also rejected. Note that we have also considered a 99 percent confidence interval based on bootstrap percentiles to take into account that the locations of the candidates and the respondents' bliss points are estimates and thus prone to estimation errors. The procedure is very close to the one presented in Subsection 4.2, except that for each bootstrap sample, once the intensity valence model is estimated, $\widehat{\gamma}^*$ is saved. The percentile method uses the 0.5th and the 99.5th percentiles of the empirical distribution of the $B = 999$ bootstrap estimates $\widehat{\gamma}^*(\eta)$, $\eta = 1, \ldots, B$. Denote by $\widehat{\gamma}^*_{0.005}$ and $\widehat{\gamma}^*_{0.995}$ the 0.05th and the 99.5 percentiles of this empirical distribution. The percentile 99 percent confidence interval for γ is then $[\widehat{\gamma}^*_{0.005}, \widehat{\gamma}^*_{0.995}] = [0.211, 0.596]$, so again $\mathrm{H}_0 : \gamma = 1$ and $\mathrm{H}_0 : \gamma = 2$ are rejected.

Note that if we had not used an Aldrich-McKelvey correction to compute the distances, we would have also obtained that the intensity valence is the sole model which is not rejected by the data. But we would have obtained a $\widehat{\gamma}$ not significantly different from one; the results are presented in Appendix A. It echoes Adams *et al.* (2005, p.17) who aptly explain that "'there is evidence that linear utility gives a better fit to thermometer scores interpreted as utilities [...]. Inferring that the utility scale itself is linear from such evidence is, however, problematic, because [...] the policy scales from which distance is measured [...] are constrained to specified finite intervals (typically 1-7 or 1-10 for the policy scales[...])." In line with Adams *et al.*, our results show that when an Aldrich-McKelvey scaling is used to solve the problem of interpersonal incomparability of responses, the utility function is no more linear in distance.

Table 4: Maximum likelihood estimates of the 4 SUR models relaxing the assumption that the utility functions are linear in distance

	[1] Unconstrained	[2] Downs $\delta_j = \delta,\ \forall j$ and $\lambda_j = \lambda,\ \forall j$	[3] Additive valence $\lambda_j = \lambda,\ \forall j$	[4] Intensity valence $\delta_j = K\lambda_j + c$
U_k (Utility if Kerry is elected)				
$\widehat{\delta}_k$	81.072***	71.325***	71.648***	
	(5.915)	(4.327)	(4.116)	
$\widehat{\lambda}_k$	36.540***	24.760***	25.857***	36.239***
	(6.274)	(4.676)	(4.384)	(6.103)
U_n (Utility if Nader is elected)				
$\widehat{\delta}_n$	50.900***	71.325***	63.471***	
	(3.033)	(4.327)	(4.119)	
$\widehat{\lambda}_n$	9.984***	24.760***	25.857***	10.106***
	(3.393)	(4.676)	(4.384)	(3.371)
U_b (Utility if Bush is elected)				
$\widehat{\delta}_b$	102.995***	71.325***	77.297***	
	(8.718)	(4.327)	(4.211)	
$\widehat{\lambda}_b$	55.284***	24.760***	25.857***	55.427***
	(9.174)	(4.676)	(4.384)	(9.173)
$\widehat{K}$				1.147***
				(0.053)
$\widehat{c}$				39.357***
				(1.681)
$\widehat{\gamma}$	0.389***	0.470***	0.484***	0.389***
	(0.076)	(0.108)	(0.100)	(0.076)
				[0.211,0.596]
N	602	602	602	602
Log-likelihood	-8279.521	-8371.640	-8311.173	-8279.542
Likelihood ratio test		184.238***	63.304***	0.043
Bootstrapped likelihood ratio test ($\widehat{ASL}$) (assuming C is the 1 percent critical value)		0***	0***	0.986
Bootstrapped likelihood ratio test ($\widehat{ASL}$) (assuming C is the 10 percent critical value)		0***	0***	0.893

Notes: i.*, ** and *** represent statistical significance at the 10, 5 and 1% levels, respectively.
ii. Standard errors are in parentheses.
iii. "Unconstrained" provides the maximum likelihood estimate of the unconstrained SUR model.
iv. Bootstrapped likelihood ratio test ($\widehat{ASL}$) provides the achieved significance level of the test, i.e., $\widehat{ASL} = \frac{\sharp\{\eta=1,\ldots,B\ ;\ LR^*(\eta)\leq C\}}{B}$ (we always consider $B = 999$ bootstrap samples). The method is discussed in Subsection 4.2, as well as in Note (iv.) at the bottom of Table 3.
v. The interval [0.211,0.596] below the estimated standard error of $\widehat{\gamma}$ for the intensity valence model corresponds to the percentile 99 percent confidence interval $[\widehat{\gamma}^*_{0.005}, \widehat{\gamma}^*_{0.995}]$ for γ. The percentile method uses the percentiles of the empirical distribution of $B = 999$ bootstrap estimates $\widehat{\gamma}^*(\eta)$, $\eta = 1, \ldots, B$.

5 Who will win the popular vote?

In the previous section, we have shown that the intensity valence utility function is not rejected by the data. Given that we have the actual locations of candidates, the bliss points of voters, and the estimated parameters of the intensity valence utility function, one can show the sources of support for each candidate according to a spatial model of electoral competition. We elaborate on it in Subsection 5.1: we make the mild assumption that Kerry, Nader and Bush are the only candidates who compete for the election, and study if it would have been possible to predict the fact that Bush won the popular vote with 50.73 percent; we compare the results with those obtained via the Downsian and the additive valence models. Then, in Subsection 5.2, we provide a counterfactual analysis. More precisely, according to the theory, the parameters of the intensity valence utility functions are "deep" parameters, i.e., parameters which are invariant to the policies proposed by the candidates. If so, one can use them to determine the percentage of vote that one candidate would have obtained if he and the other candidates had proposed other policies than $x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$. A natural question in a spatial model of electoral competition is to know if one candidate may propose some policies which insure him to obtain a majority of the popular vote whatever the policies proposed by the other candidates. Thus, we study in Subsection 5.2 if one candidate has such dominant strategies.

5.1 Support for each candidate according to the intensity valence model

The locations of Kerry, Nader and Bush are $x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$, respectively. Using these values and the estimated intensity valence utility function parameters of the three candidates obtained in Column [4] of Table 4, Figure 4 describes the estimated intensity valence utilities given by the three candidates in the policy space. Given that there is three candidates, it is possible that some voters vote strategically. In particular, those whose most preferred candidate is Nader may prefer to vote for their second most preferred candidate given that Nader is unlikely to win. Thus, we will consider two extreme cases: a first case wherein all the voters vote sincerely, and a second case wherein all the voters whose most preferred candidate is Nader vote strategically, i.e., for their second most preferred candidate.

If all the voters vote sincerely, they will vote for the candidate who gives them the highest level of utility. It is easy to see from Figure 4, that:

(i.) The voters whose bliss points are strictly between $a_2 \simeq 0.068$ and $a_3 \simeq 2.456$ obtain the highest level of utility with Bush. If so, these voters will vote for Bush. Let S_j be the fraction of voters who will vote for candidate j. Then, the relative frequency of voters who will vote for Bush is $\widehat{S}_b = \frac{\sharp\{i=1,\ldots,N\ ;\ a_2<a_i<a_3\}}{N} \simeq 0.5099$.

(ii.) The voters whose bliss points are strictly between $a_1 \simeq -1.857$ and $a_2 \simeq 0.068$ obtain the highest level of utility with Kerry. If so, these voters will vote for Kerry. The relative frequency of voters who will vote for Kerry is then $\widehat{S}_k = \frac{\sharp\{i=1,\ldots,N\ ;\ a_1<a_i<a_2\}}{N} \simeq 0.4652$.

(iii.) The voters whose bliss points are strictly less than $a_1 \simeq -1.857$, as well as those whose bliss points are higher than $a_3 \simeq 2.456$, obtain the highest level of utility with Nader. The relative frequency of voters who will vote for Nader is then $\widehat{S}_n = \frac{\sharp\{i=1,\ldots,N\ ;\ a_i<a_1 \text{ or } a_i>a_3\}}{N} \simeq 0.0249$.

Now, **if all the voters whose most preferred candidate is Nader vote for their second most preferred candidate**, then it is easy to see from Figure 4, that:

(i.) The voters whose bliss points are higher than $a_3 \simeq 2.456$ and less than $a_4 \simeq 3.933$ prefer Nader but will vote for Bush, their second most preferred candidate. If so, all the voters whose bliss points are strictly between $a_2 \simeq 0.068$ and $a_4 \simeq 3.933$ will vote for Bush. The relative frequency of voters who will vote for Bush is then $\widehat{S}_b = \frac{\sharp\{i=1,\ldots,N\ ;\ a_2<a_i<a_4\}}{N} \simeq 0.518$.

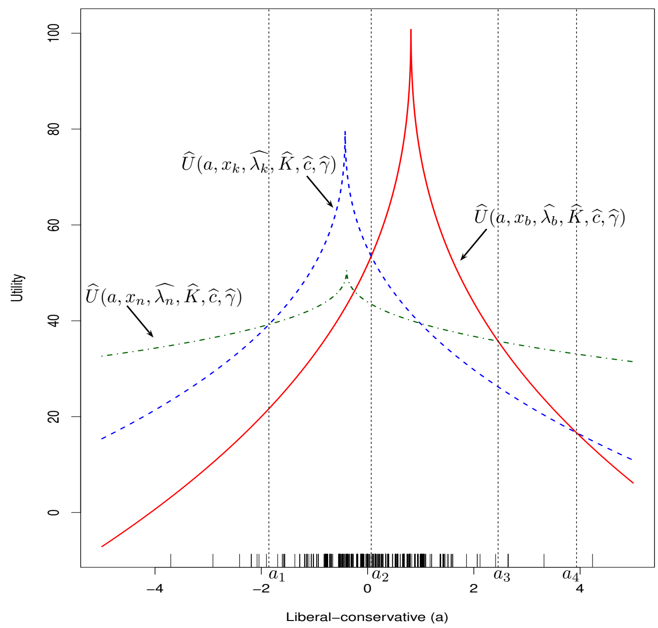


Figure 4: Estimated intensity valence utility functions

Note: This Figure depicts the three estimated intensity valence utilities in function of a obtained in Table 4. The (red) solid curve depicts the estimated utility if Bush is elected. The (blue) dashed curve depicts the estimated utility if Kerry is elected. The (green) dotdash curve depicts the estimated utility if Nader is elected. The locations of Kerry, Nader and Bush are $x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$, respectively. Finally, $a_1 \simeq -1.857$, $a_2 \simeq 0.068$, $a_3 \simeq 2.456$ and $a_4 \simeq 3.933$.

(ii.) The voters whose bliss points are higher than $a_4 \simeq 3.933$ prefer Nader but will vote for Kerry, their second most preferred candidate. The voters whose bliss points are less than $a_1 \simeq -1.857$ prefer Nader but will also vote for Kerry, their second most preferred candidate. If so, all the voters whose bliss points are strictly less than $a_2 \simeq 0.068$, as well as those whose bliss points are strictly more than $a_4 \simeq 3.933$ will vote for Kerry. The relative frequency of voters who will vote for Kerry is then $\widehat{S}_k = \frac{\sharp\{i=1,\ldots,N \ ; \ a_i<a_2 \text{ or } a_i>a_4\}}{N} \simeq 0.482$.

Table 5 provides a summary of the results. The point estimates of the fractions of vote obtained via the intensity valence model under the assumption of sincere voting (i.e., $\widehat{S}_b \simeq 0.5099$, $\widehat{S}_k \simeq 0.4652$, $\widehat{S}_n \simeq 0.0249$) or under the assumption of strategic voting (i.e., $\widehat{S}_b \simeq 0.518$, $\widehat{S}_k \simeq 0.482$, $\widehat{S}_n = 0$) are very close to the percentages of vote obtained by the candidates in reality (i.e., $S_b = 0.5073$, $S_k = 0.4827$, $S_n = 0.0038$). Table 5 also provides the point estimates of the fractions of vote obtained with the Downsian and the additive valence models (using the estimates of Table

4). Appendix B presents figures which depict the estimated utility functions of these models in the policy space and discusses in more details the results. The point estimates obtained with the Downsian model are unrealistic. Under the assumption of sincere voting, this model implies that 31.56 percent of the voters will vote for Nader (Line [2a] in Table 5); and under the assumption of strategic voting, this model implies that Kerry will win with 54.98 percent of the votes (Line [2b] in Table 5). Concerning the additive valence model, the point estimates are similar under the assumption of sincere and strategic voting (Lines [3a] and [3b] in Table 5); nobody will vote for Nader even under the assumption of sincere voting. This is due to the fact that the locations of Kerry ($x_k = -0.422$) and Nader ($x_n = -0.394$) are very close, and Kerry has an additive-valence advantage over Nader ($\widehat{\delta}_k > \widehat{\delta}_n$). Thus, the higher additive valence of Kerry implies a higher level of utility with Kerry than with Nader for all voters; see Figure B2 in Appendix B. The point estimates obtained for Kerry and Bush with the additive valence model are closer to reality than those obtained with the Downsian model, but not as close as those obtained with the intensity valence model.

Lastly, the point estimates of the fractions of votes obtain via the intensity valence model fit well with reality, but one would like to know if it would have been possible to predict with the pre-election survey the fact that Bush won the popular vote. That is, one would like to take into account sampling variation and test the null hypothesis $\mathrm{H}_0 : S_b \leq 0.50$ versus $\mathrm{H}_1 : S_b > 0.50$. If the null is rejected, one is able to predict the fact that Bush won the popular vote.

To test the null, we have drawn $B = 999$ bootstrap samples of size $N = 602$, as previously done. For each bootstrap sample $\mathbb{W}^*$, we first estimate the actual locations of the candidates, i.e., $\{x_k^*, x_n^*, x_b^*\}$, as well as the location a_i^* of each respondent i using Aldrich-McKelvey, then estimate the intensity valence SUR model, and use the estimated parameters to find the relative frequency $\widehat{S}_b^*$ of popular vote for Bush. There are $B = 999$ bootstrap estimates $\widehat{S}_b^*(\eta)$, $\eta = 1, 2, \ldots, B$; the

Table 5: Popular vote: Reality and point estimates according to the different models

	Kerry	Nader	Bush
[1] Popular vote: Reality			
	48.27%	0.38%	50.73%
[2a] Point estimates according to the Downsian model with sincere voting			
	23.75%	31.56%	44.68%
[2b] Point estimates according to the Downsian model with strategic voting			
	54.98%	0%	45.01%
[3a] Point estimates according to the additive valence model with sincere voting			
	47.51%	0%	52.49%
[3b] Point estimates according to the additive valence model with strategic voting			
	47.51%	0%	52.49%
[4a] Point estimates according to the intensity valence model with sincere voting			
	46.51%	2.49%	50.99%
[4b] Point estimates according to the intensity valence model with strategic voting			
	48.17%	0%	51.82%

Note: The results of the popular vote obtained in reality (Line [1]) are drawn from the 2004 United States presidential election Wikipedia webpage; see https://en.wikipedia.org/wiki/2004_United_States_presidential_election.

estimated achieved significance level of the test is then $\widehat{ASL} = \frac{\sharp\{\eta=1,\ldots,B\ ;\ \widehat{S}_b^*(\eta)\leq 0.50\}}{B}$.

Under the assumption that all the voters whose most preferred candidate is Nader vote for their second most preferred candidate, the achieved significance level is $\widehat{ASL} = 0.247$. Under the assumption that all the voters vote sincerely, the achieved significance level is $\widehat{ASL} = 0.384$. Hence, the null $\mathrm{H}_0 : S_b \leq 0.50$ is never rejected. It means that taking into consideration sampling variation, it would not have been possible to predict the fact that Bush obtained a majority of the popular vote with the pre-election survey.

5.2 Testing if Bush has dominant strategies which would have insured him to win the popular vote

A last question is to know if one candidate has dominant strategies which would have insured him to obtain a majority of the popular vote. As already noticed, with the intensity valence utility function, a high intensity valence is not always an advantage to win a popular vote; it depends on the heterogeneity of the distribution of voters in the policy space. More precisely, Gouret and Rossignol (2019, Propositions 3 and 5) show that, in a model with two purely-office motivated candidates, if the distribution of voters is sufficiently homogeneous, then the candidate with the highest intensity valence has dominant strategies which insure him to obtain a majority of the popular vote. In contrast, if the distribution of voters is too heterogenous, it is the candidate with the lowest intensity valence who has such dominant strategies.[18] This Section shows that the distribution of voters is sufficiently homogenous for Bush, the candidate with the highest intensity valence, to have dominant strategies. Like in Subsection 5.1, we consider two extreme cases: a case wherein all the voters vote sincerely, and a case wherein all the voters whose most preferred candidate is Nader vote for their second most preferred candidate. This second case is easier to analyze given that it reduces to a model of political competition with two candidates, Bush and Kerry; hence, we begin by studying this case.

5.2.1 Bush versus Kerry

In this first part, we consider that all the voters whose most preferred candidate is Nader vote for their second most preferred candidate, Kerry or Bush. A dominant strategy for Bush is thus defined as follows.

Definition 1 Let $S_b(x_b, x_k)$ be the fraction of voters who would have voted for Bush if Bush had located at x_b and Kerry at x_k. Consider that $x_b = \overline{x}_b$ and $S_b(\overline{x}_b, x_k)$ has a minimum at $x_k = \underline{x}_k$. Then, $\overline{x}_b$ is a dominant strategy which would have insured Bush to obtain a majority of the popular vote if $S_b(\overline{x}_b, \underline{x}_k) > \frac{1}{2}$.

Before to test if Bush has some dominant strategies, it is necessary to understand where are

[18] For moderate heterogeneity, no candidate has a dominant strategy which insures him to obtain the majority, and, more generally, no pure strategy Nash equilibrium exists; only mixed strategy equilibria exist in this intermediate case.

his plausible dominant strategies. Gouret and Rossignol (2019, Proposition 3) show that, in a model with two candidates, if the distribution of voters has a probability density function which is symmetric, strictly increasing on $(-\infty, m]$ and strictly decreasing on $(m, +\infty)$, with m the median/mean/mode of the distribution, then the candidate with the highest intensity valence has an interval of policies symmetric around m which insure him to obtain a majority of the popular vote. As shown in Table 1, the mean of the distribution of voters in the policy space is 0.153 and the median 0.113. The kernel density estimate in Figure 3, which also suggests that the distribution of voters is not symmetric, appears to be strictly increasing until a mode which is close to 0 and then strictly decreasing. These elements suggest that if Bush has dominant strategies, these dominant strategies are between –or around– 0 and 0.15. Thus, we have first studied if $\overline{x}_b = 0$ is a dominant strategy for Bush. Following Definition 1, and abstracting for the moment from sampling variation, we need to find $x_k = \underline{x}_k$ which minimizes $\widehat{S}_b(0, x_k)$, and then check that $\widehat{S}_b(0, \underline{x}_k) > \frac{1}{2}$. The steps we have followed can be summarized as:

(i.) We have used the estimated parameters in Column [4] of Table 4 to obtain the utility of each voter i if Bush had located at $\overline{x}_b = 0$, i.e., $\widehat{U}^0_{i,b} = \widehat{c} + \widehat{\lambda}_b \left(\widehat{K} - |0 - a_i|^{\widehat{\gamma}}\right)$.

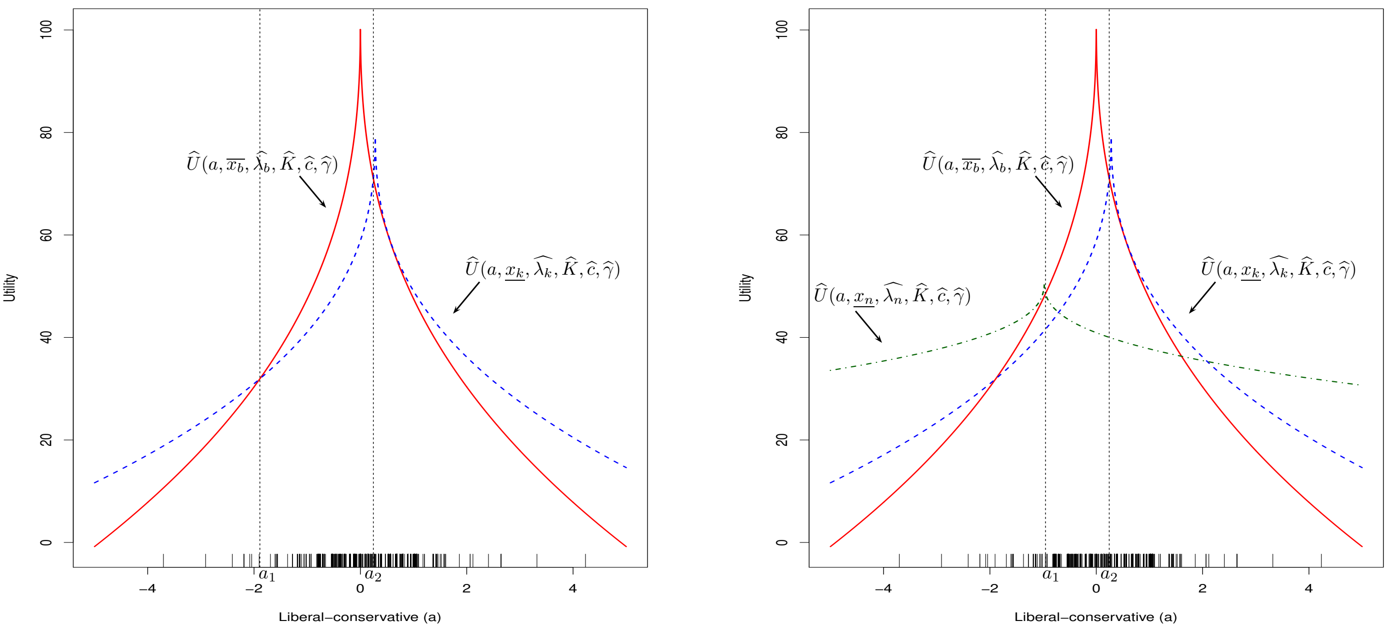
(ii.) Then, we have found $\underline{x}_k$, i.e., $\underline{x}_k \in \arg\min_{x_k \in \mathbb{R}} \widehat{S}_b(0, x_k)$. To do so, we have defined a vector of (801) possible values for x_k ranging from -3.7 to 4.3, incremented by 0.01.[19] For each possible value for x_k, we have used the estimated parameters in Column [4] of Table 4 to obtain the utility of each voter i if Kerry had located at x_k, i.e., $\widehat{U}_{i,k} = \widehat{c} + \widehat{\lambda}_k \left(\widehat{K} - |x_k - a_i|^{\widehat{\gamma}}\right)$, and have computed each time $\widehat{S}_b(0, x_k)$ by counting the proportion of voters for whom $\widehat{U}^0_{i,b} > \widehat{U}_{i,k}$, i.e., $\widehat{S}_b(0, x_k) = \frac{\sharp\{i=1,\ldots,N \ ; \ \widehat{U}^0_{i,b} > \widehat{U}_{i,k}\}}{N}$. Given that there are 801 possible values for x_k, this leads to 801 $\widehat{S}_b(0, x_k)$. Finally, among these 801 $\widehat{S}_b(0, x_k)$, we

[19] We have chosen values for x_k ranging from -3.7 to 4.3 because the respondents are located between -3.704 and 4.233; see Table 1.

have chosen the one which is minimal; $\underline{x}_k$ is the corresponding x_k.

$\widehat{S}_b(0, x_k)$ has a minimum at $x_k = \underline{x}_k = 0.28$. Panel (A) of Figure 5 depicts the utilities if Bush had located at $\overline{x}_b = 0$ and Kerry at $\underline{x}_k = 0.28$. If $\overline{x}_k = 0$ and $\underline{x}_k = 0.28$, then $\widehat{S}_b(0, 0.28) = 0.549$. Hence, abstracting from sampling variation, this result indicates that $\overline{x}_b = 0$ is a dominant strategy for Bush: it would have insured him to obtain a majority of the popular vote. To take into consideration sampling variation and test the null hypothesis that $\overline{x}_b = 0$ is not a dominant strategy (i.e., $\mathrm{H}_0 : S_b(0, \underline{x}_k) \leq 0.50$) against the alternative that it is (i.e., $\mathrm{H}_1 : S_b(0, \underline{x}_k) > 0.50$), we have drawn $B = 999$ bootstrap samples of size $N = 602$, as previously done. For each bootstrap sample $\mathbb{W}^*$, we estimate the actual locations of the candidates as well as the location of each respondent using Aldrich-McKelvey, then estimate the intensity valence SUR model, and use the estimated parameters and consider that $\overline{x}_b = 0$; finally we find the location of Kerry $\underline{x}_k^*$ which minimizes the percentage of votes for Bush, i.e., $\widehat{S}_b^*(0, \underline{x}_k^*)$. There are $B = 999$ bootstrap estimates $\widehat{S}_b^*[0, \underline{x}_k^*(\eta)]$, $\eta = 1, 2, \ldots, B$. The estimate of the achieved significance level of the test is then $\widehat{ASL} = \frac{\sharp\{\eta = 1, \ldots, B \ ; \ \widehat{S}_b^*[0, \underline{x}_k^*(\eta)] \leq 0.50\}}{B}$. Here, $\widehat{ASL} = 0.015$, so the null is rejected at the 5 percent significance level. Thus, $\overline{x}_b = 0$ is a dominant strategy for Bush.

Proceeding in a similar fashion, Part [A] of Table 6 provides the results with other plausible dominant strategies for Bush around zero. $\widehat{S}_b(\overline{x}_b, \underline{x}_k) > 0.50$ if $\overline{x}_b = \{-0.10, 0.10, 0.20, 0.30, 0.33\}$. However, the null $\mathrm{H}_0 : S_b(\overline{x}_b, \underline{x}_k) \leq 0.50$ is rejected at reasonable levels of significance only when $\overline{x}_b = \{0.10, 0.20\}$. The case $\overline{x}_b = 0.10$ is particularly striking: out of $B = 999$ bootstrap samples, we observe $\widehat{S}_b^*(0.10, \underline{x}_k^*) \leq 0.50$ only one time, giving an achieved significance level of 0.001. Thus, $\overline{x}_b = 0.10$ is clearly a dominant strategy which would have insured Bush to obtain a majority of the popular vote.



(A) Bush versus Kerry

(B) Bush versus Kerry and Nader

Figure 5: Counterfactual analyses

Notes: In Panel (A), it is assumed that all the voters whose most preferred candidate is Nader vote for their second most preferred candidate. It reduces to a model of political competition with Bush and Kerry. The (red) solid curve depicts the estimated utility with Bush, assuming that he locates at $\overline{x}_b = 0$. The (blue) dashed curve depicts the utility with Kerry, assuming that he locates at $\underline{x}_k \in \arg\min_{x_k \in \mathbb{R}} \widehat{S}_b(0, x_k)$, i.e., $\underline{x}_k = 0.28$. The relative frequency of voters who are between $a_1 \simeq -1.889$ and $a_2 \simeq 0.243$ is $\widehat{S}_b(0, 0.28) = \frac{\sharp\{i=1,\ldots,N \ ; \ a_1 < a_i < a_2\}}{N} \simeq 0.549$.
In Panel (B), it is assumed that all the voters vote sincerely. The (red) solid curve depicts the estimated utility with Bush, assuming that he locates at $\overline{x}_b = 0$. The (blue) dashed curve depicts the utility with Kerry and the (green) dotdash curve depicts the utility with Nader assuming that Kerry and Nader locate at $(\underline{x}_k, \underline{x}_n) \in \arg\min_{(x_k, x_n) \in \mathbb{R}^2} \widehat{S}_b(0, x_k, x_n)$, i.e., $(\underline{x}_k, \underline{x}_n) = (0.28, -0.98)$. The relative frequency of voters who are between $a_1 \simeq -0.954$ and $a_2 \simeq 0.243$ is $\widehat{S}_b(0, 0.28, -0.98) = \frac{\sharp\{i=1,\ldots,N \ ; \ a_1 < a_i < a_2\}}{N} \simeq 0.501$.

5.2.2 Bush versus Kerry and Nader

Finally, we have considered the case wherein all the voters vote sincerely. A dominant strategy for Bush is now defined as follows.

Definition 2 Let $S_b(x_b, x_k, x_n)$ be the fraction of voters who would have voted for Bush if he had located at x_b, Kerry at x_k and Nader at x_n. Consider that $x_b = \overline{x}_b$, and $S_b(\overline{x}_b, x_k, x_n)$ has a minimum at the pair $(\underline{x}_k, \underline{x}_n)$. If so, $\overline{x}_b$ is a dominant strategy which would have insured Bush to obtain a majority of the popular vote if $S_b(\overline{x}_b, \underline{x}_k, \underline{x}_n) > \frac{1}{2}$.

In order to compare the results with those of Subsubsection 5.2.1, Panel (B) of Figure 5 depicts the case if Bush had located at $\overline{x}_b = 0$, Kerry at $\underline{x}_k = 0.28$ and Nader at $\underline{x}_n = -0.98$;

Table 6: Counterfactual analyses: Testing if Bush has dominant strategies which would have insured him to win the popular vote

[A] Bush versus Kerry

	$\overline{x}_b$							
	-0.11	-0.10	0	0.10	0.20	0.30	0.33	0.34
$\underline{x}_k$	0.17	0.18	0.28	0.38	-0.09	0.01	0.04	0.06
$\widehat{S}_b(\overline{x}_b, \underline{x}_k)$	0.490	0.514	0.549	0.594	0.588	0.508	0.506	0.493
$\widehat{ASL}$	0.426	0.366	0.015**	0.001***	0.038**	0.306	0.432	0.489

[B] Bush versus Kerry and Nader

	$\overline{x}_b$							
	-0.11	-0.10	0	0.10	0.20	0.30	0.33	0.34
$\underline{x}_k$	0.17	0.18	0.28	0.38	0.48	0.01	0.04	0.06
$\underline{x}_n$	-1.07	-1.08	-0.98	-1.01	-0.77	1.22	1.27	1.28
$\widehat{S}_b(\overline{x}_b, \underline{x}_k, \underline{x}_n)$	0.450	0.475	0.501	0.539	0.519	0.458	0.456	0.443
$\widehat{ASL}$	0.798	0.765	0.542	0.408	0.431	0.795	0.875	0.901

Notes: i. *, ** and *** represent statistical significance at the 10, 5 and 1% levels, respectively.
ii. $\widehat{ASL}$ provides the achieved significance level of the test H_0 : "$\overline{x}_b$ is not a dominant strategy" versus H_1 : "$\overline{x}_b$ is a dominant strategy". $\widehat{ASL} = \frac{\sharp\{\eta=1,\ldots,B\ ;\ \widehat{S}^*_b[\overline{x}_b, \underline{x}^*_k(\eta)]\leq 0.50\}}{B}$ in Part [A] (Bush versus Kerry) and $\widehat{ASL} = \frac{\sharp\{\eta=1,\ldots,B\ ;\ \widehat{S}^*_b[\overline{x}_b, \underline{x}^*_k(\eta), \underline{x}^*_n(\eta)]\leq 0.50\}}{B}$ in Part [B] (Bush versus Kerry and Nader). We have always considered $B = 999$ bootstrap samples.

note that $(\underline{x}_k, \underline{x}_n)$ solves the following problem: $(\underline{x}_k, \underline{x}_n) \in \arg\min_{(x_k,x_n)\in\mathbb{R}^2} \widehat{S}_b(0, x_k, x_n)$. In this case, $\widehat{S}_b(0, 0.28, -0.98) = 0.501 > 0.50$; this result suggests that $\overline{x}_b = 0$ is a dominant strategy for Bush. However, when we take into account sampling variation via a bootstrap procedure similar to the one of Subsubsection 5.2.1, the null hypothesis $\mathrm{H}_0 : S_b(0, \underline{x}_k, \underline{x}_n) \leq 0.50$ is not rejected, so we cannot conclude that $\overline{x}_b = 0$ is a dominant strategy for Bush. Part [B] of Table 6 provides the results with other plausible dominant strategies for Bush. $\widehat{S}_b(\overline{x}_b, \underline{x}_k, \underline{x}_n) > 0.50$ when $\overline{x}_b = \{0.10, 0.20\}$, but the null $\mathrm{H}_0 : S_b(\overline{x}_b, \underline{x}_k, \underline{x}_n) \leq 0.50$ is never rejected at any reasonable level of significance. So we cannot conclude that there is a dominant strategy for Bush if all the voters vote sincerely.

6 Conclusion

This paper has taken the problem of interpersonal incomparability of responses to issue scales seriously, and tested different utility functions using the 2004 pre-election survey of the ANES. The one which is the most supported by the empirical evidence, the intensity valence utility function, is also the one which permits to make the better predictions for the 2004 United States presidential election. Furthermore, we have used the estimated intensity valence utility function and the distribution of voters in the policy space obtained via the Aldrich-McKelvey method to conduct some counterfactual analyses that assess if Bush had some dominant strategies which would have insured him to win the popular vote.

We have found that in 2004 Bush was the candidate with the highest intensity valence. The distribution of voter preferences in the policy space was sufficiently homogenous for him to have dominant strategies which would have insured him to obtain a majority of the popular vote (although they are not always significant). It is important to reiterate the fact that a model with intensity valence requires specific attention to heterogeneity in voter preferences. Too much heterogeneity may be problematic to implement intensively a policy. Above a threshold of heterogeneity, it is even known that it is the candidate with low intensity who has such dominant strategies, and he is less and less constrained by the median voter (Gouret and Rossignol, 2019, Proposition 3). It would be interesting to see if in more recent elections the distribution of voter preferences was too heterogenous for the candidate with high intensity valence to have dominant strategies. It may shed some light on the literature on mass polarization. Mass polarization is usually interpreted as a bimodal distribution of the electorate, although other interpretations exist (see, e.g., Krasa and Polborn, 2014). With this interpretation, the empirical research has struggled to find consensus on the existence of mass polarization. Fiorina and Abrams (2008) have argued that there is no evidence of mass polarization given that the American electorate remains unimodal. In contrast, Hare *et al.* (2015) have developed a Bayesian version of the Aldrich-McKelvey scaling, and, although they do not provide a multimodality test, their analysis of the 2012 ANES may suggest a bi- or even a trimodal distribution. The intensity valence model has nothing to say on the implication of a multimodal distribution. But if we interpret mass polarization as a too strong heterogeneity of the electorate, then the intensity valence model might permit to say if the distribution is too heterogenous. More precisely, consider that we find empirically that the candidate with low intensity has dominant strategies. It means that the distribution of the electorate is heterogenous, and that this candidate will be less and less constrained by the median voter if heterogeneity increases. Thus, it can explain the divergence from the median voter.

References

Adams, J., Merrill, S. and Grofman, B. (2005), *A Unified Theory of Party Competition*, Cambridge University Press.

Adams, J., Merrill, S., Simas, E. N. and Stone, W. J. (2011), "When candidates value good character: A spatial model with applications to congressional elections", *The Journal of Politics*, vol. 73: pp. 17–30.

Aldrich, J. H. and McKelvey, R. D. (1977), "A method of scaling with applications to the 1968 and 1972 Presidential elections", *American Political Science Review*, vol. 71: pp. 111–130.

Ansolabehere, S. and Snyder, J. (2000), "Valence politics and equilibrium in spatial election models", *Public Choice*, vol. 103: pp. 327–336.

Ansolabehere, S., Snyder, J. and Stewart III, C. (2001), "Candidate positioning in U.S. House elections", *American Journal of Political Science*, vol. 45: pp. 136–159.

Aragones, E. and Palfrey, T. E. (2002), "Mixed equilibrium in a Downsian model with a favored candidate", *Journal of Economic Theory*, vol. 103: pp. 131–161.

Aragonès, E. and Xefteris, D. (2012), "Candidate quality in a Downsian model with a continuous policy space", *Games and Economic Behavior*, vol. 75: pp. 464–480.

Armstrong II, D. A., Bakker, R., Carroll, R., Hare, C., Poole, K. T. and Rosenthal, H. (2014), *Analyzing Spatial Models of Choice and Judgment with R*, Chapman & Hall.

Ashworth, S. and Bueno de Mesquita, E. (2009), "Elections with platform and valence competition", *Games and Economic Behavior*, vol. 67: pp. 191–216.

Burden, B. C. (2004), "Candidate positioning in US congresional elections", *British Journal of Political Science*, vol. 34: pp. 211–227.

Câmara, O. (2012), "Economic policies of heterogeneous politicans", *mimeo University of Southern California*.

Carrillo, J. D. and Castanheira, M. (2008), "Information and strategic political polarization", *Economic Journal*, vol. 118: pp. 845–874.

Degan, A. (2007), "Candidate valence: Evidence from consecutive Presidential elections", *International Economic Review*, vol. 48: pp. 457–482.

Dix, M. and Santore, R. (2002), "Candidate ability and platform choice", *Economics Letters*, vol. 76: pp. 189–194.

Downs, A. (1957), *An Economic Theory of Democracy*, New York: Harper and Row.

EVRENK, H. (2009), "Three-candidate competition when candidates have valence: The base case", *Social Choice and Welfare*, vol. 32: pp. 157–168.

EVRENK, H. (2019), "Valence politics", in CONGLETON, R., GROFMAN, B. and VOIGT, S. (editors), *The Oxford Handbook of Public Choice*, Oxford: Oxford University Press.

FIORINA, M. P. and ABRAMS, S. J. (2008), "Political polarization in the American public", *Annual Review of Political Science*, vol. 11: pp. 563–588.

FUENTE, A. D. L. (2000), *Mathematical Methods and Models for Economists*, Cambridge: Cambridge University Press.

GOURET, F. and ROSSIGNOL, S. (2019), "Intensity valence", *Social Choice and Welfare*, vol. 53: pp. 63–112.

GOURET, F., HOLLARD, G. and ROSSIGNOL, S. (2011), "An empirical analysis of valence in electoral competition", *Social Choice and Welfare*, vol. 37: pp. 309–340.

GROSECLOSE, T. (2001), "A model of candidate location when one candidate has a valence advantage", *American Journal of Political Science*, vol. 45: pp. 862–886.

HARE, C., ARMSTRONG II, D. A., BAKKER, R., CARROLL, R. and POOLE, K. T. (2015), "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions", *American Journal of Political Science*, vol. 59: pp. 759–774.

HAYFIELD, T. and RACINE, J. S. (2008), "Nonparametric econometrics: The np package", *Journal of Statistical Software*, vol. 27 n° 5: pp. 1–32, http://www.jstatsoft.org/v27/i05/.

HOLLIBAUGH, G. E., ROTHENBERG, L. S. and RULISON, K. K. (2013), "Does it really hurt to be out of step?", *Political Research Quarterly*, vol. 66: pp. 856–867.

Hummel, P. (2010), "On the nature of equilibria in a Downsian model with candidate valence", *Games and Economic Behavior*, vol. 70: pp. 425–445.

Kartik, N. and McAfee, R. P. (2007), "Signaling character in electoral competition", *American Economic Review*, vol. 28: pp. 852–869.

Krasa, S. and Polborn, M. (2010), "Competition between specialized candidates", *American Political Science Review*, vol. 104: pp. 745–765.

Krasa, S. and Polborn, M. (2014), "Policy divergence and voter polarization in a structural model of elections", *Journal of Law and Economics*, vol. 57: pp. 31–76.

Krasa, S. and Polborn, M. K. (2012), "Political competition between differentiated candidates", *Games and Economic Behavior*, vol. 76: pp. 249–271.

Li, Q. and Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Miller, M. K. (2011), "Seizing the mantle of change: Modeling candidate quality as effectiveness instead of valence", *Journal of Theoretical Politics*, vol. 23: pp. 52–68.

Palfrey, T. R. and Poole, K. T. (1987), "The relationship between information, ideology, and voting behavior", *American Journal of Political Science*, vol. 31: pp. 511–530.

Poole, K., Lewis, J., Rosenthal, H., Lo, J. and Carroll, R. (2016), "Recovering a basic space from issue scales in R", *Journal of Statistical Software*, vol. 69 n° 7: pp. 1–21, http://www.jstatsoft.org/v69/i07/.

Ruud, P. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press.

SCHOFIELD, N., GALLEGO, M. and JEON, J. (2011), "Leaders, voters and activists in the elections in Great Britain 2005 and 2010", *Electoral Studies*, vol. 30 n° 3: pp. 484 – 496.

SOUBEYRAN, R. (2009), "Does a disadvantaged candidate choose an extremist position", *Annals of Economics and Statistics*, vol. 93-94: pp. 327–348.

STONE, W. J. and SIMAS, E. N. (2010), "Candidate valence and idelological positions in U.S. House elections", *American Journal of Political Science*, vol. 54: pp. 371–388.

TOURANGEAU, R., RIPS, L. J. and RASINSKI, K. (2000), *The Psychology of Survey Response*, Cambridge University Press.

XEFTERIS, D. (2012), "Mixed strategy equilibrium in a Downsian model with a favored candidate: A comment", *Journal of Economic Theory*, vol. 147: pp. 393–396.

XEFTERIS, D. (2014), "Mixed equilibriums in a three-candidate spatial model with candidate valence", *Public Choice*, vol. 158: pp. 101–120.

ZAKHAROVA, M. and WARWICK, P. V. (2014), "The sources of valence judgments: The role of policy distance and the structure of the left-right spectrum", *Comparative Political Studies*, vol. 27: pp. 493–517.

A Estimation results taking issue scale responses at face value

This Appendix provides maximum likelihood estimates of the four SUR models similar to those in Table 4. However, the responses to the liberal-conservative scale are taken at face value to obtain the distances, i.e., $d_{i,j} = |\widetilde{a}_i - \widetilde{x}_{i,j}|$.

Table A1 provides the results. The first result is that the Downsian and the additive valence models are rejected according to the likelihood ratio tests, while the intensity valence model is not. This first result confirms the results obtained in Section 4.

The second result concerns the estimated coefficient $\widehat{\gamma} = 1.038$ and its estimated standard error $\widehat{se}(\widehat{\gamma}) = 0.066$. The 95 percent confidence interval of γ is $[0.909, 1.168]$, so the null hypothesis $\mathrm{H}_0 : \gamma = 1$ is not rejected. However, the null $\mathrm{H}_0 : \gamma = 2$, the assumption made in Adams *et al.* (2005), is rejected. As pointed out in the main text, this result echoes Adams *et al.* (2005, p.17) who note that "'there is evidence that linear utility gives a better fit to thermometer scores interpreted as utilities". As the results with the Aldrich-McKelvey correction suggest (Table 4), and as Adams *et al.* aptly explain, the fact that the utility is linear in distance in Table A1 is due to the fact that "the policy scales from which distance is measured [...] are constrained to specified finite intervals."

B Support for each candidate according to the Downsian and additive valence models

B.1 Support for each candidate according to the Downsian model

Using the locations of Kerry, Nader and Bush ($x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$) and the Downsian utility function parameters obtained in Column [2] of Table 4, Figure B1 describes the estimated Downsian utilities given by the three candidates. As shown in Table 1, note that

Table A1: Maximum likelihood estimates of the 4 SUR models relaxing the assumption that the utility functions are linear in distance and taking liberal-conservative scale responses at face value

	[1] Unconstrained	[2] Downs $\delta_j = \delta$, $\forall j$ and $\lambda_j = \lambda$, $\forall j$	[3] Additive valence $\lambda_j = \lambda$, $\forall j$	[4] Intensity valence $\delta_j = K\lambda_j + c$
U_k (Utility if Kerry is elected)				
$\widehat{\delta}_k$	71.458***	70.081***	68.507***	
	(1.530)	(1.209)	(1.326)	
$\widehat{\lambda}_k$	9.228***	8.477***	7.054***	9.358***
	(1.350)	(1.135)	(1.044)	(1.148)
U_n (Utility if Nader is elected)				
$\widehat{\delta}_n$	53.348***	70.081***	61.006***	
	(1.617)	(1.209)	(1.484)	
$\widehat{\lambda}_n$	4.591***	8.477***	7.054***	4.503***
	(0.745)	(1.135)	(1.044)	(0.735)
U_b (Utility if Bush is elected)				
$\widehat{\delta}_b$	82.520***	70.081***	72.414***	
	(1.836)	(1.209)	(1.487)	
$\widehat{\lambda}_b$	12.620***	8.477***	7.054***	12.466***
	(1.501)	(1.135)	(1.044)	(1.485)
$\widehat{K}$				3.693***
				(0.396)
$\widehat{c}$				36.745***
				(1.915)
$\widehat{\gamma}$	1.038***	1.051***	1.139***	1.038***
	(0.066)	(0.075)	(0.083)	(0.066)
N	602	602	602	602
Log-likelihood	-8101.348	-8190.552	-8141.887	-8101.747
Likelihood ratio test		178.408***	81.077***	0.798

Notes: i.*, ** and *** represent statistical significance at the 10, 5 and 1% levels, respectively.
ii. Standard errors are in parentheses.
iii. "Unconstrained" provides the maximum likelihood estimate of the unconstrained SUR model.

voters are located between -3.704 and 4.233 on the liberal-conservative space. However, Figure B1 only considers the interval of the liberal-conservative space ranging from -1 to 1. We do so in order to clearly visualize when the estimated utilities intersect. Indeed, Kerry and Nader are so close ($x_k = -0.422$ and $x_n = -0.394$) that it would have been difficult to visualize these intersections if we had shown all the possible values taken by the respondents in the policy space. The main objective of this Appendix is to fully understand why when the voters vote sincerely, the Downsian model predicts unrealistically that Nader will obtain 31.56 percent of the votes. Thus, we focus on the situation wherein all the voters vote sincerely.

It is easy to see from Figure B1, that:

(i.) The voters whose bliss points are strictly higher than $a_2 \simeq 0.211$ obtain the highest level of utility with Bush. If so, these voters will vote for Bush. Then, according to the Downsian model, the relative frequency of voters who will vote for Bush is $\widehat{S}_b = \frac{\sharp\{i=1,\ldots,N\ ;\ a_i > a_2\}}{N} \simeq 0.4468$.

(ii.) The voters whose bliss points are strictly between $a_1 \simeq -0.408$ and $a_2 \simeq 0.211$ obtain the highest level of utility with Nader. If so, these voters will vote for Nader. The relative frequency of voters who will vote for Nader is then $\widehat{S}_n = \frac{\sharp\{i=1,\ldots,N\ ;\ a_1 < a_i < a_2\}}{N} \simeq 0.3156$.

(iii.) The voters whose bliss points are strictly less than $a_1 \simeq -0.408$ obtain the highest level of utility with Kerry. The relative frequency of voters who will vote for Kerry is then $\widehat{S}_k = \frac{\sharp\{i=1,\ldots,N\ ;\ a_i < a_1\}}{N} \simeq 0.2375$.

Thus, the Downsian model predicts unrealistically that Nader will obtain 31.56 percent of the vote (under the assumption of sincere voting) because he is located on the right of Kerry ($x_n = -0.394$ and $x_k = -0.422$) and attracts all the voters located between $a_1 \simeq -0.408$ and $a_2 \simeq 0.211$. This is an interval wherein a massive heap of voters are located as shown in Figure 3.

B.2 Support for each candidate according to the additive valence model

Using the locations of Kerry, Nader and Bush ($x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$) and the estimated additive valence utility function parameters obtained in Column [3] of Table 4, Figure B2 describes the estimated additive valence utilities given by the three candidates. The main objective is to understand why, even under the assumption of sincere voting, nobody will

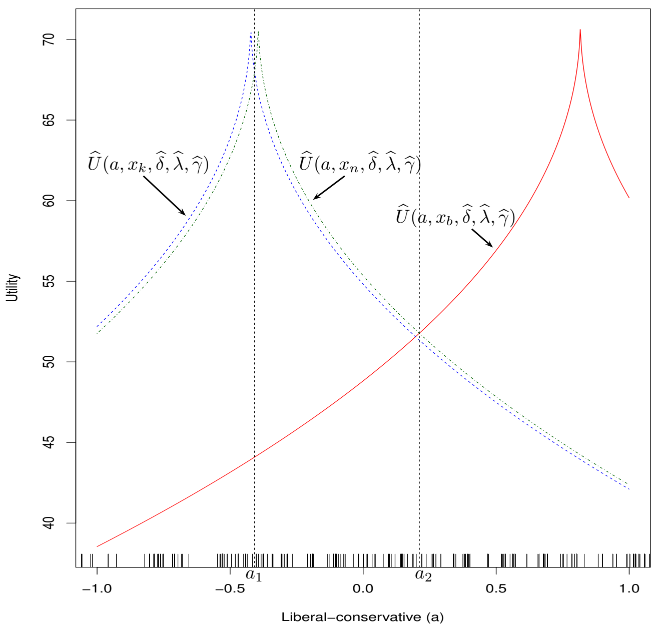


Figure B1: Estimated Downsian utility functions

Note: This Figure depicts the three estimated Downsian utilities in function of a obtained in Table 4. The (red) solid curve depicts the estimated utility if Bush is elected. The (blue) dashed curve depicts the estimated utility if Kerry is elected. The (green) dotdash curve depicts the estimated utility if Nader is elected. The locations of Kerry, Nader and Bush are $x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$, respectively. Finally, $a_1 \simeq -0.408$ and $a_2 \simeq 0.211$.

vote for Nader.

It is easy to see from Figure B2 that:

(i.) The voters whose bliss points are strictly higher than $a_1 \simeq 0.022$ obtain the highest level of utility with Bush. If so, these voters will vote for Bush. Then, according to the Downsian model, the relative frequency of voters who will vote for Bush is $\widehat{S}_b = \frac{\sharp\{i=1,\ldots,N\ ;\ a_i > a_1\}}{N} \simeq 0.5249$.

(ii.) The voters whose bliss points are strictly less than $a_1 \simeq 0.022$ obtain the highest level of utility with Kerry. The relative frequency of voters who will vote for Kerry is then $\widehat{S}_k = \frac{\sharp\{i=1,\ldots,N\ ;\ a_i < a_1\}}{N} \simeq 0.4751$.

The results are similar under the assumptions of sincere voting and strategic voting because the locations of Kerry ($x_k = -0.422$) and Nader ($x_n = -0.394$) are very close, and Kerry has an additive-valence advantage over Nader ($\widehat{\delta}_k > \widehat{\delta}_n$), as shown in Column [3] of Table 4. Consequently, the higher additive valence of Kerry implies a higher level of utility with Kerry than with Nader for all voters. So even under the assumption of sincere voting, nobody will vote for Nader according to the additive valence model.

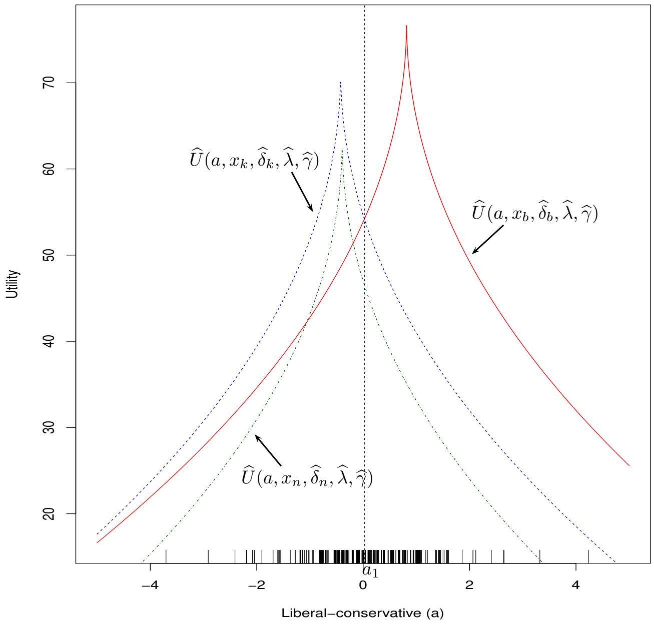


Figure B2: Estimated additive utility functions

Note: This Figure depicts the three estimated additive utilities in function of a obtained in Table 4. The (red) solid curve depicts the estimated utility if Bush is elected. The (blue) dashed curve depicts the estimated utility if Kerry is elected. The (green) dotdash curve depicts the estimated utility if Nader is elected. The locations of Kerry, Nader and Bush are $x_k = -0.422$, $x_n = -0.394$ and $x_b = 0.816$, respectively. Finally, $a_1 \simeq 0.022$.