THEMA

théorie économique, modélisation et applications

THEMA Working Paper n°2024-05
CY Cergy Paris Université, France

Additive valence and the single-crossing property

Fabian Gouret

May 2024

CERGY PARIS UNIVERSITÉ

cnrs

Additive valence and the single-crossing property*

Fabian Gouret†

May 27, 2024

Abstract

To enhance the realism of the spatial model of voting, several authors have added a valence parameter into a Downsian utility function. However, when doing so, they rarely discuss the value that the exponent on the distance between voters and candidates should take. For some values of the exponent and the valence-advantage of one candidate over another one, the single-crossing property cannot be assumed. This paper underscores the importance of this consideration by providing first a necessary and sufficient condition for this property not being satisfied. I then discuss the identification of the key parameters in two econometric frameworks to realize various hypothesis tests related to the single-crossing property. I use data from pre-election surveys of the American National Election Studies. I mainly focus on the 2008 Presidential election, and find some evidence against the single-crossing hypothesis. I also discuss the results with more recent US Presidential elections, but it is more difficult to find evidence against this hypothesis.

JEL Classification: D72, C81
Keywords: spatial models of voting, valence, single-crossing property, survey.

*I am grateful to William Kengne and Aristide Houndetoungan for discussions on computational issues, as well as Agustín Pérez-Barahona, Guillaume Chapelle, Ashley Piggins and Jean-Luc Prigent. All remaining errors are mine.

†Théma UMR8184, CY Cergy Paris University, 33 Bvd du Port, 95011 Cergy-Pontoise Cedex, France (Email: fabiangouret@gmail.com). Phone: +33134252275. Homepage: https://sites.google.com/site/fabgouret/

1 Introduction

Spatial models of voting have dominated formal political theory since the seminal work of Downs (1957). This work and the literature stemming from it have typically considered that each candidate j commits to a policy x_j in the policy space $\mathbb{R}$. And each elector votes on the basis of her "Downsian" utility function: if voter i's bliss point is $a_i \in \mathbb{R}$, then her utility if candidate j is elected is $u(a_i, x_j) = -\mathrm{L}(|a_i - x_j|)$, where L() is any strictly increasing loss function; each voter i votes for the candidate whose policy minimizes her loss. Such a theory fails to take into account that candidates possess valence characteristics, i.e., characteristics unrelated to policy selection and unanimously evaluated by voters (e.g., charisma, competence). Thus, to add realism into the spatial model, various authors have included an additive valence into the Downsian utility function and explored its implications. More precisely, if $\theta_j \in \mathbb{R}$ is the valence associated to candidate j, then the literature has usually considered that voter i's utility if candidate j is elected is:

$$u(a_i, x_j, \theta_j, \gamma) = \theta_j - |a_i - x_j|^\gamma \tag{1}$$

where $\gamma \in \mathbb{R}_+^*$ is commonly assumed to be equal to 2, i.e., the loss function $|a_i - x_j|^2$ is quadratic (e.g., Dix and Santore, 2002, Aragonès and Xefteris, 2012), or equal to 1, i.e., the loss function $|a_i - x_j|$ is absolute (e.g., Aragones and Palfrey, 2002, Hummel, 2010). But the value that γ should take is almost never discussed. This paper argues that if the objective of adding a valence parameter is to add realism into the spatial model, choosing an adequate value for γ is also crucial to confront reality because it may have strong implications on voter preferences.

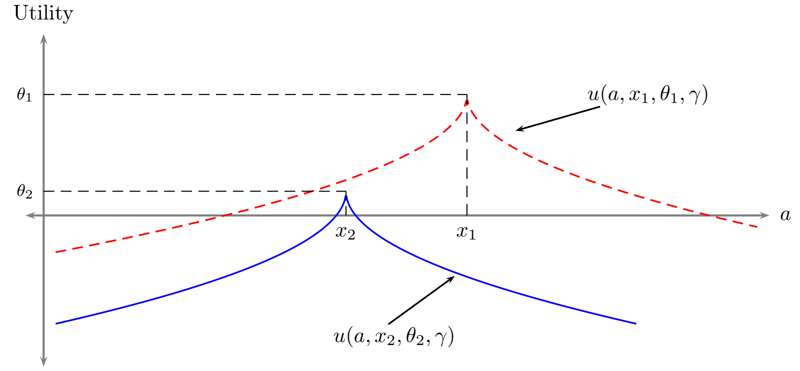
More precisely, this paper provides first a necessary and sufficient condition for the single-crossing property not being satisfied. A part of this condition states that the single-crossing property does not hold if $\gamma < 1$ *and* the distance between the policies proposed by the candidates is high enough compared to the valence-advantage that one candidate has over the other. Second, I use this condition to provide various hypothesis tests. In particular, I show that $\widehat{\gamma}$ is significantly less than one in the 2008 United States (US) Presidential election, using data from the pre-election survey of the American National Election Studies (ANES). The single-crossing property is also rejected in this wave.

Let me first provide a definition of the single-crossing property, and then explain why it is important to test if this property is satisfied or not.

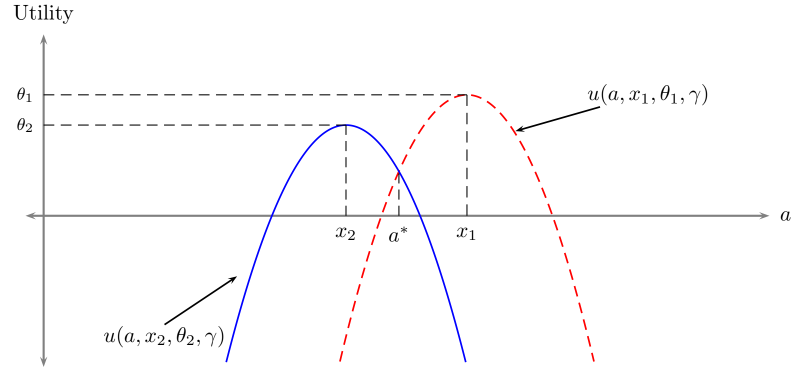
Definition 1 (Single-crossing property) Consider two candidates, indexed by $j = 1, 2$. The single-crossing property is satisfied if, for any distinct policy pair x_1, $x_2 \in \mathbb{R}$, the equation $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has at most one solution in a.

This definition is standard in models of political competition (e.g., Roemer, 1994, p.360; Ortuño-Ortín, 1997, p.431), and more generally in economic theory (e.g., Milgrom and Shannon, 1994, p.160; Quah and Strulovici, 2012, p.2333). The different panels of Figure 1 illustrate Definition 1, assuming that $x_1 > x_2$, i.e., candidate 1 locates on the right of candidate 2, and $\theta_1 > \theta_2$, i.e., candidate 1 has a valence-advantage over candidate 2. As depicted in Panel (A), when the single-crossing property is satisfied but the solution to the equation in Definition 1 is empty, *all* the voters prefer the valence-advantaged candidate. In Panel (B), the single-crossing property is also satisfied but there is one solution in a, denoted a^*. In this case, the voters whose bliss point is a^* are indifferent between the two candidates, all the voters whose bliss points are to the left of a^* prefer candidate 2, while those to the right of a^* prefer candidate 1. Thus, if the single-crossing holds, there is a "nice" separation of the set of voters for and the set of voters against a given candidate.

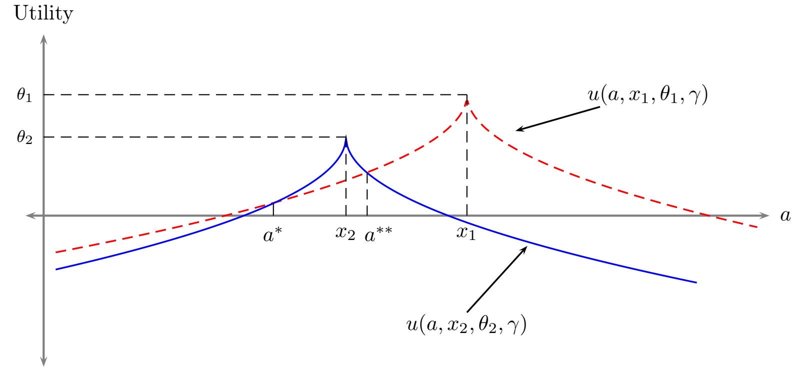
Conversely, Panel (C) depicts a situation wherein the single-crossing property is not satisfied: $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has two solutions in a, denoted a^* and a^{**}. If so, the set of voters who prefer the disadvantaged candidate in terms of valence is now the



(A) The single-crossing property is satisfied: $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has no solution in a (parameters: $\gamma = 0.5$, $x_1 = 1$, $\theta_1 = 2$, $x_2 = -1$, $\theta_2 = 0.4$)



(B) The single-crossing property is satisfied: $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has one solution in a (parameters: $\gamma = 2$, $x_1 = 1$, $\theta_1 = 2$, $x_2 = -1$, $\theta_2 = 1.5$)



(C) The single-crossing property is not satisfied: $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has two solutions in a (parameters: $\gamma = 0.5$, $x_1 = 1$, $\theta_1 = 2$, $x_2 = -1$, $\theta_2 = 1.3$)

Figure 1: The single-crossing property

open interval (a^*, a^{**}); and the set of voters who prefer the valence-advantaged candidate is a non-convex set: this candidate is supported by voters whose bliss points are on the left of a^* and on the right of a^{**}. Thus, if the single-crossing property is not satisfied, voters at the ideological ends vote together in opposition to moderates. Hence, voter preferences differ strongly from the case wherein the single-crossing property is satisfied.

When authors assume $\gamma = 2$ in Equation (1), this is an assumption of convenience, made to obtain a model which is more mathematically tractable. As I will show, the problem is that the single-crossing property is also implicitly assumed in additive-valence models which consider $\gamma = 2$ (and more generally $\gamma > 1$). One can argue that it is not a problem: it is common to assume the single-crossing property in models of spatial competition. As far as only the distance matters, the single-crossing property can be directly assumed. But if one adds a valence parameter to the Downsian framework, there are instances where the property cannot be assumed or derived, as this paper shows. One should have in mind that the single-crossing property is only an assumption that may hold in some settings; it is not a universal truth.

Authors are usually silent about the fact that the single-crossing property may not hold in an additive-valence model, or, in some cases, do not understand it at all. For instance, Hollard and Rossignol (2008, p.443), who propose a model with multiplicative valence[1], write: "This paper investigates the consequences of moving from an additive form to a multiplicative form. [...] [F]or dimension 1 [...] the electorates of both sides on the equilibrium spectrum support the most favored candidate. [...] This is in sharp contrast with the additive case which predicts a split into two intervals." I will show that this last assertion is false; as already noticed, the set of voters who prefer the valence-advantaged candidate may be a non-convex set in a model with additive valence.

To the best of my knowledge, Evrenk (2019, p.273) and Groseclose (2001, p.865) are the sole authors who explicitly mention that the single-crossing property *may not* hold if $\gamma < 1$, as well as if $\gamma = 1$; see Section 2 for a discussion of this latter case. The sufficient

[1]Under the assumptions of a unidimensional policy space and an exponent parameter equal to one, a utility function with multiplicative valence is expressed as follows: $u(a_i, x_j, \delta_j) = -\frac{1}{\delta_j}|a_i - x_j|$, where $\delta_j \in \mathbb{R}_+^*$ represents the multiplicative valence.

and necessary condition that I provide in Section 2 is obviously more complete because one of my objectives is to perform hypothesis testing of the single-crossing property; this is not the objective of Evrenk (2019) and Groseclose (2001).[2]

The general idea of this paper is simple and straightforward. It nevertheless appears to be new. Indeed, I am not aware of research that tests the single-crossing hypothesis in additive valence models. I am not aware of research that tests the hypothesis that γ is less than one either.[3] One reason of this oversight may be the fact that the parameter γ enters utility nonlinearly. To test the hypotheses of interest, one of the model that I will use is a discrete choice model. Standard computer packages, like, e.g., Stata, or those in the R environment (Croissant, 2013), only provide routines for discrete choice models with linear-in-parameters utility. Hence, the difficulty of writing a code to take account of a nonlinear-in-parameter utility may be the reason of this oversight.

Alternatively, the reason may be the complexity to place candidates and voters in a common ideological space to obtain the crucial regressor $|a_i - x_j|$. To compute it, it is natural to use survey items which ask respondents to place themselves and candidates on issue scales –typically a liberal-conservative scale. But respondents usually interpret and answer this scale differently. I will show that this problem of interpersonal incomparability of responses occurs in the data used in this paper. If so, taking these responses at face value might bias the computed distances and the final results. To solve this problem, I will use the Aldrich and McKelvey's (1977) scaling method. It is considered as an extremely satisfactory approach to correcting for interpersonal incomparability and placing candidates and voters in a common issue space (King *et al.*, 2004, p.192). In particular, it has been shown that the Aldrich-McKelvey procedure permits to recover

[2]See also Martin *et al.* (2022, p.314) who discuss the fact that $\gamma = 2$ in a multidimensional policy space is a convenient assumption because it ensures an equal-utility hyperplane.

[3]In fact, the empirical literature often assumes a squared distance, i.e., $\gamma = 2$ is imposed; see, e.g., Adams *et al.* (2005, p.17), Alvarez and Nagler (1995, p.725) or Schofield *et al.* (2011, p.492).

an accurate location of the candidates (Aldrich and McKelvey, 1977, pp.117-121). This is true, even when one of the Gauss-Markov assumptions of the method, an assumption of homoscedasticity which means that respondents are in practice equally imperfectly informed about the candidates, is strongly rejected (Palfrey and Poole, 1987, pp.514-516). This method is sometimes used; see Hollibaugh *et al.* (2013) and Gouret (2021), as well as Zakharova and Warwick (2014) who use a Bayesian version due to Hare *et al.* (2015). But generally, empirical work on valence does not deal with the problem of interpersonal incomparability, despite the fact that correcting for this problem is far better than taking the responses at face value.[4]

I arrange my presentation in the following way. Section 2 provides the necessary and sufficient condition for the single-crossing property not being satisfied. This condition permits to know the parameters that I need to identify to realize various hypothesis tests related to the single-crossing property. Section 3 discusses the identification of the parameters of interest in two frameworks that I will exploit to realize these tests. The first one is a discrete choice model –a conditional logit– where the regressand is stated choice. I show that under the assumption of sincere voting, the identification of the parameters is possible in this framework. It contrasts with a "traditional" conditional logit model where the utility is linear in all its unrestricted parameters, and where these unrestricted parameters and the scale parameter are not separately identified. The second

[4]Degan (2007) is another notable exception who corrects for problems of interpersonal incomparability. However, she does not use the Aldrich-McKelvey method. 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. This method is based on roll call votes, like in Poole and Rosenthal (1991) and Heckman and Snyder (1997). But if the NOMINATE data give the position of candidates on a common space, they do not give the position of voters. So Degan estimates a parametric distribution of voters' positions using their stated choices in two presidential elections and their characteristics. Note as well that Henry and Mourifié (2013) study how a Downsian model permits to partially identify the distribution of voter positions, assuming that each voter faces multiple elections and knows candidate positions via NOMINATE, like the econometrician. Henry and Mourifié show that voting profiles in the multiple elections are incompatible with the Downsian model. To reconcile this model with the data, they add an unobserved valence parameter to the utility, assuming that $\gamma = 2$ (see Equation (1), p.656).

framework is a seemingly unrelated regressions (SUR) model where the regressands are feeling thermometers, i.e., respondents' affect toward candidates to an election. I explain the advantages that one framework can have over the other. To keep my presentation manageable, Sections 4 and 5 focus on the 2008 US Presidential election. Section 4 describes the 2008 pre-election survey of the ANES. Section 5 provides the results. Section 6 provides some robustness checks and discusses the results with other US Presidential elections. Following all of this, Section 7 concludes. Some additional results are relegated to various appendixes.

2 Theoretical framework

Consider an election between two candidates indexed by $j = 1, 2$. Each candidate j chooses a policy platform x_j in the policy space $\mathbb{R}$. Each voter i has a bliss point $a_i \in \mathbb{R}$. Recall the utility function (1) of voter i if candidate j is elected:

$$u(a_i, x_j, \theta_j, \gamma) = \theta_j - |x_j - a_i|^\gamma$$

where $\theta_j \in \mathbb{R}$ is the valence associated to candidate j, and $\gamma \in \mathbb{R}^*_+$. I assume that candidate 1 has a valence-advantage over candidate 2, i.e., $\theta \equiv \theta_1 - \theta_2 > 0$. By Definition 1, recall that the single-crossing property is satisfied or not for any distinct pair x_1 and x_2. Without loss of generality (w.l.o.g.), I assume that $x_1 > x_2$, i.e., candidate 1 locates on the right of candidate 2.

Proposition 1 The single-crossing property is not satisfied if and only if:

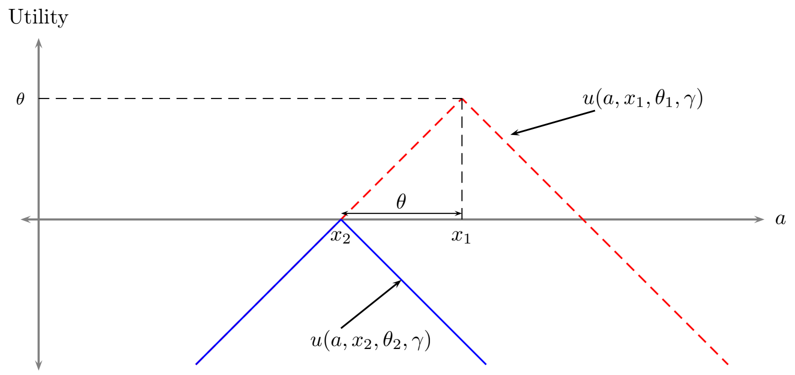
$$(|x_1 - x_2|^\gamma > \theta \quad \text{and} \quad \gamma < 1) \text{ or } (|x_1 - x_2|^\gamma = \theta \quad \text{and} \quad \gamma = 1) \tag{2}$$

Proof: see Appendix A.

Proposition 1 states that Property (2) is a necessary and sufficient condition for the

single-crossing property not being satisfied. Property (2) is a composite one, and it is useful to comment its different parts as well as the proof in Appendix A to have a good understanding of this condition. The proof is divided in three parts. It first shows that $(|x_1 - x_2|^\gamma > \theta \quad \text{and} \quad \gamma < 1)$ is a sufficient condition for the single-crossing not being satisfied. This case is exactly the one described in Panel (C) of Figure 1. Indeed, by proving that $(|x_1 - x_2|^\gamma > \theta \quad \text{and} \quad \gamma < 1)$ implies that $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has two solutions in a, denoted a^* and a^{**}, Appendix A also shows that $a^* \in (-\infty, x_2]$ and $a^{**} \in (x_2, x_1)$, as in Panel (C) of Figure 1. Note that if one had assumed that candidate 2 had the valence-advantage over candidate 1, i.e., $\theta \equiv \theta_2 - \theta_1 > 0$, he would have obtained a symmetric result: $(|x_1 - x_2|^\gamma > \theta \quad \text{and} \quad \gamma < 1)$ implies that $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has two solutions in a, which are $a^* \in (x_2, x_1)$ and $a^{**} \in [x_1, +\infty)$.
The second part of the proof shows that $(|x_1 - x_2|^\gamma = \theta \quad \text{and} \quad \gamma = 1)$ is also a sufficient condition for the single-crossing property not being satisfied. More precisely, it shows that $(|x_1 - x_2|^\gamma = \theta \quad \text{and} \quad \gamma = 1)$ implies $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ $\forall a \in (-\infty, x_2]$, i.e., there are an infinite number of solutions. It corresponds to the situation depicted by Figure 2 wherein $a \mapsto u(a, x_1, \theta_1, \gamma)$ and $a \mapsto u(a, x_2, \theta_2, \gamma)$ are coincident lines on $(-\infty, x_2]$. Obviously, if one had assumed that candidate 2 had the valence-advantage over candidate 1, i.e., $\theta \equiv \theta_2 - \theta_1 > 0$, he would have obtained that $(|x_1 - x_2|^\gamma = \theta \quad \text{and} \quad \gamma = 1)$ implies $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ $\forall a \in [x_1, +\infty)$.

The third part of the proof shows that Property (2) is a necessary condition for the single-crossing property not being satisfied. To do so, I use the fact that a statement and its contrapositive are equivalent. Thus, the proof first provides the negation of Property (2), denoted $\neg P$; see Equations (A8)-(A9) in Appendix A. Then it shows that $\neg P$ implies the single-crossing property. Now, given that Property (2) is a necessary and sufficient condition for the single-crossing property not being satisfied, and because, again, a state-



The single-crossing property is not satisfied: $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$ has infinitely many solutions in a (parameters: $\gamma = 1$, $x_1 = 1$, $\theta_1 = 2$, $x_2 = -1$, $\theta_2 = 0$)

Figure 2: $|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$

ment and its contrapositive are equivalent, Corollary 1 can be deduced from Proposition 1; Property (3) is $\neg P$ as shown in Equation (A9) of Appendix A.

Corollary 1 The single-crossing property is satisfied if and only if:

$$\begin{aligned}(\gamma > 1) \quad \text{or} \quad (|x_1 - x_2|^\gamma < \theta) \quad \text{or} \quad (|x_1 - x_2|^\gamma \leq \theta \quad \text{and} \quad \gamma \neq 1) \\ \text{or} \qquad (\gamma \geq 1 \quad \text{and} \quad |x_1 - x_2|^\gamma \neq \theta)\end{aligned} \tag{3}$$

Lastly, I think it prudent to emphasize that it is only the introduction of a difference in quality among candidates which makes that the single-crossing property may not hold when the loss function $|a_i - x_j|^\gamma$ in Equation (1) is concave, i.e., $\gamma < 1$, or linear, i.e., $\gamma = 1$. If there is no difference in quality among candidates, i.e., $\theta = 0$, the utility function (1) is Downsian, i.e., $u(a_i, x_j, \gamma) = -|x_j - a_i|^\gamma$. As it is well known, if the utility function is Downsian, the single-crossing property is satisfied for all $\gamma \in \mathbb{R}_+^*$; see, e.g., Evrenk (2019).

3 Econometrics

This Section presents the two frameworks that I will exploit to provide various statistical tests related to the single-crossing property. To realize these tests, I first need to discuss the identification of the parameters γ and θ which appear in Proposition 1 and

Corollary 1. Subsection 3.1 discusses the identification of these parameters in my first framework, a discrete choice model where the regressand is stated choice. I then present the statistical tests that these identified parameters permit to realize. Subsection 3.2 discusses this identification in a second framework, a SUR model where the regressands are feeling thermometers. Subsection 3.2 also explains the advantages that one framework can have over the other.

3.1 Stated choice and discrete choice analysis

Consider a survey of N voters, $i = 1, \ldots, N$, representative of the electorate of an election. J candidates compete for this election, where $J \geq 2$. Observe that if $J > 2$, a voter may vote strategically, i.e., she may be willing to vote for another candidate than her most preferred candidate if this latter is unlikely to win. However, in the discrete choice model, I must assume that for all $J \geq 2$, each voter i is sincere, i.e., she votes for her most preferred candidate. Let y_i denote the stated choice of respondent i and $U_{i,j}$ her utility if candidate j is elected; $U_{i,j} = u(a_i, x_j, \theta_j, \gamma) + \varepsilon_{i,j}$, where $u(a_i, x_j, \theta_j, \gamma)$ is the deterministic component of the utility given by Equation (1) and $\varepsilon_{i,j}$ a random component to utility. The probability that respondent i states that she will vote for candidate j is:

$$\begin{aligned} \mathrm{P}\left[y_i = j\right] &= \mathrm{P}\left[U_{i,j} > U_{i,k}, \quad \forall k \neq j\right] \\ &= \mathrm{P}\left[\theta_j - |a_i - x_j|^\gamma + \varepsilon_{i,j} > \theta_k - |a_i - x_k|^\gamma + \varepsilon_{i,k}, \quad \forall k \neq j\right] \end{aligned} \tag{4}$$

Before to step any further, two remarks are in order concerning Equation (4). First, the valence parameters θ_j and θ_k are not identified; what is identified is the difference in valence. To fully understand, consider w.l.o.g. that $\theta_j > \theta_k$. Note now that $U_{i,j} = \theta_j^0 - |a_i - x_j|^\gamma + \varepsilon_{i,j}$ and $U_{i,k} = \theta_k^0 - |a_i - x_k|^\gamma + \varepsilon_{i,k}$, with $\theta = \theta_j^0 - \theta_k^0$, is equivalent to a model with $U_{i,j} = \theta_j^1 - |a_i - x_j|^\gamma + \varepsilon_{i,j}$ and $U_{i,k} = \theta_k^1 - |a_i - x_k|^\gamma + \varepsilon_{i,k}$, where the difference in valence is the same, i.e., $\theta_j^1 - \theta_k^1 = \theta = \theta_j^0 - \theta_k^0$. Indeed, it is impossible to identify the valence parameters because an infinite number of values of the two valence parameters have the same difference, so result in the same choice probabilities. To account for this fact, we must normalize the valence of one candidate. But one should have in mind that the impossibility to identify the valence parameters is of no practical consequence because it is the valence-advantage θ of one candidate over another one which matters in Proposition 1, not the level of valence. I will present the results with the valence of the candidate with the lowest valence normalized to zero. So, for the ease of exposition, I consider that candidates are ranked according to their valence as such: $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_{J-1} \geq \theta_J = 0$, with at least one inequality which must be strict (otherwise, the model is Downsian).

The second remark concerns the scale of the utility and the distribution for the disturbances. To recover the parameters $\{\theta_j\}_{j=1}^{J-1}$ and γ which will be crucial to form appropriate test statistics, it is key to understand that the econometrician must scale the utility $U_{i,j}$ in the estimation procedure. To fully understand why, note that the scale is not identified (e.g., Ruud, 2000, pp.765-766), i.e., multiplying or dividing the utility of each alternative by a common strictly positive constant α does not affect the probability in Equation (4): $\mathrm{P}\left[U_{i,j} > U_{i,k},\ \forall k \neq j\right] = \mathrm{P}\left[\frac{U_{i,j}}{\alpha} > \frac{U_{i,k}}{\alpha},\ \forall k \neq j\right]$. Hence, the econometrician must choose the scale, and this is done by normalizing the variance of the disturbances. I assume, as it is commonly done, that the J disturbances are independent and identically distributed (i.i.d.) with type I extreme value distribution. The variance of the type I extreme value distribution is $\frac{\pi^2}{6}$. Obviously, there is no reason to assume that the variance of $\varepsilon_{i,j}$ is $\frac{\pi^2}{6}$: the disturbance $\varepsilon_{i,j}$ has a variance which may be any positive number. However, this variance can be expressed w.l.o.g. as a multiple of $\frac{\pi^2}{6}$, i.e., $\mathrm{Var}(\varepsilon_{i,j}) = \sigma^2\frac{\pi^2}{6}$. If so, setting the variance to $\frac{\pi^2}{6}$ implies to divide the utility of each alternative by the scale parameter σ: $\frac{U_{i,j}}{\sigma} = \frac{u(a_i,x_j,\theta_j,\gamma)}{\sigma} + \varepsilon^*_{i,j}$, with $\varepsilon^*_{i,j} \equiv \frac{\varepsilon_{i,j}}{\sigma}$. It does not affect the probability in Equation (4), as described above, and the variance of the scaled disturbance is $\text{Var}\left(\frac{\varepsilon_{i,j}}{\sigma}\right) = \text{Var}(\varepsilon^*_{i,j}) = \frac{\pi^2}{6}$. Now, if $\varepsilon^*_i \equiv (\varepsilon^*_{i,j};\ j = 1, 2, \ldots, J)$ are independent and identically distributed (i.i.d.) with type I extreme value distribution, the probability of the ith individual stating that she will vote for candidate j has the following conditional logit form (McFadden, 1974):

$$\text{P}\left[y_i = j\right] = \frac{\exp\left(\frac{\theta_j}{\sigma} - \frac{1}{\sigma}|a_i - x_j|^\gamma\right)}{\sum_{k=1}^{J} \exp\left(\frac{\theta_k}{\sigma} - \frac{1}{\sigma}|a_i - x_k|^\gamma\right)} \tag{5}$$

It is crucial to include σ in the specification to avoid any bias. Note also that σ^{-1} is identified, so are σ, $\{\theta_j\}_{j=1}^{J-1}$ and γ. It contrasts with a "traditional" conditional logit where the utility is specified to be linear in all its unrestricted parameters; in this case, the unrestricted coefficients and the scale parameter are not separately identified (Train, 2009, p.41).[5]

Define $d_{i,j} = 1$ if $y_i = j$ and zero otherwise. Then, the log-likelihood function is:

$$\ell = \ln(L_N) = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{i,j} \ln\left[\text{P}\left(y_i = j\right)\right] \tag{6}$$

Substituting (5) into (6), the log-likelihood function is rewritten as:

$$\ell = \sum_{i=1}^{N} \left[\left(\sum_{j=1}^{J} d_{i,j} \left(\frac{\theta_j}{\sigma} - \frac{1}{\sigma}|a_i - x_j|^\gamma \right) \right) - \ln\left(\sum_{k=1}^{J} \exp\left(\frac{\theta_k}{\sigma} - \frac{1}{\sigma}|a_i - x_k|^\gamma \right) \right) \right]$$

[5] When the utility is specified to be linear in all its unrestricted coefficients, i.e., $U_{i,j} = X_{i,j}\beta + \varepsilon_{i,j}$, where $X_{i,j}$ is a vector of variables that relate to alternative j, β an unrestricted vector of coefficients of these variables, and $\varepsilon_{i,j}$ a disturbance whose variance may be any positive number, the conditional logit is $\text{P}\left[y_i = j\right] = \frac{\exp(X_{i,j}\frac{\beta}{\sigma})}{\sum_{k=1}^{J}\exp(X_{i,k}\frac{\beta}{\sigma})}$. However, such a conditional logit is usually expressed in its scaled form, with $\beta^* = \beta/\sigma$: $\text{P}\left[y_i = j\right] = \frac{\exp(X_{i,j}\beta^*)}{\sum_{k=1}^{J}\exp(X_{i,k}\beta^*)}$. The parameters β^* are estimated, but some generalist textbooks in econometrics do not mention explicitly that these parameters are actually the "true" coefficients β divided by σ; they only say that the estimates of β^* in a logit model are the probit estimated coefficients multiplied by $\frac{\pi}{\sqrt{3}}$. This difference is due to the variance of $\varepsilon^*_{i,j}$ which is normalized to one in a probit and $\frac{\pi^2}{3}$ in a logit. The fact that some generalist textbooks in econometrics are not explicit on this issue is probably due to the fact that only the ratio $\frac{\beta}{\sigma}$ may be estimated if the utility is linear in all its unrestricted parameters, i.e., β and σ are not separately identified in that case. Train (2009, p.24), whose seminal book focuses on discrete choice models, is a notable exception.

I code this log-likelihood in the R environment and use the optim function to maximize it.

Now, recall that the estimation of this discrete choice model and the identification of the parameters γ and $\{\theta_j\}_{j=1}^{J-1}$ will serve to realize various statistical tests related to the single-crossing property. Consider two of the J candidates, say candidates 1 and 2. There are three hypotheses to be tested concerning these two candidates. First, the single-crossing property holds for all $\gamma \in \mathbb{R}_+^*$ if there is no difference in valence among these two candidates. This is the introduction of a valence difference between candidates 1 and 2 which makes that the single-crossing property may not hold. So the first null hypothesis is $\mathrm{H}_0: \ \theta_1 = \theta_2$ versus $\mathrm{H}_A: \ \theta_1 \neq \theta_2$. Recall that the valence-advantage of one candidate over another one is denoted θ, so the hypothesis can be restated as $\mathrm{H}_0: \ \theta = 0$ versus $\mathrm{H}_A: \ \theta \neq 0$.

The second hypothesis to be tested concerns γ. When one candidate has a valence-advantage over another one, the single-crossing property may not hold if $\gamma < 1$ or if $\gamma = 1$. I believe that the most interesting case is when $\gamma < 1$ because it is when $u(a, x_1, \theta, \gamma) = u(a, x_2, \gamma)$ may have two solutions in a, and an "ends against the middle" behavior may occur, as depicted in Panel (C) of Figure 1. When $\gamma = 1$, the single-crossing property may not hold because $u(a, x_1, \theta, \gamma) = u(a, x_2, \gamma)$ may have an infinity of solutions as depicted in Figure 2. But this case is not a situation where voters at the ideological ends vote together for the same candidate; it is only that an infinity of voters are indifferent between the two candidates. So the second null hypothesis that I will test is $\mathrm{H}_0: \ \gamma \geq 1$ versus $\mathrm{H}_A: \ \gamma < 1$.

Last but not least, the third hypothesis to be tested concerns the single-crossing property per se. Consider again two of the J candidates, candidates 1 and 2. I want to test H_0 : "the single-crossing property holds" versus H_A : "the single-crossing property does not hold". Using the conditions in Proposition 1 and Corollary 1, the null hypothesis and its alternative can be restated formally as:

$$\mathrm{H}_0 : (\gamma > 1) \quad \text{or} \quad (|x_1 - x_2|^\gamma < \theta) \quad \text{or} \quad (|x_1 - x_2|^\gamma \leq \theta \quad \text{and} \quad \gamma \neq 1)$$

$$\text{or} \quad (\gamma \geq 1 \quad \text{and} \quad |x_1 - x_2|^\gamma \neq \theta)$$

$$\text{versus}$$

$$\mathrm{H}_A : [(|x_1 - x_2|^\gamma > \theta \text{ and } \gamma < 1) \text{ or } (|x_1 - x_2|^\gamma = \theta \text{ and } \gamma = 1)]$$

To test these hypotheses, I will exploit bootstrap methods. This approach is particularly attractive for the third test hypothesis, given the complexity of the null hypothesis. But the approach is also useful for the other test hypotheses. Indeed, the distances between the candidates and the respondents' bliss points, i.e., the key regressors in Equation (5), will be based on a first-step estimate, the Aldrich-McKelvey scaling method to solve problems of interpersonal incomparability of responses. Bootstrap methods permit to take into consideration any estimation error in this first-step estimation; see Appendix D for further details.

3.2 Thermometer scores and system of regression equations

The second framework that I use is a SUR model where the regressands are feeling thermometers. Feeling thermometers are sometimes regarded as the best available measures of voters' utility from political alternatives (e.g., Armstrong II *et al.*, 2014, p.147). There are at least two advantages to using thermometer scores. First, it is often considered (implicitly) that responses to feeling thermometer questions reveal sincere preferences, so they are often exploited when there are problems of strategic voting (e.g., Abramson *et al.*, 1992, Black, 1978, Blais and Nadeau, 1996, Gouret, 2021). It contrasts with discrete choice models with stated choices as the one described in Subsection 3.1 which might generate biased estimates of the parameters of the utility functions if $J > 2$. Indeed, when $J > 2$, a voter may not vote for her preferred candidate because of strategic voting. One may argue that using stated choice data from Presidential elections in the US permits to avoid this bias because these elections are often described as two-candidate elections. In reality, it only minimizes this bias, but does not eliminate it, because the US Presidential elections have experienced the rise of third-candidate challengers (Alvarez and Nagler, 2000, pp.58-59): Perot in 1992 or Nader in 2000 and 2004 are notable examples. Anticipating Section 4, note, however, that if there were more than two candidates at the 2008 Presidential election, each respondent in the ANES was asked if she would vote for Obama, McCain, none, or another candidate but without specifying his or her name. Similarly, the respondents were asked to provide thermometer scores for Obama and McCain, but not for the other candidates. So in practice, we will operate as if there were only the two main candidates, Obama and McCain. It implies that by using stated choice, the respondents who answered that they would not vote or vote for another candidate will appear as missing observations. In contrast, even if a respondent indicated that she would not vote or vote for another candidate, she could provide thermometer scores for Obama and McCain. This is the second advantage of using thermometer scores: it reduces the number of missing observations. Given these advantages, the next step is to know if one can identify accurately the parameters of interest in the SUR model; I explain below.

One should have in mind that the different utility functions considered are unique up to positive affine transformations. Thus, the utility function $u(a_i, x_j, \theta_j, \gamma) = \theta_j - |x_j - a_i|^\gamma$ is equivalent to $v(a_i, x_j, \theta_j, \gamma, c, \beta) = c + \beta\,(\theta_j - |x_j - a_i|^\gamma)$ for some scalar c and some scalar $\beta > 0$ independent of i and j. The values of these scaling parameters c and β will depend on the scales used in the questions of the survey; for instance, feeling thermometer questions in the ANES ask respondents to rate on a 100-point scale their affect toward different candidates.

Let $V_{i,j}$ denote the thermometer score or the utility of respondent i if candidate j is elected; $V_{i,j} = v(a_i, x_j, \theta_j, \gamma, c, \beta) + \xi_{i,j}$, where $v(a_i, x_j, \theta_j, \gamma, c, \beta)$ is the deterministic component of the utility and $\xi_{i,j}$ a random component to utility. The model is the following system of equations:

$$\begin{cases} V_{i,1} & = c + \beta\theta_1 - \beta|x_1 - a_i|^\gamma + \xi_{i,1} \\ V_{i,2} & = c + \beta\theta_2 - \beta|x_2 - a_i|^\gamma + \xi_{i,2} \\ \vdots & = \quad \vdots \\ V_{i,J} & = c + \beta\theta_J - \beta|x_J - a_i|^\gamma + \xi_{i,J} \end{cases} \tag{7}$$

Two remarks are in order concerning System (7). First, it imposes some cross-equation restrictions: c, β and γ are not equation-specific. 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, i.e., $\mathbb{E}\left[\varepsilon_{i,j}\varepsilon_{i',j'}\right] = \sigma_{j,j'}$ if $i = i'$ and 0 otherwise. The second remark concerns the fact that System (7) is in fact not estimable as such. To understand why, note first that System (7) is linear in the parameter c, but it is not linear in the other parameters β, θ_1, θ_2, ..., θ_J and γ. However, if one lets $\delta_j = \beta\theta_j$ for all j, then the system becomes linear in c, δ_1, δ_2, ..., δ_J. Now, stacking all J utilities for the ith voter, we get:

$$\begin{bmatrix} V_{i,1} \\ V_{i,2} \\ \vdots \\ V_{i,J} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & |a_i - x_1|^\gamma \\ 1 & 0 & 1 & \dots & 0 & |a_i - x_2|^\gamma \\ & & & \vdots & & \\ 1 & 0 & 0 & \dots & 1 & |a_i - x_J|^\gamma \end{bmatrix} \begin{bmatrix} c \\ \delta_1 \\ \delta_2 \\ \vdots \\ \delta_J \\ -\beta \end{bmatrix} + \begin{bmatrix} \xi_{i,1} \\ \xi_{i,2} \\ \vdots \\ \xi_{i,J} \end{bmatrix}$$

Consider the first matrix on the right-hand side. The sum from the second to the $J+1$ columns of this $J \times (J+2)$ matrix is equal to one, and reproduces the first column –a case of perfect multicollinearity. Then, in order to estimate the model, one must drop one of these variables, or, this is the same, normalized c or one of the δ_j, $j = 1 \ldots J$, to zero. By normalizing one δ_j to zero, one considers that the corresponding θ_j is equal to zero, given that $\delta_j = \beta\theta_j$. I will present the results with the valence of the candidate with the lowest valence normalized to zero, to compare the estimated parameters $\{\widehat{\theta}_j\}_{j=1}^{J-1}$ and $\widehat{\gamma}$ of the SUR model with those of the discrete choice model.

I will provide maximum likelihood estimates of the SUR models. They are obtained by iterated feasible generalized least squares (see, e.g., Ruud, 2000, p.706). The computation is carried out in the R environment and makes use of the nlsur package realized by Jan Marvin Garbuszus.[6] Concerning the hypotheses to be tested, they are the same as those described in Section 3.1. Bootstrap methods will be used to take into consideration the fact that the key regressors, i.e., the distances between the respondents and the candidates, are estimates based on the Aldrich-McKelvey scaling method; see Appendix D for further details.

4 The data

The data used in Sections 4 and 5 are drawn from the 2008 pre-election ANES. This survey is produced by Stanford University and the University of Michigan. It began on September 2, 2008 and ended November 3, 2008. No interviewing was conducted on Election Day, November 4. The sample was structured to be representative of the electorate. 2322 respondents were interviewed. Each respondent i was asked four key questions for the analysis. First, each respondent i was asked if she will vote for Barack

[6] The package is available at: `https://rdrr.io/github/JanMarvin/nlsur/`.

Obama ($y_i = 1$) or John McCain ($y_i = 0$). The question reveals if the respondent will vote for one of the four other candidates at this election, but without specifying his or her name; the four other candidates were Cynthia McKinney (from the Green party), Charles Baldwin (from the Constitution party), Robert Barr Jr. (from the Libertarian party) and Ralph Nader who ran his campaign independently.[7] The discrete variable y_i will be the regressand of the discrete choice model described in Section 3.1. Second, each respondent was asked to rate her affect toward Obama ($V_{i,o}$) and McCain ($V_{i,m}$); respondents were not asked to rate their affect toward the other candidates.[8] These feeling thermometers will be the regressands of the SUR model described in Section 3.2. Third, 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).[9] Lastly, each respondent i was

[7]More precisely, the question is:

> Who do you think you will vote for in the election for President?

or, if the respondent said previously that he will not vote:

> If you were going to vote, who do you think you would vote for in the election for President?
> [1] Barack Obama, [2] John McCain, [5] None, [7] Other.

[8]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 the meaning of 0, 15, 30, 40, 60, 70, 85 and 100 degrees. Then, the interviewer asked respondent i to rate her affect toward the two main Presidential candidates. For instance, for Obama, the question was:

> How would you rate BARACK OBAMA?

[9]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

also asked to place Obama ($\widetilde{x}_{i,o}$) and McCain ($\widetilde{x}_{i,m}$) on this 7-point liberal-conservative scale; respondents were not asked to place the other candidates.[10] 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 provided a "Don't know" to the stated choice question).

It would have been tempting to use the self-placement $\widetilde{a}_i$ and the reported placement of the candidates $\widetilde{x}_{i,j}$, $j = m, o$, to obtain the two distances of interest $\widetilde{d}_{i,j} = |\widetilde{x}_{i,j} - \widetilde{a}_i|$ for each respondent i. However, using $\widetilde{x}_{i,j}$ and $\widetilde{a}_i$ to compute the distances 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. For example, a liberal respondent may place a conservative candidate more on the right than do conservative respondents to exaggerate the distance between her and this candidate she views unfavorably. Appendix B shows that this form of interpersonal incomparability of responses is particularly true for McCain. Thus, any econometrics using reported locations of candidates and self-placements at face value cannot be considered as credible because the distances will be biased. I follow Aldrich and McKelvey (1977) to recover the underlying locations of the candidates and the respondents on a common dimension, the real line $\mathbb{R}$. Their method permits to obtain estimates of the actual positions of the candidates x_j, $j = m, o$, and respondents' bliss points a_i, $i = 1, \ldots, N$, on this dimension, using respondents' reported positions of

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, [-7] Haven't thought much about it, [-8] Don't know, [-9] Refused.

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

Where would you place BARACK OBAMA on this scale?

the stimuli (candidates and parties) as well as their self-placement. As already noticed, this method is known to produce accurate locations of the candidates; see Appendix C for a careful presentation of this method. The method minimizes a sum over all the respondents of squared errors subject to a technical assumption (i) which is that the sum of the true locations of the stimuli is zero and the sum of squares equal to one (Equation (C3) in Appendix C). However, as shown in Remark C2 of Appendix C, if one considers the respondents' reported positions of only two stimuli, Obama and McCain here, the Aldrich-McKelvey technical assumption (i) fully determines the locations of the stimuli if it is combined with an extra assumption (ii) that the researcher knows the stimulus the more conservative. Under these two assumptions, the obtained locations are $-\sqrt{\frac{1}{2}}$ and $\sqrt{\frac{1}{2}}$ whatever the two stimuli considered. In other words, with only two stimuli, the reported positions of these stimuli by the respondents become irrelevant in determining their actual location; the locations of these stimuli are no longer the outcome of the minimization of a sum of squared errors. For a sound analysis, the policy space should be built based on an estimator having an optimality property. This is the reason why I also include the reported placement of the Democratic ($\widetilde{x}_{i,dem}$) and Republican ($\widetilde{x}_{i,rep}$) parties in the Aldrich and McKelvey (1977) procedure to obtain the estimated locations of Obama (x_o) and McCain (x_m). An additional advantage of incorporating these two extra stimuli in the procedure is that if one only includes the reported positions of the two stimuli $\widetilde{x}_{i,o}$ and $\widetilde{x}_{i,m}$, it will be impossible to obtain the bliss point of respondent i if she locates Obama and McCain at the same place (i.e., $\widetilde{x}_{i,o} = \widetilde{x}_{i,m}$); see in particular Remark C3 in Appendix C. Indeed, the absence of variability in the reported position of the stimuli makes it impossible to estimate the Aldrich-McKelvey respondent-specific distortion parameters denoted c_i and w_i in Appendix C. Without these distortion parameters, it is then impossible to obtain the respondents' bliss points in the same policy space as the actual locations of the candidates. By introducing two additional stimuli into the procedure, the likelihood of encountering a lack of variability in the reported stimulus locations diminishes. Consequently, this enhances the likelihood of placing respondent i's bliss point within the same policy space as the actual locations of the candidates. That is why Table 1 not only includes the reported placements of Obama and McCain but also those of the Democratic and the Republican parties.

Table 1: Descriptive statistics

	Original sample					**Discrete choice sample**				
Variable	Obs.	Mean	Median	Min	Max	Obs.	Mean	Median	Min	Max
$y_i = 1$ (Obama)	2017	0.673	1	0	1	1320	0.615	1	0	1
$U_{i,o}$	2293	64.272	70	0	100					
$U_{i,m}$	2283	48.588	50	0	100					
$\widetilde{a}_i$	1626	4.139	4	1	7	1320	4.151	4	1	7
$\widetilde{x}_{i,o}$	2099	3.287	3	1	7	1320	2.950	2	1	7
$\widetilde{x}_{i,m}$	2097	4.857	5	1	7	1320	5.113	6	1	7
$\widetilde{x}_{i,dem}$	2127	3.313	3	1	7	1320	3.003	2	1	7
$\widetilde{x}_{i,rep}$	2110	4.953	5	1	7	1320	5.222	6	1	7
Candidates' locations and respondents' bliss points according to the Aldrich-McKelvey method										
a_i						1320	-0.046	-0.004	-3.600	4.347
x_o						1	-0.518			
x_m						1	0.478			

	SUR sample				
Variable	Obs.	Mean	Median	Min	Max
$y_i = 1$ (Obama)					
$U_{i,o}$	1449	62.501	70	0	100
$U_{i,m}$	1449	51.500	50	0	100
$\widetilde{a}_i$	1449	4.137	4	1	7
$\widetilde{x}_{i,o}$	1449	2.966	2	1	7
$\widetilde{x}_{i,m}$	1449	5.107	6	1	7
$\widetilde{x}_{i,dem}$	1449	3.022	3	1	7
$\widetilde{x}_{i,rep}$	1449	5.223	5	1	7
Candidates' locations and respondents' bliss points according to the Aldrich-McKelvey method					
a_i	1449	-0.054	-0.012	-3.652	4.572
x_o	1	-0.518			
x_m	1	0.478			

I will consider two estimation samples, as described in Table 1: one, called the Discrete choice sample, where the regressand is stated choice (i.e., $y_i = 1$ if respondent i will vote for

Obama, and 0 if he will vote for McCain), and another one, called the SUR sample, where the regressands are the thermometer scores; the names used to describe these samples refer to the econometric methods employed to identify the parameters of interest within each sample, as described in Section 3. I explain below how they have been built.

Concerning the Discrete choice sample, observations where either the stated choice (y_i), the self-placement ($\widetilde{a}_i$), or the reported location of one of the stimuli ($\widetilde{x}_{i,o}$, $\widetilde{x}_{i,m}$, $\widetilde{x}_{i,dem}$ or $\widetilde{x}_{i,rep}$) are missing are dropped. It reduces the sample size to 1346. Furthermore, 26 additional observations are excluded. These 26 respondents locate the four stimuli at the same place (i.e., $\widetilde{x}_{i,o} = \widetilde{x}_{i,m} = \widetilde{x}_{i,dem} = \widetilde{x}_{i,rep}$). As already explained, this absence of variability in the reported location of the four stimuli makes it impossible to obtain via the Aldrich-McKelvey procedure these respondents' bliss points in the same policy space as the actual locations of the candidates, and then compute accurate distances. Thus, the Discrete choice sample is composed of 1320 respondents ($= 1346 - 26$), implying an effective response rate of about 57% $\left(\simeq \frac{1320}{2322}\right)$.

Concerning the SUR sample, observations where either one of the thermometer scores ($V_{i,o}$ or $V_{i,m}$), the self-placement ($\widetilde{a}_i$), or the reported location of one of the stimuli ($\widetilde{x}_{i,o}$, $\widetilde{x}_{i,m}$, $\widetilde{x}_{i,dem}$ or $\widetilde{x}_{i,rep}$) are missing are dropped. It reduces the sample size to 1483. Furthermore, 34 additional observations are excluded because these respondents locate the four stimuli at the same place (i.e., $\widetilde{x}_{i,o} = \widetilde{x}_{i,m} = \widetilde{x}_{i,dem} = \widetilde{x}_{i,rep}$); so, again, the absence of variability makes it impossible to obtain these respondents' bliss points. Thus, the SUR sample is composed of 1449 respondents ($= 1483 - 34$), implying an effective response rate of about 63% $\left(\simeq \frac{1449}{2322}\right)$.

As reported in Table 1, the number of observations in the SUR sample is slightly higher than the one in the Discrete choice sample (1449 versus 1320). Despite of this difference, the summary statistics are very similar. The estimated locations of Obama and McCain are almost similar: $x_o = -0.5180862$ and $x_m = 0.4780555$ in the Discrete choice sample, while $x_o = -0.5180191$ and $x_m = 0.4783012$ in the SUR sample.[11] Concerning the bliss points of the respondents, they range from a low of -3.600 to a high of 4.347 in the Discrete choice sample, and from a low of -3.652 to a high of 4.572 in the SUR sample.

Note that the missing observations in the Discrete choice and SUR samples are mainly due to the self-placement question: of the 2322 respondents of the survey, only 1626 provided a valid answer, i.e., one of the 7 points of the scale, as shown in the Original sample of Table 1. Of the 696 other respondents ($= 2322 - 1626$), 675 answered that they "Haven't thought much about it", 17 that they "Don't know", and 4 refused to answer ($696 = 675 + 17 + 4$); see the question described in Footnote 9. It is a priori possible to reduce the number of missing observations. Indeed, the self-placement question is followed by another question which asks those who answered they "Haven't thought much about it" or "Don't know", as well as those who considered themselves as "Moderate/middle of the road" (i.e., $\widetilde{a}_i = 4$), how they would consider themselves if they had to choose. This follow-up question proposes three possible valid answers: "Liberal", "Conservative" and "Moderate". Exploiting the answers to the follow-up question to reduce the missing observations implies to make some assumptions about the coding. For instance, think about someone who answered initially "Moderate/middle of the road" or "Haven't thought much about it" to the self-placement question and who then responded "Liberal" to the follow-up question. Should one code this respondent as liberal (i.e., $\widetilde{a}_i = 2$) or as slightly liberal (i.e., $\widetilde{a}_i = 3$)? If one codes this respondent as liberal, he will consider that she is more liberal than those who answered slightly liberal to the self-placement question. Yet, when a respondent initially answered that she hasn't thought much about it or that she

[11]The computations of the actual locations of the stimuli are carried out in the R environment and make use of the basicspace package (Poole *et al.*, 2016). Table 1 does not report the estimated locations of the Democratic and the Republican parties, given that they are useless for the estimations of the discrete choice and the SUR models. Remark C4 in Appendix C provides more details.

is moderate, categorizing this person as more liberal than those who responded slightly liberal to the self-placement question can be challenging. I set these concerns aside for the moment. I shall present a set of robustness checks in Section 6 wherein I will exploit this follow-up question to reduce the missing observations.

5 Estimation results

Table 2 provides the results for the 2008 Presidential election. Column (A) provides the discrete choice model estimates. The first result concerns the estimated valence-advantage which is $\widehat{\theta} = 0.269$; Obama is the candidate who has this advantage. If this valence-advantage parameter adds realism to the spatial model, then the null hypothesis $\mathrm{H}_0 : \ \theta = 0$ should be rejected. The estimated standard error of $\widehat{\theta}$ is $\widehat{se}(\widehat{\theta}) = 0.033$. If so, the test statistic is $t = \frac{\widehat{\theta}-0}{\widehat{se}(\widehat{\theta})} = 7.995$, which is far larger than the critical values of the standard normal distribution for conventional levels of significance. Thus, the null is rejected, and one may conclude that valence matters. However, a problem with this test is that it treats the distances between a_i and x_j as observed variables, i.e., ignoring any estimation error in these variables. But the uncertainty in the estimates of these variables in a first step can influence the test statistic in the second step. To take into account this problem, Column (A) in Table 2 also provides a 95 percent confidence interval based on bootstrap percentiles; the complete procedure is detailed in Appendix D. The percentile method uses the 2.5th and the 97.5th percentiles of the empirical distribution of $B = 999$ bootstrap estimates $\widehat{\theta}^*_b$, $b = 1, \ldots, B$. Denote by $\widehat{\theta}^*_{0.025}$ and $\widehat{\theta}^*_{0.975}$ these percentiles. The percentile 95 percent confidence interval for θ is then $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$=[0.206, 0.338]. Given that $\mathrm{H}_0 : \theta = 0$ lies outside this interval, the null is rejected again.

The second result concerns $\widehat{\gamma} = 0.818$. In the case of γ, the interest is in the one-tailed test $\mathrm{H}_0 : \gamma \geq 1$ versus $\mathrm{H}_A : \gamma < 1$, as already noticed in Subsection 3.1. It is also possible to build a percentile 95 percent confidence interval based on a lower-tailed test. Such an interval, which is open all the way out to infinity in one direction, is build using the 95th percentile of the empirical distribution of $B = 999$ bootstrap estimates $\widehat{\gamma}^*_b$, $b = 1, \ldots, B$. The upper 0.05 quantile of the bootstrap estimates is 0.988, so the percentile 95 confidence interval is $(-\infty, 0.988]$ in this case. The null $\mathrm{H}_0 : \gamma \geq 1$ falls outside this interval, so $\widehat{\gamma}$ is significantly less than 1 at the 5 percent significance level. If one had ignored any estimation error in a_i and x_j and had used a simple t-test, he would have made the same conclusion: the null is also rejected at the 5 percent significance level. Indeed, the estimated standard error of $\widehat{\gamma}$ is $\widehat{se}(\widehat{\gamma}) = 0.097$, so the test statistic is $t = \frac{\widehat{\gamma}-1}{\widehat{se}(\widehat{\gamma})} = -1.876$, which is less than -1.64, the one-tailed 5 percent critical value. So, again, $\widehat{\gamma}$ is significantly less than 1 at the 5 percent significance level.

The third result concerns the single-crossing property. Abstracting for the moment from sampling variation, it is easy to see that the single-crossing does not hold when considering the estimates from either the discrete choice model or the SUR model. Indeed, if one considers the discrete choice model, note that $\widehat{\gamma} = 0.818 < 1$ and $|x_o - x_m|^{\widehat{\gamma}} = |-0.518 - 0.478|^{0.818} = 0.997 > \widehat{\theta} = 0.269$. So Property (2) in Proposition 1 holds, which means that the single-crossing property is not satisfied. With the SUR model, $\widehat{\gamma} = 0.371 < 1$ and $|x_o - x_m|^{\widehat{\gamma}} = |-0.518 - 0.478|^{0.371} = 0.998 > \widehat{\theta} = 0.227$. So, again, the single-crossing property does not hold. To take into consideration sampling variation and test the null that the single-crossing holds (versus the alternative that it does not), the last line in Table 2 provides the achieved significance levels of the test. In the case of the discrete choice model, out of $B = 999$ bootstrap samples, the single-crossing property is satisfied as many as 37 times, giving a achieved significance level of 0.037. So the single-crossing property is rejected at the 5 percent significance level when the discrete choice model is considered. In the case of the SUR model, out of $B = 999$ bootstrap samples,

Table 2: Discrete choice and SUR estimates for the 2008 Presidential election

	(A) Discrete choice model (Stated choice)	(B) SUR model (Thermometer scores)
Obama's valence-advantage		
$\widehat{\theta}$	0.269	0.227
	(0.033)	(0.037)
	$[0.206, 0.338]^{ii.}$	[0.168, 0.286]
Exponent parameter		
$\widehat{\gamma}$	0.818	0.371
	(0.097)	(0.055)
	$(-\infty, 0.988]^{iii.}$	$(-\infty, 0.458]$
Scale parameters		
$\widehat{\sigma}$	0.358	
	(0.031)	
$\widehat{c}$		86.615
		(4.928)
$\widehat{\beta}$		42.655
		(5.070)
N	1320	1449
Log-likelihood	-540.161	-13274.940
Bootstrap single-crossing test$^{iv.}$		
$\widehat{ASL}$	0.037	0.000

Notes: i. Standard errors are in parentheses.
ii. The intervals in brackets below the estimated standard errors of $\widehat{\theta}$ correspond to the percentile 95 percent confidence intervals for the parameters. These intervals are equal-tailed: they are the distance between the lower 0.025 and upper 0.025 quantiles of $B = 999$ bootstrap estimates of the parameter of interest. For instance, the interval [0.206,0.338] below the estimated standard error of $\widehat{\theta}$ in the case of the discrete choice model corresponds to the percentile 95 percent confidence interval $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$ for θ, where $\widehat{\theta}^*_{0.025} = 0.206$ is the lower 0.025 and $\widehat{\theta}^*_{0.975} = 0.338$ the upper 0.025 quantiles of the $B = 999$ bootstrap estimates $\widehat{\theta}^*_b$, $b = 1, \ldots, B$.
iii. The intervals below the estimated standard errors of $\widehat{\gamma}$ correspond to the percentile 95 percent confidence intervals based on a lower one-tailed alternative test. For instance, the interval $(-\infty, 0.988]$ below the estimated standard error of $\widehat{\gamma}$ in the case of the discrete choice model corresponds to the percentile 95 percent confidence interval $(-\infty, \widehat{\gamma}^*_{0.95}]$ for γ, where $\widehat{\gamma}^*_{0.95} = 0.988$ is the upper 0.05 quantiles of the $B = 999$ bootstrap estimates $\widehat{\gamma}^*_b$, $b = 1, \ldots, B$.
iv. Bootstrap single-crossing test ($\widehat{ASL}$) provides the achieved significance level of the test H_0 : "the single-crossing property holds" versus H_A : "the single-crossing property does not hold". Given that there are $B = 999$ bootstrap samples, there are B sets of parameter estimates $[x^*_{ob}, x^*_{mb}, \widehat{\theta}^*_b, \widehat{\gamma}^*_b]$, $b = 1, \ldots, B$. I check for each set of parameter estimates if the single-crossing property holds. The proportion of bootstrap samples for which the single-crossing property holds is the estimate of the achieved significance level:

$$\widehat{ASL} = 1 - \frac{\sharp\{b = 1, \ldots, B \ ; \ [(|x^*_{ob} - x^*_{mb}|^{\widehat{\gamma}^*_b} > \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b < 1) \text{ or } (|x^*_{ob} - x^*_{mb}|^{\widehat{\gamma}^*_b} = \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b = 1)]\}}{B}$$

The method is discussed in Appendix D.

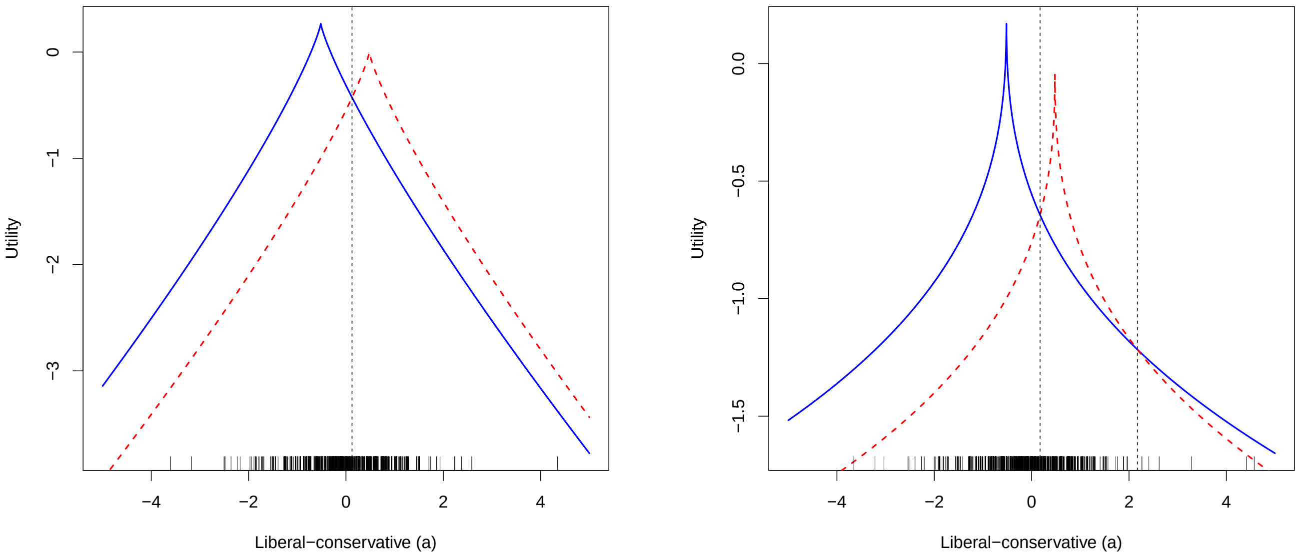
the single-crossing is never observed, giving a achieved significance level of zero. So the single-crossing property is also rejected when the SUR model is considered.

The rejection of the single-crossing property with both the discrete choice model and the SUR model suggests consistent results. However, there is a problem with the discrete choice model: if we consider the locations of Obama and McCain obtained with the Discrete choice sample, and the estimated parameters $\widehat{\theta}$ and $\widehat{\gamma}$ in Column (A) of Table 2, the equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has two solutions in a which are $a^* \simeq 0.124$ and $a^{**} \simeq 442.458$; a^{**} belongs to the theoretical policy space $\mathbb{R}$, but it is outside the observed data-range [-3.600, 4.347] reported in Table 1. In other words, the single-crossing property is rejected with the discrete choice model, but, in the data-range, the equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has only one solution in a, $a^* \simeq 0.124$, as described in Panel (A) of Figure 3.

This issue does not occur with the SUR model. When we consider the locations of Obama and McCain obtained with the SUR sample, and the estimated parameters in Column (B) of Table 2, $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has two solutions in a, $a^* \simeq 0.172$ and $a^{**} \simeq 2.173$, and these two solutions are in the data-range [-3.652, 4.572] reported in Table 1. So according to the estimated SUR model depicted in Panel (B) of Figure 3, voters whose bliss points are to the right of 2.173 are closer to McCain, but their estimated utility is higher for Obama. A reader may naturally ask if these respondents do rank Obama above McCain by providing a higher thermometer score to Obama. 7 respondents have their bliss points to the right of 2.173. Out of these 7 respondents, 3 rate Obama higher than McCain, 3 rate them equally, and only 1 rates McCain higher.

This difference between the discrete choice and the SUR models is due to the estimated exponent parameter $\widehat{\gamma}$. Indeed, the estimated valence-advantage parameter $\widehat{\theta}$ is broadly similar in the two models, as shown in Table 2. The estimated locations of Obama and

McCain are also comparable in both the Discrete choice and SUR samples, as shown in Table 1. Concerning $\widehat{\gamma}$, if this estimate is significantly less than one in both models, it is notably lower in the SUR model.



(A) Discrete choice model

(B) SUR model

Figure 3: Estimated utility functions (Discrete choice and SUR samples)

Note: This figure depicts the estimated utilities in function of a obtained in Table 2. In Panel (A), the estimated coefficients $\widehat{\theta}$ and $\widehat{\gamma}$ of the discrete choice model are used, while in Panel (B), these are those of the SUR model. The (blue) solid curve depicts the estimated utility if Obama is elected. The (red) dashed curve depicts the estimated utility if McCain is elected.

6 Additional results

This section includes several robustness checks. It also examines the results in the context of other US Presidential elections besides the 2008 election.

6.1 Expanding the size of the Discrete choice and SUR samples

As previously noted in Section 4, missing observations in the Discrete Choice and SUR samples primarily result from respondents failing to answer the self-placement question. Respondents who answered they “Haven’t thought much about it” or “Don’t know” were prompted with a follow-up question, with three possible responses: “Liberal”, “Conservative” and “Moderate”. This subsection leverages this follow-up question to reduce the number of missing observations. Recall that this follow-up question was also posed to those who selected “Moderate/middle of the road” in response to the self-placement question (i.e., $\widetilde{a}_i = 4$). I briefly describe the specific assumptions used in the coding and the results.

I assume that those who initially answered “Moderate/middle of the road” to the self-placement question (i.e., $\widetilde{a}_i = 4$) are moderate, even if they responded that they were “Liberal” or “Conservative” to the follow-up question. Concerning those who responded initially they “Haven’t thought much about it” or “Don’t know” to the self-placement question, I assume that they are moderate (i.e., $\widetilde{a}_i = 4$) if they answered “Moderate” to the follow-up question. I assume that they are slightly liberal (i.e., $\widetilde{a}_i = 3$) if they answered “Liberal” to the follow-up question, and slightly conservative (i.e., $\widetilde{a}_i = 5$) if they answered “Conservative” to the follow-up question. The reason of this choice is that if someone responded initially she hasn’t thought much about it or doesn’t know, it suggests that she had doubts about her liberal-conservative inclination. So, if she answered “Liberal” or “Conservative” to the follow-up question, it indicates only a slight inclination.

I proceed as I did in Section 4 to build the expanded Discrete choice and SUR samples. For both samples, I initially remove any remaining missing observations, which could occur if, e.g., the reported location of one of the stimuli ($\widetilde{x}_{i,o}$, $\widetilde{x}_{i,m}$, $\widetilde{x}_{i,dem}$, or $\widetilde{x}_{i,rep}$) is unanswered. I then obtain via the Aldrich-McKelvey procedure the true locations of Obama and McCain which are very similar to those obtained in Table 1: these locations are $x_o \simeq -0.520$ and $x_m \simeq 0.482$ in the expanded Discrete choice sample, and $x_o \simeq -0.520$ and $x_m \simeq 0.483$ in the expanded SUR sample. Finally, I obtain the respondents’ bliss points a_i, $i = 1, \ldots, N$, in the same dimension as the actual locations of the candidates.

Note that these bliss points range in the interval [-3.729, 4.495] in the expanded Discrete choice sample, and [-3.786, 4.753] in the expanded SUR sample. As for the number of observations, the expanded Discrete choice sample consists of 1721 respondents, and the expanded SUR sample has 1895 respondents. Hence the effective response rates are about 74% $\left(\simeq \frac{1721}{2322}\right)$ and 82% $\left(\simeq \frac{1895}{2322}\right)$, against 57 and 63% in Sections 4 and 5.

Table E1 in Appendix E provides the results. Once again, I find that Obama is the candidate with a valence-advantage. I obtain $\widehat{\theta} = 0.386$ with the discrete choice model(versus 0.269 in Table 2), and $\widehat{\theta} = 0.337$ with the SUR model (versus 0.227 in Table 2). The null hypothesis $\mathrm{H}_0: \ \theta = 0$ is rejected whether we treat the distances between a_i and x_j as observed variables (i.e., ignoring any estimation error in these variables) or if we consider the 95 percent confidence interval based on bootstrap percentiles.

As for $\widehat{\gamma}$, the results closely resemble those obtained in Table 2: $\widehat{\gamma} = 0.846$ with the discrete choice model (versus 0.818 in Table 2), and $\widehat{\gamma} = 0.333$ with the SUR model (versus 0.371 in Table 2). All the one-sided tests for $\mathrm{H}_0: \ \gamma \geq 1$ versus $\mathrm{H}_A: \ \gamma < 1$ reject the null hypothesis.

Finally, the single-crossing property is rejected at the 5 percent significance level in both the discrete choice and the SUR models. Similar to Section 5, the equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ based on the discrete choice estimates has two solutions in a, $a^* \simeq 0.185$ and $a^{**} \simeq 166.164$, and a^{**} falls outside the data-range [-3.729, 4.495]. This issue, again, does not occur with the SUR model: $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has two solutions for a, $a^* \simeq 0.277$ and $a^{**} \simeq 1.040$, and both of them fall within the data-range [-3.786, 4.753]. So voters whose bliss points are to the right of 1.040 are closer to McCain, but their estimated utility is higher for Obama. A reader may naturally ask if these respondents do rank Obama above McCain by providing a higher thermometer score to Obama. 105 respondents have their bliss points to the right of 1.040. Out of these 105 respondents, 21 rate Obama higher than McCain and 15 rate them equally.

6.2 Constructing the policy space using the reported positions of only two stimuli

In Section 4, I have contended that it is crucial to consider more than just two stimuli when constructing the policy space. Nevertheless, by focusing on respondents' reported positions for just two stimuli, namely Obama and McCain, the Aldrich-McKelvey technical assumption (i) which is that the sum of the true locations is zero and the sum of squares equal to one fully determines the locations of the stimuli if it is combined with an extra assumption (ii) that the researcher knows that McCain is located to the right of Obama, i.e., $x_m > x_o$. Under these two assumptions, the obtained locations are $x_o = -\sqrt{\frac{1}{2}}$ and $x_m = \sqrt{\frac{1}{2}}$; see Remark C2 in Appendix C for more details. Some readers may be curious about the results when using only these two stimuli to build the policy space. This subsection does so, providing a brief description of the specific issues in the coding and the results.

First, I think it prudent to emphasize that if one constructs the policy space using only two stimuli, the basicspace package in the R environment cannot be used because the Aldrich-McKelvey procedure per se is not applicable. Indeed, the additional assumption (ii) $x_m > x_o$ is not part of the Aldrich-McKelvey assumptions. While it is a priori credible to assume that the researcher knows McCain's position is to the right of Obama, when working with only two stimuli to define the policy space, a customized (simple) coding procedure becomes necessary. I have first to assign the values of $-\sqrt{\frac{1}{2}}$ and $\sqrt{\frac{1}{2}}$ to x_o and x_m. With the reported positions $\widetilde{x}_{i,o}$ and $\widetilde{x}_{i,m}$ of the stimuli and their location x_o and x_m, I then estimate the respondent-specific distortion parameters c_i and w_i using OLS, following the Aldrich-McKelvey method. Finally, I use these distortion parameters and each respondent i's self-placement $\widetilde{a}_i$ to obtain the bliss points a_i, $i = 1, \ldots, N$, in the common space.

As mentioned in Section 4, considering only two stimuli increases the number of instances where there is no variability in the reported positions of these stimuli (i.e., $\widetilde{x}_{i,o} = \widetilde{x}_{i,m}$). Specifically, 89 respondents are affected when constructing the Discrete Choice sample (versus 26 in Section 4), and 109 when creating the SUR sample (versus 34 in Section 4). Consequently, it is not surprising that both the Discrete choice and SUR samples consist of fewer respondents than in Sections 4 and 5 (1278 versus 1320 and 1398 versus 1449, respectively).

Table E2 in Appendix E provides the results. Observe that the bootstrap method is still attractive for testing the single-crossing hypothesis given its analytical complexity. But if the number of stimuli to build the policy space is two, the reported positions of these stimuli by the respondents become irrelevant to determine their location in a first step: whatever the sample, $x_o = -\sqrt{\frac{1}{2}}$ and $x_m = \sqrt{\frac{1}{2}}$. Hence there is no randomness in x_o and x_m, nor in a_i, $i = 1, \ldots, N$: it is as if these variables were observed variables, and there is no two-step estimation problem here. So the bootstrap procedure is only applied to the second-stage regressions. Given that, the percentile confidence intervals for θ and γ are not necessary. Nevertheless, Table E2 includes these percentile confidence intervals to have a homogeneous presentation of the different tables of results. Besides, they are provided because they could have lead to different conclusions compared to the simple t-tests. However, this is not the case, and the results are broadly similar to those of Tables 2 and E1. First, the estimated valence advantage for Obama always differs significantly from zero. Second, $\widehat{\gamma}$ is always significantly less than one. Ultimately, the single-crossing property is also rejected in both the discrete choice and SUR models. Similar to Sections 5 and 6.1, the equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ based on the discrete choice estimates has two solutions in a, a^* and a^{**}, but a^{**} falls outside the data-range. This issue, again, does not occur with the SUR model: $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has two solutions for

a, and both of them fall within the data-range; more details are provided at the end of Appendix E.

6.3 Other waves of the ANES

In Section 5, as well as in Subsections 6.1 and 6.2, I have shown that adding a valence-advantage parameter to Obama over McCain in the spatial model is supported by both the discrete choice and SUR models. Additionally, both econometric models provide results which are consistent with a γ which is less than 1. The various tests also reject the single-crossing hypothesis in both models. Thus, the results show qualitative similarity across both models. However, it is worth noting that if the equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has two solutions in a in both models, the "ends against the middle" split of the voters is primarily confirmed by the SUR estimates. Specifically, if the equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ based on the discrete choice estimates has two solutions in a, and these two solutions belong to the theoretical policy space $\mathbb{R}$, one of this solution is always outside the data-range. This discrepancy does not occur with the SUR estimates where the two solutions in a always belong to the data-range.

Nevertheless, one should have in mind that these results are only true for the 2008 US Presidential election. Nothing insures that the data for other US Presidential elections are consistent with a γ which is less than 1, nor that the single-crossing property is rejected. I have also investigated the 2012, 2016 and 2020 US Presidential elections, and I document the results below. Before to step any further, note that I proceed like in Section 4 to build the different estimation samples. In particular, I always consider four stimuli to build the policy space of each election (the Democratic and the Republican candidates, as well as their respective party).

Table 3 provides the discrete choice estimates. The first result concerns the estimated exponent parameter $\widehat{\gamma}$. It is higher than one in the 2016 Presidential election (Column

(B)). In the other elections (Columns (A) and (C)), $\widehat{\gamma}$ is less than one, but the null $H_0 : \gamma \geq 1$ is never rejected at the 5 percent significance level. So the data in these different waves of the ANES are consistent with the hypothesis that γ is higher than one. If so, given Property (3) in Corollary 1, the single-crossing property is always satisfied. In line with this finding, note that the achieved significance level of the bootstrap single-crossing test is always larger than 0.05. So we conclude that the Discrete choice samples in waves 2012, 2016 and 2020 are consistent with the single-crossing hypothesis.

Observe, however, that the 2012 Presidential election is a borderline case in Table 3: we have failed to reject the single-crossing hypothesis at the 5 percent significance level, but we would have rejected this hypothesis at the 10 percent significance level given that the achieved significance level is 0.051. Similarly, the percentile 95 percent confidence interval based on a lower tailed test for γ is $(-\infty, 1.003]$ when the Discrete choice sample is considered, but the percentile 90 percent confidence interval based on a lower tailed test is $(-\infty, 0.969]$. So $\widehat{\gamma}$ is not significantly less than one at level 0.05, but it is at level 0.10.

Concerning the SUR estimates, the results are reported in Table 4. As for the 2008 Presidential election, the estimates $\widehat{\gamma}$ with the SUR models are much lower than those found for the discrete choice models. These estimates are between 0.301 and 0.338, depending on the election, and they are always significantly less than one. Finally, the single-crossing property is also rejected in the different SUR models. These results seem at odds with those of the different discrete choice models presented in Table 3. Nonetheless, consider the equation based on the different SUR estimates $\widehat{u}(a, x_j, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_k, \widehat{\gamma})$, where x_j is the location of the valence-advantaged candidate j and x_k the location of the opponent in a given election. This equation has always two solutions in a, but, again, one of this solution may be outside the observed data-range. This is the case of the 2016 Presidential election: the data-range for a is [-4.140, 4.140], but the two solutions are

Table 3: Discrete choice estimates for the 2012, 2016 and 2020 Presidential elections

	(A) 2012 Obama vs Romney	(B) 2016 Clinton vs Trump	(C) 2020 Biden vs Trump
Democrat's valence advantage			
$\widehat{\theta}$	0.200	0.151	0.098
	(0.015)	(0.021)	(0.012)
	$[0.169, 0.232]^{ii.}$	[0.107, 0.194]	[0.0733, 0.122]
Exponent parameter			
$\widehat{\gamma}$	0.881	1.045	0.953
	(0.062)	(0.081)	(0.063)
	$(-\infty, 1.003]^{iii.}$	$(-\infty, 1.228]$	$(-\infty, 1.089]$
Scale parameter			
$\widehat{\sigma}$	0.249	0.274	0.253
	(0.011)	(0.014)	(0.009)
N	3836	2295	5733
Log-likelihood	-1069.398	-610.061	-1456.822
Bootstrap single-crossing test$^{iv.}$			
$\widehat{ASL}$	0.051	0.736	0.235

Notes: i. Standard errors are in parentheses.
ii. The intervals in brackets below the estimated standard errors of $\widehat{\theta}$ correspond to the percentile 95 percent confidence intervals for the parameter. These intervals are equal-tailed: they are the distance between the lower 0.025 and upper 0.025 quantiles of $B = 999$ bootstrap estimates of the parameter of interest. For instance, the interval [0.169,0.232] below the estimated standard error of $\widehat{\theta}$ in Column (A) corresponds to the percentile 95 percent confidence interval $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$ for θ, where $\widehat{\theta}^*_{0.025} = 0.169$ is the lower 0.025 and $\widehat{\theta}^*_{0.975} = 0.232$ the upper 0.025 quantiles of the $B = 999$ bootstrap estimates $\widehat{\theta}^*(b)$, $b = 1, \ldots, B$.
iii. The intervals below the estimated standard errors of $\widehat{\gamma}$ correspond to the percentile 95 percent confidence intervals based on a lower one-tailed alternative test. For instance, the interval $(-\infty, 1.003]$ below the estimated standard error of $\widehat{\gamma}$ in Column (A) corresponds to the percentile 95 percent confidence interval $(-\infty, \widehat{\gamma}^*_{0.95}]$ for γ, where $\widehat{\gamma}^*_{0.95} = 1.003$ is the upper 0.05 quantiles of the $B = 999$ bootstrap estimates $\widehat{\gamma}^*(b)$, $b = 1, \ldots, B$.
iv. Bootstrap single-crossing test ($\widehat{ASL}$) provides the achieved significance level of the test H_0 : "the single-crossing property holds" versus H_A : "the single-crossing property does not hold". For instance, in the 2012 Presidential election, for each bootstrap sample, I obtain first the actual location of Obama (x^*_o) and Romney (x^*_r) and the bliss point a^*_i of each respondent i, and then I estimate the conditional logit to obtain $\widehat{\theta}^*$ and $\widehat{\gamma}^*$. Given that there are $B = 999$ bootstrap samples, there are B sets of parameter estimates $[x^*_{ob}, x^*_{rb}, \widehat{\theta}^*_b, \widehat{\gamma}^*_b]$, $b = 1, \ldots, B$. I check for each set of parameter estimates if the single-crossing property holds. The proportion of bootstrap samples for which the single-crossing property holds is the estimate of the achieved significance level:

$$\widehat{ASL} = 1 - \frac{\sharp\{b = 1, \ldots, B \ ; \ [(|x^*_{ob} - x^*_{rb}|^{\widehat{\gamma}^*_b} > \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b < 1) \text{ or } (|x^*_{ob} - x^*_{rb}|^{\widehat{\gamma}^*_b} = \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b = 1)]\}}{B}$$

The method is discussed in Appendix D.

$a^* \simeq 0.058$ and $a^{**} \simeq 8.124$; so a^{**} falls outside the data-range. In the case of the 2020 Presidential election, the data-range for a is [-6.214, 4.434], and the two solutions in a are $a^* \simeq 0.136$ and $a^{**} \simeq 3.617$. If this result seems to confirm the "ends against the middle" split of the voters, it does not. Indeed, there is only one respondent who has her bliss point $a_i = 4.434$ greater than 3.617; this is negligible. Concerning the 2012 Presidential election, the data-range is [-5.235, 6.398], and the two solutions in a are $a^* \simeq 0.146$ and $a^{**} \simeq 2.514$. Voters whose bliss points are to the right of 2.514 are closer to Romney, but their estimated utility is higher for Obama. 13 respondents have their bliss point to the right of 2.514. Out of these 13 respondents, 3 provide a higher thermometer score to Obama, and 3 rate Obama and Romney equally.

Thus, in contrast to the other SUR estimates of Table 4 where the single-crossing hypothesis is rejected but broadly satisfied in the data-range, the single-crossing hypothesis is rejected and is not satisfied in the data-range in 2012. It raises the question: what is so special about the 2012 (and 2008) Presidential election? The short answer is: Obama. His estimated valence-advantage over Romney in 2012 (0.159) is higher than the one of Clinton over Trump in 2016 (0.080). As stated by Property (3) in Corollary 1, if θ is excessively high, i.e., $\theta > |x_j - x_k|^\gamma$, the single-crossing property is satisfied (even if $\gamma < 1$). Nevertheless, suppose that θ is relatively high, but the condition for the single-crossing property not being satisfied holds, i.e., $|x_j - x_k|^\gamma > \theta$ (and $\gamma < 1$), as stated by Property (2) of Proposition 1, which is the case in the 2012 Presidential election. If so, given that $|x_j - x_k|^\gamma > \theta$ (and $\gamma < 1$), the single-crossing property is not satisfied in the theoretical policy space $\mathbb{R}$. Moreover, as I will explain below, given that θ is relatively high, it is also not satisfied in the data-range. In contrast, if $|x_j - x_k|^\gamma > \theta$ (and $\gamma < 1$), but θ is too low, as is the case in the 2016 Presidential election, the single-crossing property is not satisfied in the theoretical policy space *but* it is satisfied in the data-range. As depicted in Figure F1

Table 4: SUR estimates for the 2012, 2016 and 2020 Presidential elections

	(A) 2012 Obama vs Romney	(B) 2016 Clinton vs Trump	(C) 2020 Biden vs Trump
Democrat's valence advantage			
$\widehat{\theta}$	0.159	0.080	0.115
	(0.014)	(0.017)	(0.010)
	$[0.136, 0.187]^{ii.}$	[0.049,0.113]	[0.094,0.140]
Exponent parameter			
$\widehat{\gamma}$	0.305	0.338	0.301
	(0.023)	(0.033)	(0.018)
	$(-\infty, 0.343]^{iii.}$	$(-\infty, 0.381]$	$(-\infty, 0.337]$
Scale parameter			
$\widehat{c}$	99.303	82.795	95.652
	(3.909)	(4.369)	(3.222)
$\widehat{\beta}$	65.352	56.273	67.948
	(3.981)	(4.506)	(3.301)
N	4938	3125	6540
Log-likelihood	-45766.23	-29479.11	-60960.36
Bootstrap single-crossing test[iv.]			
$\widehat{ASL}$	0	0	0

Notes: i. Standard errors are in parentheses.
ii. The intervals in brackets below the estimated standard errors of $\widehat{\theta}$ correspond to the percentile 95 percent confidence intervals for the parameter. These intervals are equal-tailed: they are the distance between the lower 0.025 and upper 0.025 quantiles of $B = 999$ bootstrap estimates of the parameter of interest. For instance, the interval [0.136, 0.187] below the estimated standard error of $\widehat{\theta}$ in Column (A) corresponds to the percentile 95 percent confidence interval $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$ for θ, where $\widehat{\theta}^*_{0.025} = 0.136$ is the lower 0.025 and $\widehat{\theta}^*_{0.975} = 0.187$ the upper 0.025 quantiles of the $B = 999$ bootstrap estimates $\widehat{\theta}^*(b)$, $b = 1, \ldots, B$.
iii. The intervals below the estimated standard errors of $\widehat{\gamma}$ correspond to the percentile 95 percent confidence intervals based on a lower one-tailed alternative test. For instance, the interval $(-\infty, 0.343]$ below the estimated standard error of $\widehat{\gamma}$ in Column (A) corresponds to the percentile 95 percent confidence interval $(-\infty, \widehat{\gamma}^*_{0.95}]$ for γ, where $\widehat{\gamma}^*_{0.95} = 0.343$ is the upper 0.05 quantiles of the $B = 999$ bootstrap estimates $\widehat{\gamma}^*(b)$, $b = 1, \ldots, B$.
iv. Bootstrap single-crossing test ($\widehat{ASL}$) provides the achieved significance level of the test H_0 : "the single-crossing property holds" versus H_A : "the single-crossing property does not hold". For instance, in the 2012 Presidential election, for each bootstrap sample, I obtain first the actual location of Obama (x^*_o) and Romney (x^*_r) and the bliss point a^*_i of each respondent i, and then I estimate the SUR to obtain $\widehat{\theta}^*$ and $\widehat{\gamma}^*$. Given that there are $B = 999$ bootstrap samples, there are B sets of parameter estimates $[x^*_{ob}, x^*_{rb}, \widehat{\theta}^*_b, \widehat{\gamma}^*_b]$, $b = 1, \ldots, B$. I check for each set of parameter estimates if the single-crossing property holds. The proportion of bootstrap samples for which the single-crossing property holds is the estimate of the achieved significance level:

$$\widehat{ASL} = 1 - \frac{\sharp\{b = 1, \ldots, B \;;\; [(|x^*_{ob} - x^*_{rb}|^{\widehat{\gamma}^*_b} > \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b < 1) \text{ or } (|x^*_{ob} - x^*_{rb}|^{\widehat{\gamma}^*_b} = \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b = 1)]\}}{B}$$

The method is discussed in Appendix D.

of Appendix F, this phenomenon occurs because, given that the single-crossing property is not satisfied in the theoretical policy space, the length of the interval (a^*, a^{**}) diminishes as the valence-advantage parameter θ increases. More precisely, the lower bound a^* increases while the upper bound a^{**} decreases. Thus, in the 2016 Presidential election where the estimated valence-advantage $\widehat{\theta}$ of Clinton over Trump is low, the length of the interval (a^*, a^{**}) is large and the upper bound a^{**} falls outside the observed data-range. And in the 2012 Presidential election where the estimated valence-advantage $\widehat{\theta}$ of Obama over Romney is relatively high, the length of the interval (a^*, a^{**}) is smaller, so a^* and a^{**} fall within the observed data-range.

7 Conclusion

In order to add realism into the spatial model of voting, various authors have added a valence parameter into the Downsian utility function. In this framework, the value that the exponent on the distance should take is almost never discussed. Typically, many theorists, but also empiricists, assume an exponent of 2 to obtain a mathematically more tractable model. The problem is that the single-crossing property is also implicitly assumed in valence models wherein the distance appears with an exponent of 2, as demonstrated in this paper. More generally, this paper first establishes a necessary and sufficient condition for the single-crossing property not being satisfied in an additive-valence model. A part of this condition highlights that the single-crossing may not be satisfied if the exponent is less than one.

It is not merely a theoretical peculiarity. We show, through two distinct econometric frameworks applied to the 2008 US Presidential election, that the estimated exponent parameter is significantly less than one. Furthermore, the single-crossing property is rejected in both frameworks. Nevertheless, in one framework, if the single-crossing property is not satisfied in the theoretical policy space, i.e., the real line, it is satisfied in the data-range. This discrepancy does not occur in the other framework wherein the single-crossing is not satisfied in the data-range, and the "ends against the middle" split of the voters is confirmed.

These results highlight that an exponent parameter which is less than one or the fact that the single-crossing property does not hold may occur. Nevertheless, it is essential to emphasize that these results are not common in more recent US Presidential elections, or, at least, are much more ambiguous. In the 2012, 2016, and 2020 US Presidential elections, one of our frameworks always provides results consistent with the hypothesis of an exponent parameter higher than one, which is a sufficient condition for the single-crossing property being satisfied. In line with these results, the single-crossing hypothesis is not rejected. In the other framework, the results are consistent with an exponent parameter which is less than one. However, an "ends against the middle" split of the voters is not found, except in the 2012 election.

References

ABRAMSON, P. R., ALDRICH, J. H., PAOLINO, P. and ROHDE, D. W. (1992), " "Sophisticated" voting in the 1988 presidential primaries", *American Political Science Review*, vol. 86: pp. 55–69.

ADAMS, J., MERRILL, S. and GROFMAN, B. (2005), *A Unified Theory of Party Competition*, Cambridge University Press.

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.

ALVAREZ, R. M. and NAGLER, J. (1995), "Economics, issues and the Perot candidacy: Voter choice in the 1992 Presidential election", *American Journal of Political Science*, vol. 39 n° 3: pp. 714–744.

ALVAREZ, R. M. and NAGLER, J. (2000), "A new approach for modelling strategic voting in multiparty elections", *British Journal of Political Science*, vol. 30: pp. 57–75.

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.

BLACK, J. H. (1978), "The multicandidate calculus of voting: Application to Canadian federal elections", *American Journal of Political Science*, vol. 22: pp. 609–638.

BLAIS, A. and NADEAU, R. (1996), "Measuring strategic voting: A two-step procedure", *Electoral Studies*, vol. 15: pp. 39–52.

CROISSANT, Y. (2013), "mlogit: Multinomial logit model", http://CRAN.R-project.org/package=mlogit/.

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. (2019), "Valence politics", in Congleton, R., Grofman, B. and Voigt, S. (editors), *The Oxford Handbook of Public Choice*, Oxford: Oxford University Press.

Fuente, A. d. l. (2000), *Mathematical Methods and Models for Economists*, Cambridge: Cambridge University Press.

Gouret, F. (2021), "Empirical foundation of valence using Aldrich-McKelvey scaling", *Review of Economic Design*, vol. 25: pp. 177–226.

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.

Heckman, J. J. and Snyder, J. M. (1997), "Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators", *RAND Journal of Economics*, vol. 28: pp. S142–S189.

Henry, M. and Mourifié, I. (2013), "Euclidian revealed preferences: Testing the spatial voting model", *Journal of Applied Econometrics*, vol. 28: pp. 650–666.

Hollard, G. and Rossignol, S. (2008), "An alternative approach to valence advantage in spatial competition", *Journal of Public Economic Theory*, vol. 10: pp. 441–454.

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.

King, G., Murray, C. J. L., Salomon, J. A. and Tandon, A. (2004), "Enhancing the validity and cross-cultural comparability of measurement in survey research", *American Political Science Review*, vol. 98: pp. 191–207.

Martin, M., Nganmeni, Z., Piggins, A. and Tchouante, E. F. (2022), "Pure-strategy Nash equilibrium in the spatial model with valence: existence and characterization", *Public Choice*, vol. 190: pp. 301–316.

McFadden, D. (1974), "Conditional logit analysis of qualitative choice behavior", in Zarembka, P. (editor), *Frontiers in Econometrics*, Academik Press, New York London, pp. 105–142.

Milgrom, P. and Shannon, C. (1994), "Monotone comparative statics", *Econometrica*, vol. 62: pp. 157–180.

Ortuño-Ortín, I. (1997), "A spatial model of political competition and proportional representation", *Social Choice and Welfare*, vol. 14: pp. 427–438.

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: pp. 1–21, http://www.jstatsoft.org/v69/i07/.

Poole, K. T. and Rosenthal, H. (1991), "Patterns of congressional voting", *American Journal of Political Science*, vol. 35: pp. 228–278.

Quah, J. K.-H. and Strulovici, B. (2012), "Aggregating the single crossing property", *Econometrica*, vol. 80: pp. 2333–2348.

Roemer, J. E. (1994), "A theory of policy differentiation in single issue electoral politics", *Social Choice and Welfare*, vol. 11: pp. 355–380.

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: pp. 484 – 496.

Tourangeau, R., Rips, L. J. and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge University Press.

Train, K. E. (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press, 2nd edition.

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 Proof of Proposition 1

(1) *We prove first* that ($|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1$) is a sufficient condition for the single-crossing property not being satisfied. A voter whose bliss point is a strictly prefers candidate 1 if $u(a, x_1, \theta_1, \gamma) - u(a, x_2, \theta_2, \gamma) > 0$. Let $f(a) = u(a, x_1, \theta_1, \gamma) - u(a, x_2, \theta_2, \gamma) = \theta - |x_1 - a|^\gamma + |x_2 - a|^\gamma$. Note that $a \mapsto u(a, x_1, \theta_1, \gamma)$ and $a \mapsto -u(a, x_2, \theta_2, \gamma)$ are continuous on $\mathbb{R}$. The sum of two continuous functions on $\mathbb{R}$ is also continuous on $\mathbb{R}$, so $a \mapsto f(a)$ is continuous on $\mathbb{R}$.

Now, recall that $x_1 > x_2$, and note the three preliminary results (A1), (A2) and (A3):

$$\text{If } a \in (-\infty, x_2], \text{ then } f(a) = \theta - (x_1 - a)^\gamma + (x_2 - a)^\gamma \tag{A1}$$

$$\text{If } a \in (x_2, x_1), \text{ then } f(a) = \theta - (x_1 - a)^\gamma + (a - x_2)^\gamma \tag{A2}$$

$$\text{If } a \in [x_1, +\infty), \text{ then } f(a) = \theta - (a - x_1)^\gamma + (a - x_2)^\gamma \tag{A3}$$

We study $f(a)$ on these three intervals, and show that $f(a) = 0$ has two solutions in a if $(|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1)$.

- If $a \in (-\infty, x_2]$, then, using Equation (A1), $f'(a) = \gamma(x_1 - a)^{\gamma-1} - \gamma(x_2 - a)^{\gamma-1}$. Given that $x_1 > x_2$, we have $x_1 - a > x_2 - a \Leftrightarrow \frac{1}{x_1-a} < \frac{1}{x_2-a}$. Given that $\gamma < 1$, we obtain $\left(\frac{1}{x_1-a}\right)^{1-\gamma} < \left(\frac{1}{x_2-a}\right)^{1-\gamma} \Leftrightarrow (x_1 - a)^{\gamma-1} < (x_2 - a)^{\gamma-1}$. Hence, $f'(a) < 0$, and $f(a)$ is strictly decreasing on $(-\infty, x_2]$. Thus, $f(a)$ has a minimum at $a = x_2$ and a maximum when $a \to -\infty$. If $a = x_2$, Equation (A1) becomes $f(x_2) = \theta - (x_1 - x_2)^\gamma$. Given that $(x_1 - x_2)^\gamma > \theta$, then $f(x_2) < 0$. Now, if $a \to -\infty$, the limit of Equation (A1) is indeterminate: $\lim_{a\to-\infty} \theta - (x_1 - a)^\gamma + (x_2 - a)^\gamma = \theta - \infty + \infty$. To evaluate this indeterminate form, we consider a linear Taylor series expansion. First, we rewrite Equation (A1) as:

$$f(a) = \theta - (-a)^\gamma \left[1 + \left(\frac{-x_1}{a}\right)\right]^\gamma + (-a)^\gamma \left[1 + \left(\frac{-x_2}{a}\right)\right]^\gamma \tag{A4}$$

  The linear Taylor series expansion of a function $g(z) = (1 + z)^\alpha$ around the point $z = 0$ is $g(z) = 1 + \alpha z + o(z)$, where o is the little-o notation. Now, note that $\lim_{a\to-\infty} \left(\frac{-x_1}{a}\right) = 0$ and $\lim_{a\to-\infty} \left(\frac{-x_2}{a}\right) = 0$. If so, Equation (A4) can be rewritten as:

$$f(a)=\theta-(-a)^{\gamma}\left[1+\gamma\left(\frac{-x_1}{a}\right)+o\left(\frac{-x_1}{a}\right)\right]+(-a)^{\gamma}\left[1+\gamma\left(\frac{-x_2}{a}\right)+o\left(\frac{-x_2}{a}\right)\right] \text{ as } a\rightarrow-\infty$$

Given that $o\left(\frac{-x_1}{a}\right)=o\left(\frac{1}{a}\right)$ and $o\left(\frac{-x_2}{a}\right)=o\left(\frac{1}{a}\right)$, and $o\left(\frac{1}{a}\right)+o\left(\frac{1}{a}\right)=o\left(\frac{1}{a}\right)$, we obtain:

$$f(a)=\theta+(-a)^{\gamma}\left[-1+\gamma\left(\frac{x_1}{a}\right)+1-\gamma\left(\frac{x_2}{a}\right)+o\left(\frac{1}{a}\right)\right] \text{ as } a\rightarrow-\infty \tag{A5}$$

Simplifying and rearranging Equation (A5), we have:

$$\begin{aligned}
f(a) &= \theta+(-a)^{\gamma}\left[\frac{\gamma}{a}(x_1-x_2)+o\left(\frac{1}{a}\right)\right] \text{ as } a\rightarrow-\infty \\
&= \theta+\frac{(-a)^{\gamma}}{a}\left[\gamma(x_1-x_2)+o\left(1\right)\right] \text{ as } a\rightarrow-\infty \\
&= \theta+\frac{1}{(-a)^{1-\gamma}}\left[\gamma(x_1-x_2)+o\left(1\right)\right] \text{ as } a\rightarrow-\infty
\end{aligned} \tag{A6}$$

Now, note that $\frac{1}{(-a)^{1-\gamma}}=o(1)$, so $f(a)=\theta+o(1)$, with $o(1)$ a function whose limit is zero when $a\rightarrow-\infty$. If so, we have shown that $\lim_{a\rightarrow-\infty}f(a)=\theta$.

To sum up, we have shown that $f(a)$ is strictly decreasing on $(-\infty,x_2]$ given that $\gamma<1$, with $\lim_{a\rightarrow-\infty}f(a)=\theta>0$ and $f(x_2)=\theta-(x_1-x_2)^{\gamma}<0$ given that $\theta<(x_1-x_2)^{\gamma}$. If so, there exists a unique $a^*\in(-\infty,x_2]$ such that $f(a^*)=0$.

- If $a\in(x_2,x_1)$, then, using Equation (A2), $f'(a)=\gamma(x_1-a)^{\gamma-1}+\gamma(a-x_2)^{\gamma-1}$. Given that $x_1>a>x_2$, we obtain $f'(a)>0$, so $f(a)$ is strictly increasing on (x_2,x_1). Thus, $f(a)$ has a minimum at $a=x_2$ and a maximum at $a=x_1$. If $a=x_2$, Equation (A2) becomes $f(x_2)=\theta-(x_1-x_2)^{\gamma}$. Given that $(x_1-x_2)^{\gamma}>\theta$, then $f(x_2)<0$. If $a=x_1$, Equation (A2) becomes $f(x_1)=\theta+(x_1-x_2)^{\gamma}$. Given that $x_1>x_2$, we obtain $f(x_1)>0$. Since $f(a)$ is strictly increasing on (x_2,x_1), with $f(x_2)<0$ and $f(x_1)>0$, there exists a unique $a^{**}\in(x_2,x_1)$ such that $f(a^{**})=0$.

- If $a \in [x_1, +\infty)$, then, using Equation (A3), $f'(a) = -\gamma(a - x_1)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > x_2$, we have $a - x_1 < a - x_2 \Leftrightarrow \frac{1}{a-x_1} > \frac{1}{a-x_2}$. Given that $\gamma < 1$, we obtain $\left(\frac{1}{a-x_1}\right)^{1-\gamma} > \left(\frac{1}{a-x_2}\right)^{1-\gamma} \Leftrightarrow (a - x_1)^{\gamma-1} > (a - x_2)^{\gamma-1}$. Hence, $f'(a) < 0$, so $f(a)$ is strictly decreasing on $[x_1, +\infty)$. Thus, $f(a)$ has a maximum at $a = x_1$ and a minimum when $a \to +\infty$. If $a = x_1$, Equation (A3) becomes $f(x_1) = \theta + (x_1 - x_2)^\gamma$. Given that $x_1 > x_2$, $f(x_1) > 0$. Now, if $a \to +\infty$, the limit of Equation (A3) is indeterminate: $\lim_{a\to+\infty} \theta - (a - x_1)^\gamma + (a - x_2)^\gamma = \theta - \infty + \infty$. To evaluate this indeterminate form, I consider a linear Taylor series expansion. First, Equation (A3) is rewritten as: $f(a) = \theta - a^\gamma \left[1 + \left(\frac{-x_1}{a}\right)\right]^\gamma + a^\gamma \left[1 + \left(\frac{-x_2}{a}\right)\right]^\gamma$. The linear Taylor series expansion of a function $g(z) = (1+z)^\alpha$ around the point $z = 0$ is $g(z) = 1 + \alpha z + o(z)$. Note that $\lim_{a\to+\infty} \left(\frac{-x_1}{a}\right) = 0$ and $\lim_{a\to+\infty} \left(\frac{-x_2}{a}\right) = 0$. If so, $f(a)$ can be rewritten as: $f(a) = \theta - a^\gamma \left[1 + \gamma\left(\frac{-x_1}{a}\right) + o\left(\frac{-x_1}{a}\right)\right] + a^\gamma \left[1 + \gamma\left(\frac{-x_2}{a}\right) + o\left(\frac{-x_2}{a}\right)\right]$ as $a \to +\infty$. Given that $o\left(\frac{-x_1}{a}\right) = o\left(\frac{1}{a}\right)$ and $o\left(\frac{-x_2}{a}\right) = o\left(\frac{1}{a}\right)$, and $o\left(\frac{1}{a}\right) + o\left(\frac{1}{a}\right) = o\left(\frac{1}{a}\right)$, we obtain:

$$f(a) = \theta + \frac{1}{a^{1-\gamma}} \left[\gamma\left(x_1 - x_2\right) + o\left(1\right)\right] \text{ as } a \to +\infty \tag{A7}$$

Given that $\frac{1}{(a)^{1-\gamma}} = o(1)$, we have $f(a) = \theta + o(1)$, with $o(1)$ a function whose limit is zero when $a \to +\infty$. If so, we have shown that $\lim_{a\to+\infty} f(a) = \theta$.

To sum up, we have shown that $f(a)$ is strictly decreasing on $[x_1, +\infty)$ given that $\gamma < 1$, with $\lim_{a\to+\infty} f(a) = \theta > 0$. If so, $f(a) > 0$ $\forall a \in [x_1, +\infty)$.

Conclusion: we have shown that ($|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1$) is a sufficient condition for $f(a) = 0$ having two solutions in a: there exists a unique $a^* \in (-\infty, x_2]$ such that $f(a^*) = 0$, and there exists a unique $a^{**} \in (x_2, x_1)$ such that $f(a^{**}) = 0$. Thus, ($|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1$) is a sufficient condition for the single-crossing property not being satisfied.

(2) *We prove now* that ($|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$) is also a sufficient condition for the single-crossing property not being satisfied. Consider the interval given by (A1).

- If $a \in (-\infty, x_2]$, then, using Equation (A1), $f'(a) = \gamma(x_1 - a)^{\gamma-1} - \gamma(x_2 - a)^{\gamma-1}$. Given that $\gamma = 1$, we obtain $f'(a) = 0$, so $f(a)$ is a constant for all $a \in (-\infty, x_2]$. It means that $\forall a \in (-\infty, x_2]$, $f(a) = f(x_2)$. Hence, Equation (A1) is $f(a) = \theta - (x_1 - x_2)^\gamma$. Given that $|x_1 - x_2|^\gamma = \theta$, we obtain $f(a) = 0$ $\forall a \in (-\infty, x_2]$, i.e., there are infinitely many $a \in (-\infty, x_2]$ such that $f(a) = 0$. This case is depicted by Figure 2 in the main text.

Conclusion: we have shown that ($|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$) is also a sufficient condition for $f(a) = 0$ having infinitely many solutions in $a \in (-\infty, x_2]$. Thus, ($|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$) is a sufficient condition for the single-crossing property not being satisfied.

(3) *It remains to prove* that [($|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1$) or ($|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$)] is a necessary condition for the single-crossing property not being satisfied, i.e., if the single-crossing property is not satisfied, then [($|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1$) or ($|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$)]. Denote $P =$ [($|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1$) or ($|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1$)]. The negation of this composite property, $\neg P$, is, using De Morgan's laws:

$$\begin{aligned} \neg P &= \neg[(|x_1 - x_2|^\gamma > \theta \text{ and } \gamma < 1) \text{ or } (|x_1 - x_2|^\gamma = \theta \text{ and } \gamma = 1)] \\ &= [\neg(|x_1 - x_2|^\gamma > \theta \text{ and } \gamma < 1) \text{ and } \neg(|x_1 - x_2|^\gamma = \theta \text{ and } \gamma = 1)] \\ &= [(|x_1 - x_2|^\gamma \leq \theta \text{ or } \gamma \geq 1) \text{ and } (|x_1 - x_2|^\gamma \neq \theta \text{ or } \gamma \neq 1)] \end{aligned} \tag{A8}$$

Rearranging this property by using the distributive law, we obtain:

$$\begin{aligned} \neg P &= [(|x_1 - x_2|^\gamma \leq \theta \text{ or } \gamma \geq 1) \text{ and } |x_1 - x_2|^\gamma \neq \theta] \text{ or } [(|x_1 - x_2|^\gamma \leq \theta \text{ or } \gamma \geq 1) \text{ and } \gamma \neq 1] \\ &= [(|x_1 - x_2|^\gamma \leq \theta \text{ and } |x_1 - x_2|^\gamma \neq \theta) \text{ or } (\gamma \geq 1 \text{ and } |x_1 - x_2|^\gamma \neq \theta)] \\ &\text{or } [(|x_1 - x_2|^\gamma \leq \theta \text{ and } \gamma \neq 1) \text{ or } (\gamma \geq 1 \text{ and } \gamma \neq 1)] \end{aligned}$$

Note that ($\gamma \geq 1$ and $\gamma \neq 1$) = ($\gamma > 1$), as well as ($|x_1 - x_2|^\gamma \leq \theta$ and $|x_1 - x_2|^\gamma \neq \theta$) = ($|x_1 - x_2|^\gamma < \theta$). So simplifying and rearranging $\neg P$, we obtain:

$$
\begin{aligned}
\neg P \quad &= \quad (\gamma > 1) \text{ or } (|x_1 - x_2|^\gamma < \theta) \text{ or } (\gamma \neq 1 \text{ and } |x_1 - x_2|^\gamma \leq \theta) \\
&\text{or} \quad (\gamma \geq 1 \text{ and } |x_1 - x_2|^\gamma \neq \theta)
\end{aligned} \tag{A9}
$$

Now, note that a statement and its contrapositive are equivalent. So proving the statement "the single-crossing property is not satisfied implies P" is equivalent to prove its contrapositive "$\neg P$ implies the single-crossing property". We proceed in four steps. We first show that ($\gamma > 1$) implies the single-crossing property. Second, we show that ($|x_1 - x_2|^\gamma < \theta$) implies the single-crossing property. Third, we show that ($\gamma \neq 1$ and $|x_1 - x_2|^\gamma \leq \theta$) implies the single-crossing property. Fourth, we show that ($\gamma \geq 1$ and $|x_1 - x_2|^\gamma \neq \theta$) implies the single-crossing property.

First step. We first show that ($\gamma > 1$) implies the single-crossing property.

- If $a \in (-\infty, x_2]$, then, using Equation (A1), $f'(a) = \gamma(x_1 - a)^{\gamma-1} - \gamma(x_2 - a)^{\gamma-1}$. Given that $x_1 > x_2$ and $\gamma > 1$, we obtain that $f'(a) > 0$. So $f(a)$ is strictly increasing on $(-\infty, x_2]$ if $\gamma > 1$.

- If $a \in (x_2, x_1)$, then, using Equation (A2), $f'(a) = \gamma(x_1 - a)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > a > x_2$, we have $f'(a) > 0$, so $f(a)$ is strictly increasing on (x_2, x_1).

- If $a \in [x_1, +\infty)$, then, using Equation (A3), $f'(a) = -\gamma(a - x_1)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > x_2$ and $\gamma > 1$, we have $a - x_1 < a - x_2 \Leftrightarrow (a - x_1)^{\gamma-1} < (a - x_2)^{\gamma-1}$. Hence, $f'(a) > 0$, so $f(a)$ is strictly increasing on $[x_1, +\infty)$.

  To sum up, we have shown that $f(a)$ is strictly increasing on $\mathbb{R}$ if $\gamma > 1$. Recall that $f(a)$ is continuous on $\mathbb{R}$. Therefore, there is at most one $a^* \in \mathbb{R}$ such that $f(a^*) = 0$ if $\gamma > 1$. Thus, $\gamma > 1$ implies the single-crossing property.

Second step. We now show that $|x_1 - x_2|^\gamma < \theta$ implies the single-crossing property. we have shown in the First step that $\gamma > 1$ implies the single-crossing property. Thus, $(|x_1 - x_2|^\gamma < \theta$ and $\gamma > 1)$ implies the single-crossing property. If so, we only need to show that $(|x_1 - x_2|^\gamma < \theta$ and $\gamma \leq 1)$ implies the single-crossing property.

- If $a \in (-\infty, x_2]$, then, using Equation (A1), $f'(a) = \gamma(x_1 - a)^{\gamma-1} - \gamma(x_2 - a)^{\gamma-1}$. Given that $x_1 > x_2$, we have $x_1 - a > x_2 - a \Leftrightarrow \frac{1}{x_1-a} < \frac{1}{x_2-a}$. Given that $\gamma \leq 1$, we obtain $\left(\frac{1}{x_1-a}\right)^{1-\gamma} \leq \left(\frac{1}{x_2-a}\right)^{1-\gamma} \Leftrightarrow (x_1 - a)^{\gamma-1} \leq (x_2 - a)^{\gamma-1}$. Hence, $f'(a) \leq 0$, and $f(a)$ is weakly decreasing on $(-\infty, x_2]$. Thus, $f(a)$ has a minimum at $a = x_2$ on the interval $(-\infty, x_2]$. If $a = x_2$, we obtain from Equation (A1) that $f(x_2) = \theta - (x_1 - x_2)^\gamma$. Given that $(x_1 - x_2)^\gamma < \theta$, then $f(x_2) > 0$. Hence, $f(a) > 0$ $\forall a \in (-\infty, x_2]$.

- If $a \in (x_2, x_1)$, then, using Equation (A2), $f'(a) = \gamma(x_1 - a)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > a > x_2$, we have $f'(a) > 0$, so $f(a)$ is strictly increasing on (x_2, x_1). Thus, $f(a)$ has a minimum at $a = x_2$. We have $f(x_2) = \theta - (x_1 - x_2)^\gamma > 0$, given that $(x_1 - x_2)^\gamma < \theta$. So $f(a) > 0$ $\forall a \in (x_2, x_1)$.

- If $a \in [x_1, +\infty)$, then, using Equation (A3), $f'(a) = -\gamma(a - x_1)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > x_2$, we have $a - x_1 < a - x_2 \Leftrightarrow \frac{1}{a-x_1} > \frac{1}{a-x_2}$. Given that $\gamma \leq 1$, we obtain $\left(\frac{1}{a-x_1}\right)^{1-\gamma} \geq \left(\frac{1}{a-x_2}\right)^{1-\gamma} \Leftrightarrow (a - x_1)^{\gamma-1} \geq (a - x_2)^{\gamma-1}$. Hence, $f'(a) \leq 0$, and $f(a)$ is weakly decreasing on $[x_1, +\infty)$. Thus, $f(a)$ has a minimum when $a \to +\infty$. As already shown (see Equation (A7)), $\lim_{a\to+\infty} f(a) = \theta > 0$. So $f(a) > 0$ $\forall a \in [x_1, +\infty)$.

  To sum up, we have shown that $f(a) > 0$ on $\mathbb{R}$ if $(\gamma \leq 1$ and $|x_1 - x_2|^\gamma < \theta)$. Thus, $(\gamma \leq 1$ and $|x_1 - x_2|^\gamma < \theta)$ implies the single-crossing property.

Third step. We now show that ($\gamma \neq 1$ and $|x_1 - x_2|^\gamma \leq \theta$) implies the single-crossing property. Given that we have already shown in the First step that $\gamma > 1$ implies the single-crossing property, and given that we have already shown in the Second step that $|x_1 - x_2|^\gamma < \theta$ also implies the single-crossing property, we only need to show that ($\gamma < 1$ and $|x_1 - x_2|^\gamma = \theta$) implies the single-crossing property.

- If $a \in (-\infty, x_2]$, then, using Equation (A1), $f'(a) = \gamma(x_1 - a)^{\gamma-1} - \gamma(x_2 - a)^{\gamma-1}$. Given that $x_1 > x_2$, we have $x_1 - a > x_2 - a \Leftrightarrow \frac{1}{x_1-a} < \frac{1}{x_2-a}$. Given that $\gamma < 1$, we obtain $\left(\frac{1}{x_1-a}\right)^{1-\gamma} < \left(\frac{1}{x_2-a}\right)^{1-\gamma} \Leftrightarrow (x_1 - a)^{\gamma-1} < (x_2 - a)^{\gamma-1}$. Hence, $f'(a) < 0$, and $f(a)$ is strictly decreasing on $(-\infty, x_2]$. Thus, $f(a)$ has a minimum at $a = x_2$ on the interval $(-\infty, x_2]$. If $a = x_2$, we obtain from Equation (A1) that $f(x_2) = \theta - (x_1 - x_2)^\gamma$. Given that $(x_1 - x_2)^\gamma = \theta$, then $f(x_2) = 0$. Given that $f(a)$ is strictly decreasing in a, there exists a unique $a^* \in (-\infty, x_2]$ such that $f(a^*) = 0$ and it is $a^* = x_2$.

- If $a \in (x_2, x_1)$, then, using Equation (A2), $f'(a) = \gamma(x_1 - a)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > a > x_2$, we have $f'(a) > 0$, so $f(a)$ is strictly increasing on (x_2, x_1). So $f(a) > 0\ \forall a \in (x_2, x_1)$.

- If $a \in [x_1, +\infty)$, then, using Equation (A3), $f'(a) = -\gamma(a - x_1)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > x_2$, we have $a - x_1 < a - x_2 \Leftrightarrow \frac{1}{a-x_1} > \frac{1}{a-x_2}$. Given that $\gamma < 1$, we obtain $\left(\frac{1}{a-x_1}\right)^{1-\gamma} > \left(\frac{1}{a-x_2}\right)^{1-\gamma} \Leftrightarrow (a - x_1)^{\gamma-1} > (a - x_2)^{\gamma-1}$. Hence, $f'(a) < 0$, and $f(a)$ is strictly decreasing on $[x_1, +\infty)$. Thus, $f(a)$ has a minimum when $a \to +\infty$. As already shown (see Equation (A7)), $\lim_{a\to+\infty} f(a) = \theta > 0$. So $f(a) > 0\ \forall a \in [x_1, +\infty)$.

  To sum up, we have shown that if ($|x_1 - x_2|^\gamma = \theta$ and $\gamma < 1$), there exists a unique $a^* \in \mathbb{R}$ such that $f(a^*) = 0$ and it is $a^* = x_2$. So ($|x_1 - x_2|^\gamma = \theta$ and $\gamma < 1$) implies the single-crossing property.

Fourth step. We now show that ($\gamma \geq 1$ and $|x_1 - x_2|^\gamma \neq \theta$) implies the single-crossing property. Given that we have shown in the First step that $\gamma > 1$ implies the single-crossing property, and given that we have also shown in the Second step that $|x_1 - x_2|^\gamma < \theta$ also implies the single-crossing property, we only need to show that ($\gamma = 1$ and $|x_1 - x_2|^\gamma > \theta$) implies the single-crossing property.

- If $a \in (-\infty, x_2]$, then, using Equation (A1), $f'(a) = \gamma(x_1 - a)^{\gamma-1} - \gamma(x_2 - a)^{\gamma-1}$. Given that $\gamma = 1$, we obtain $f'(a) = 0$, so $f(a)$ is a constant for all $a \in (-\infty, x_2]$. It means that $\forall a \in (-\infty, x_2]$, $f(a) = f(x_2)$. Hence, Equation (A1) becomes $f(a) = \theta - (x_1 - x_2)^\gamma$. Given that $|x_1 - x_2|^\gamma > \theta$, we obtain $f(a) < 0$ $\forall a \in (-\infty, x_2]$.

- If $a \in (x_2, x_1)$, then, using Equation (A2), $f'(a) = \gamma(x_1 - a)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $x_1 > a > x_2$, we have $f'(a) > 0$, so $f(a)$ is strictly increasing on (x_2, x_1). Thus, $f(a)$ has a minimum at $a = x_2$ and a maximum at $a = x_1$. Using Equation (A2), we obtain $f(x_2) = \theta - (x_1 - x_2)^\gamma < 0$, given that $(x_1 - x_2)^\gamma > \theta$, and $f(x_1) = \theta + (x_1 - x_2)^\gamma > 0$, given that $\theta > 0$ and $x_1 > x_2$. If so, and given that $f(a)$ is strictly increasing on (x_2, x_1), there exists a unique $a^* \in (x_2, x_1)$ such that $f(a^*) = 0$.

- If $a \in [x_1, +\infty)$, then, using Equation (A3), $f'(a) = -\gamma(a - x_1)^{\gamma-1} + \gamma(a - x_2)^{\gamma-1}$. Given that $\gamma = 1$, we obtain $f'(a) = 0$, so $f(a)$ is a constant for all $a \in [x_1, +\infty)$. It means that $\forall a \in [x_1, +\infty)$, $f(a) = f(x_1)$. Hence, Equation (A3) becomes $f(a) = \theta + (x_1 - x_2)^\gamma > 0$ $\forall a \in [x_1, +\infty)$.

  To sum up, we have shown that if ($\gamma = 1$ and $|x_1 - x_2|^\gamma > \theta$), there exists a unique $a^* \in \mathbb{R}$ such that $f(a^*) = 0$. Thus ($\gamma = 1$ and $|x_1 - x_2|^\gamma > \theta$) implies the single-crossing property.

Conclusion: we have shown that $\neg P$ implies the single-crossing property. Given that a statement and its contrapositive are equivalent, we have also shown that if the single-crossing property is not satisfied, then Property P holds, with $P =[(|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1)$ or $(|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1)]$. So $[(|x_1 - x_2|^\gamma > \theta$ and $\gamma < 1)$ or $(|x_1 - x_2|^\gamma = \theta$ and $\gamma = 1)]$ is a necessary condition for the single-crossing property not being satisfied.■

B Interpersonal incomparability of responses

This appendix shows that respondents interpret issue scales differently. This problem of interpersonal incomparability of responses can stem from different reasons. For instance, some respondents may lack some relevant knowledge on political matters. If so, their self-placement as well as the placement of the candidates are more likely to be noisy answers. It is also possible that some respondents have relevant information and are able to provide accurate responses in principle, but they have only a finite amount of time to process this information and provide answers in practice; so again, their self-placement as well as the one of the candidates are more likely to be noisy answers. Another reason is that some respondents may exaggerate the distance between themselves and candidates they view unfavorably. For example, a liberal respondent may place a conservative candidate more on the right than do conservative respondents, probably to exaggerate the distance between her and this candidate she views unfavorably. I show below that the latter problem occurs in the Discrete choice and the SUR samples described in Section 4 and exploited in Section 5.

Concerning the Discrete choice sample, Panel (B) in Figure B1 suggests that, on average, voters who consider themselves as liberal ($\widetilde{a} = 2$) or extremely liberal ($\widetilde{a} = 1$) place McCain more on the right than those who consider themselves as conservative ($\widetilde{a} = 6$) or extremely conservative ($\widetilde{a} = 7$). Indeed, $\text{Mean}(\widetilde{x}_m|\widetilde{a} = 2) = 5.91$ and $\text{Mean}(\widetilde{x}_m|\widetilde{a} =$

1) = 5.70 while Mean($\widetilde{x}_m|\widetilde{a} = 6$) = 5.24 and Mean($\widetilde{x}_m|\widetilde{a} = 7$) = 4.56. Note that the conditional means are depicted by (red) solid triangles in the box-and-whisker diagrams of Figure B1. The conditional $(0.25, 0.50, 0.75)$ quantiles of the box-and-whisker diagram in Panel (B) provide a similar story, so this trend seems to be robust to outliers.

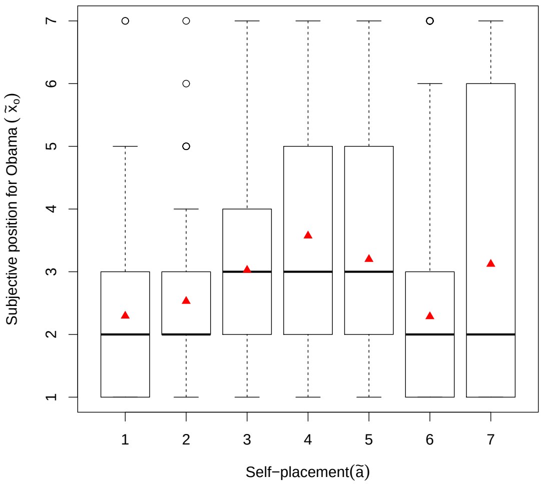
If one considers the SUR sample, the results are broadly equivalent: Panel (D) in Figure B1 shows that voters who consider themselves as liberal ($\widetilde{a} = 2$) or extremely liberal ($\widetilde{a} = 1$) place McCain more on the right than those who consider themselves as conservative ($\widetilde{a} = 6$) or extremely conservative ($\widetilde{a} = 7$). Indeed, Mean($\widetilde{x}_m|\widetilde{a} = 2$) = 5.88 and Mean($\widetilde{x}_m|\widetilde{a} = 1$) = 5.76 while Mean($\widetilde{x}_m|\widetilde{a} = 6$) = 5.19 and Mean($\widetilde{x}_m|\widetilde{a} = 7$) = 4.47. Again, this trend seems to be robust to outliers, given that the conditional $(0.25, 0.50, 0.75)$ quantiles in Panel (D) give a similar story.

Table B1: Statistical tests of interpersonal comparability of responses

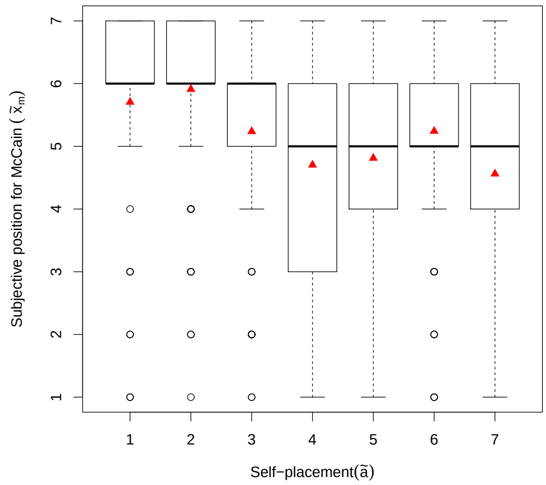
	Discrete choice sample		SUR sample	
	(A)	(B)	(C)	(D)
	Obama ($\widetilde{x}_o$)	McCain ($\widetilde{x}_m$)	Obama ($\widetilde{x}_o$)	McCain ($\widetilde{x}_m$)
[1] OLS estimates				
$\widehat{\beta}_{1,j}$	-0.001	-0.162***	-0.001	-0.173***
	(0.030)	(0.025)	(0.029)	(0.024)
	[0.960]	[0.000]	[0.950]	[0.000]
$\widehat{\beta}_{0,j}$	2.956***	5.786***	2.974***	5.823***
	(0.119)	(0.111)	(0.115)	(0.106)
[2] Spearman's rank correlations				
$\widehat{\rho}(\widetilde{a}, \widetilde{x}_j)$	-0.103***	-0.206***	-0.096***	-0.211***
	[0.000]	[0.000]	[0.000]	[0.000]
N	1320	1320	1449	1449

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 $[\cdot]$.

Table B1 provides formal tests of this specific form of interpersonal incomparability. I 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

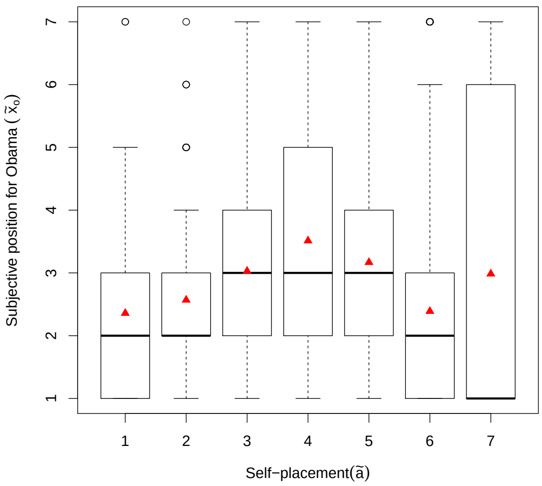


(A) Obama

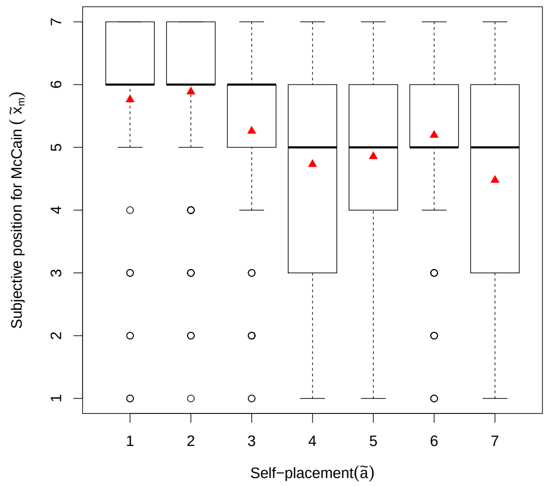


(B) McCain

Discrete choice sample



(C) Obama



(D) McCain

SUR sample

Figure B1: Reported location of Obama ($\widetilde{x}_o$) and McCain ($\widetilde{x}_m$) conditional on self-placement ($\widetilde{a}$) (Discrete choice and SUR samples)

Notes: The four 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.

there is interpersonal incomparability of responses for candidate j, then the null hypothesis $H_0: \beta_{1,j} = 0$ should be rejected. Part [1] of Table B1 shows that $\widehat{\beta}_{1,j}$ is negative and significantly different from zero for McCain ($j = m$), but not for Obama ($j = o$). This is true if one considers the Discrete choice or the SUR sample. It means that liberal respondents place McCain more on the right than do conservative respondents but conservative respondents do not place Obama more on the left than do liberal respondents. Given the ordinal nature of the data, a Spearman's rho between $\widetilde{x}_j$ and $\widetilde{a}$ has also been considered to test the null of interpersonal comparability $H_0: \rho(\widetilde{a}, \widetilde{x}_j) = 0$. Part [2] of Table B1 shows that the null is again rejected for McCain, whatever the sample; the null is also rejected for Obama. All these results indicate a pronounced issue of interpersonal incomparability in responses.

C The Aldrich-McKelvey procedure

In this appendix, I give a short but formal presentation of the Aldrich-McKelvey procedure. It is a technique for estimating the positions of political stimuli (i.e., candidates and parties) and respondent bliss points on a common issue space, using the reported positions of the stimuli by the respondents as well as their self-placement. This appendix permits in particular to be explicit about a technical assumption of this procedure, i.e., that the sum of the true locations of the stimuli is zero and the sum of squares equal to one. I also show that if one considers only two stimuli, this technical assumption is not sufficient to determine the locations of the stimuli. However, it fully determines these locations if it is combined with the extra assumption that the researcher knows the stimulus the more conservative; see Remark C2. The appendix also permits to explain why the absence of variability in the reported location of the stimuli makes it impossible to estimate the Aldrich-McKelvey respondent-specific distortion parameters, denoted c_i and

w_i. As noticed in the main text, these distortion parameters, which appear in Equation (C1), are crucial to obtain the respondents' bliss points in the same policy space as the actual locations of the stimuli.

For each respondent i, the data reveal her reported positions of the stimuli locations, i.e., the responses $\widetilde{x}_{i,j}$, $j = 1, \ldots, J$. These responses follow a two-step process. The first step has to do with respondent i's perception of the stimuli locations $x_{i,j}$, $j = 1, \ldots, J$. Respondent i retrieves relevant information on the actual locations x_j, $j = 1, \ldots, J$, of the stimuli on $\mathbb{R}$. However, her perception is distorted. One reason of this distortion is that respondent i pushes the stimuli she views unfavorably toward the extremes, as shown in Appendix B. But other reasons may also generate a distortion in her perception, as described at the beginning of Appendix B. Whatever the reason for this distortion, her perception $x_{i,j}$ of the stimulus 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 stimulus j to the interviewer. This answer is assumed to be a (linear) distortion of her perception $x_{i,j}$ since there is not a common metric for placing the stimuli. Indeed, the response options offered by the question may appear as vague: the exact boundary between slightly liberal (i.e., [3] on the 7-point scale) and liberal (i.e., [2]) may be unclear. If so, there are the already mentioned distortion parameters c_i and w_i for each respondent i such that (Aldrich and McKelvey, 1977, Equation (3), p.114):

$$x_{i,j} = x_j + \varepsilon_{i,j} = c_i + w_i \widetilde{x}_{i,j} \tag{C1}$$

Remark C1 The assumption that the response is a two-step or more generally a multi-step process is in line with the current literature on the psychology of survey responses.

For instance, Tourangeau *et al.* (2000) consider that a survey response process involves four stages: (i) understanding the question, (ii) retrieving relevant information, (iii) using this information to make a judgment, and (iv) selecting and reporting of an answer. Each stage can add a level of noise to the responses. Stages (i)-(iii) correspond to the first step in the Aldrich-McKelvey procedure, and stage (iv) corresponds to the second step.

Now, consider the following matrix notation:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix} \quad \widetilde{X}_i = \begin{bmatrix} 1 & \widetilde{x}_{i,1} \\ 1 & \widetilde{x}_{i,2} \\ \vdots & \vdots \\ 1 & \widetilde{x}_{i,J} \end{bmatrix} \quad \delta_i = \begin{bmatrix} c_i \\ w_i \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{C2}$$

If the vector X of actual locations were known, then the best linear unbiased estimator $\widehat{\delta}_i$ of the distortion parameters for respondent i would be $\widehat{\delta}_i = (\widetilde{X}_i'\widetilde{X}_i)^{-1}\widetilde{X}_i'X$, and the sum of squared residuals for this respondent $(X - \widetilde{X}_i\widehat{\delta}_i)'(X - \widetilde{X}_i\widehat{\delta}_i)$.

To obtain the vector X of actual locations, a technical assumption is made: the sum of the true locations of the stimuli is zero and the sum of squares equal to one, i.e.,

Assumption (i) $\sum_{j=1}^{J} x_j = X'F = 0$ *and* $\sum_{j=1}^{J} x_j^2 = X'X = 1$

Then the total sum of squared residuals of all the respondents is minimized subject to Assumption (i). That is, a Lagrangian multiplier problem is set up as follows:

$$\mathcal{L}(\widehat{\delta}_i, X, \alpha_1, \alpha_2) = \sum_{i=1}^{N}(X - \widetilde{X}_i\widehat{\delta}_i)'(X - \widetilde{X}_i\widehat{\delta}_i) + 2\alpha_1 X'F + \alpha_2(X'X - 1) \tag{C3}$$

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 $[A - NI_J]X = \alpha_2 X$, where I_J is the $J \times J$ identity matrix (Aldrich and McKelvey, 1977, Equation (24), p.115). By definition, α_2 is an eigenvalue of $[A - NI_J]$ and X an eigenvector of $[A - NI_J]$ (Fuente, 2000, p.146). It can then be shown that $-X'[A - NI_J]X = \sum_{i=1}^{N}(X - \widetilde{X}_i\widehat{\delta}_i)'(X - \widetilde{X}_i\widehat{\delta}_i) = -\alpha_2$ (Aldrich and McKelvey, 1977, Equation (26), p.116). In words, the solution X is the eigenvector of $[A - NI_J]$ with the highest (negative) nonzero eigenvalue.

Remark C2 If the number of stimuli is 2 ($J = 2$), then Assumption (i) fully determines the locations of the stimuli if it is combined with an extra assumption:

Assumption (ii) $x_1 > x_2$

Assumption (ii) states that the researcher knows the stimulus the more conservative; it is assumed w.l.o.g. that stimulus 1 is more conservative than stimulus 2. Under Assumptions (i) and (ii), the respondents' reported positions of the two stimuli $\widetilde{x}_{i,1}$ and $\widetilde{x}_{i,2}$ for $i = 1, \ldots, N$ do not matter to determine the locations of the two stimuli; there is no need to set up the Lagrange multiplier problem (C3). Indeed, if $J = 2$, the part of Assumption (i) which states that the sum of the true locations of the stimuli is zero permits to write that $\sum_{j=1}^{2} x_j = 0$, so $x_1 = -x_2$. Then, given that $x_1 = -x_2$, the part of Assumption (i) which states that the sum of squares equal to one can be rewritten as $\sum_{j=1}^{2} x_j^2 = x_1^2 + (-x_1)^2 = 2x_1^2 = 1$. Thus, $x_1 = \pm\sqrt{\frac{1}{2}}$. Given that $x_1 = -x_2$, if $x_1 = \sqrt{\frac{1}{2}}$, then $x_2 = -\sqrt{\frac{1}{2}}$; and if $x_1 = -\sqrt{\frac{1}{2}}$, then $x_2 = \sqrt{\frac{1}{2}}$. Finally, Assumption (ii) permits to obtain that $x_1 = \sqrt{\frac{1}{2}}$ and $x_2 = -\sqrt{\frac{1}{2}}$.

Having obtained the stimuli locations X, it is then possible to obtain each respondent's bliss point in the common space, the real line. Indeed, one can estimate the distortion parameters by calculating $\widehat{\delta}_i = (\widetilde{X}_i'\widetilde{X}_i)^{-1}\widetilde{X}_i'X$. Then, one has to subject each respondent's bliss point to the same transformation that her reported positions of the stimuli are subjected to. Given that $\widetilde{a}_i$ is respondent i's self-placement, her 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).

Remark C3 If the number of stimuli is 2, i.e., $J = 2$, and if the reported placement of stimuli 1 and 2 by respondent i is identical, i.e., $\widetilde{x}_{i,1} = \widetilde{x}_{i,2}$, then one cannot obtain the respondent-specific distortion parameters $\widehat{c}_i$ and $\widehat{w}_i$. The reason is that in such case, the two columns of the 2×2 respondent-specific data matrix $\widetilde{X}_i$ in (C2) are perfectly collinear. It is now obvious that by adding additional stimuli, so $J > 2$, the two columns of the $J \times 2$ respondent-specific data matrix $\widetilde{X}_i$ are less likely to be perfectly collinear.

Remark C4 In Sections 4 and 5, the computations are carried out in the R environment and make use of the basicspace package (Poole *et al.*, 2016). I need to obtain the actual locations of Obama (x_o) and McCain (x_m), the two main candidates of the 2008 Presidential election. As pointed out in Section 4, other candidates competed at this election. But the respondents were not asked to report their position. So the data reveal the reported positions of only two candidates, i.e., the responses $\widetilde{x}_{i,o}$ and $\widetilde{x}_{i,m}$ for each respondent i. Given Remarks C2 and C3, I add the reported positions of the Democratic party ($\widetilde{x}_{i,dem}$) and the Republican party ($\widetilde{x}_{i,rep}$) in the Aldrich-McKelvey procedure to obtain the actual locations of Obama and McCain. So I consider the reported positions of four stimuli: $\widetilde{x}_{i,m}$, $\widetilde{x}_{i,o}$, $\widetilde{x}_{i,dem}$ and $\widetilde{x}_{i,rep}$. Using the Discrete choice sample, the actual locations of the stimuli obtained via the Aldrich-McKelvey method are:

$$x_o = -0.5180862 \quad x_m = 0.4780555 \quad x_{dem} = -0.4811076 \quad x_{rep} = 0.5211383$$

Using the SUR sample, the results are almost similar:

$$x_o = -0.5180191 \quad x_m = 0.4783012 \quad x_{dem} = -0.4811876 \quad x_{rep} = 0.5209055$$

D Bootstrap methods

The locations of the candidates and the respondents' bliss points are estimates based on the Aldrich-McKelvey procedure described in Appendix C. The uncertainty in the estimates of these variables can influence the different test statistics considered in the conditional logit and the SUR model. To take into account these two-step estimation problems, I consider bootstrap tests. In this appendix, I describe the bootstrap methods that I exploit to test the different hypotheses considered in this paper. Observe that in the main text, in Section 5, I also consider for $\widehat{\theta}$ and $\widehat{\gamma}$ test statistics which treat the distances between a_i and x_j as observed variables, i.e., ignoring any estimation error in these variables. I consider in the presentation of the bootstrap methods that the two candidates of interest are Obama ($j = o$) and McCain ($j = m$), as in Sections 4 and 5. Finally, before to step any further, recall that to find the actual location of Obama (x_o) and McCain (x_m), I include in the Aldrich-McKelvey procedure the reported positions of Obama ($\widetilde{x}_{i,o}$) and McCain ($\widetilde{x}_{i,m}$) of each respondent i, as well as her reported positions of the Democratic party ($\widetilde{x}_{i,dem}$) and the Republican party ($\widetilde{x}_{i,rep}$). A full explanation is provided in Appendix C, in particular in Remark C4, as well as in Remarks C2 and C3.

Difference in valence. Let me begin by the first hypothesis which concerns the difference in valence among Obama and McCain. This is the introduction of a valence difference between Obama and McCain which makes that the single-crossing property may not hold. So the first null hypothesis is $\mathrm{H}_0 : \theta = 0$ against $\mathrm{H}_A : \theta \neq 0$. I describe the method for the discrete choice model. The method remains the same for the SUR model, with the only difference being the size of each bootstrap sample: $N = 1449$ for the SUR and $N = 1320$ for the discrete choice model. My approach for this two-sided test at level α is to consider a bootstrap percentile method: I find the lower $\frac{\alpha}{2}$ and upper $\frac{\alpha}{2}$ quantiles of B bootstrap estimates $\widehat{\theta}^*_b$, $b = 1, \ldots, B$; if $\mathrm{H}_0 : \theta = 0$ falls outside this region, the null is rejected. I fully explain the procedure below.

Let $\omega_i = (y_i, \widetilde{x}_{i,o}, \widetilde{x}_{i,m}, \widetilde{x}_{i,dem}, \widetilde{x}_{i,rep}, \widetilde{a}_i)$ be the respondent i's answers to the set of questions which are needed for the estimations of the discrete choice model, and $\mathbb{W} = (\omega_1, \omega_2, \ldots, \omega_N)'$ the sample. A bootstrap sample $\mathbb{W}^*$ of size $N = 1320$ is obtained by sampling from $\omega_1, \omega_2, \ldots, \omega_N$ with replacement. The steps to obtain the bootstrap test can be summarized as:

(i.) Draw $B = 999$ bootstrap samples $\mathbb{W}^*$ of size $N = 1320$.

(ii.) For each bootstrap sample $\mathbb{W}^*$: first, estimate the actual locations of the stimuli, i.e., $\{x_o^*, x_m^*, x_{dem}^*, x_{rep}^*, \}$, as well as the bliss point a_i^* of each respondent i, using the Aldrich-McKelvey method. Second, use the values obtained for x_o^*, x_m^* and a_i^* to compute the distances $|x_o^* - a_i^*|$ and $|x_m^* - a_i^*|$ for each respondent i in the bootstrap sample, and estimate the conditional logit. Given that there are B bootstrap samples, this leads to B estimates $\widehat{\theta}_b^*$, $b = 1, \ldots, B$.

(iii.) Consider the empirical distribution of the B bootstrap estimates $\widehat{\theta}_b^*$, $b = 1, \ldots, B$. Denote by $\widehat{\theta}_{0.025}^*$ and $\widehat{\theta}_{0.975}^*$ the 2.5th and the 97.5th percentiles of this empirical distribution. The percentile 95 percent confidence interval for θ is then $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$.

(iv.) Reject the null hypothesis if $\mathrm{H}_0 : \theta = 0$ falls outside $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$.

The exponent. The second hypothesis concerns γ. When one candidate has a valence-advantage, the single-crossing property may not hold if $\gamma < 1$ or if $\gamma = 1$. As noticed in Section 3, the most interesting case is when $\gamma < 1$ because it is when $u(a, x_o, \theta, \gamma) = u(a, x_m, \gamma)$ may have two solutions in a, and the ends may be against the middle. So the second null hypothesis is $\mathrm{H}_0 : \ \gamma \geq 1$ against $\mathrm{H}_A : \ \gamma < 1$. My approach for this lower one-tailed test is very similar to the one presented for the difference in valence. The only difference is that for each bootstrap sample, once the conditional logit model is estimated, $\widehat{\gamma}^*$ is saved, and the percentile 95 percent confidence interval for γ is constructed based on a lower one-tailed test. Denote by $\widehat{\gamma}^*_{0.95}$ the 95th percentiles of the empirical distribution of $B = 999$ bootstrap estimates $\widehat{\gamma}^*_b$, $b = 1, \ldots, B$. The percentile 95 confidence interval is $(-\infty, \widehat{\gamma}^*_{0.95}]$ in this case. And the null hypothesis is rejected if $\mathrm{H}_0 : \ \gamma \geq 1$ falls outside this interval.

The single-crossing property. The third hypothesis concerns the single-crossing property per se. This third null hypothesis is H_0 : "the single-crossing property holds" versus H_A :

"the single-crossing property does not hold". Recall Proposition 1: the single-crossing property does not hold if and only if $[(|x_o - x_m|^\gamma > \theta$ and $\gamma < 1)$ or $(|x_o - x_m|^\gamma = \theta$ and $\gamma = 1)]$. The general approach for this test is very similar to the one presented for the difference in valence. For each bootstrap sample $\mathbb{W}^*$, I first estimate the actual locations of the stimuli, i.e., $\{x_o^*, x_m^*, x_{dem}^*, x_{rep}^*, \}$, and the bliss point a_i^* of each respondent i using the Aldrich-McKelvey method, then I use the values obtained for x_o^*, x_m^* and a_i^* to compute the distances $|x_o^* - a_i^*|$ and $|x_m^* - a_i^*|$ for each respondent i, and I estimate the discrete choice model to obtain $\widehat{\theta}^*$ and $\widehat{\gamma}^*$. Thus, for each bootstrap sample $\mathbb{W}^*$, I obtain the set of parameter estimates $[x_o^*,\ x_m^*,\ \widehat{\theta}^*,\ \widehat{\gamma}^*]$ to see if the single-crossing holds. Given that there are $B = 999$ bootstrap samples, there are B sets of parameter estimates $[x_{ob}^*,\ x_{mb}^*,\ \widehat{\theta}_b^*,\ \widehat{\gamma}_b^*]$, $b = 1, \ldots, B$, and I count the proportion of bootstrap samples for which the single-crossing property holds. It permits to obtain an estimate of the achieved significance level (or p-value) of the test which is:

$$\widehat{ASL} = 1 - \frac{\sharp\{b = 1, \ldots, B\ ;\ [(|x_{ob}^* - x_{mb}^*|^{\widehat{\gamma}_b^*} > \widehat{\theta}_b^* \text{ and } \widehat{\gamma}_b^* < 1) \text{ or } (|x_{ob}^* - x_{mb}^*|^{\widehat{\gamma}_b^*} = \widehat{\theta}_b^* \text{ and } \widehat{\gamma}_b^* = 1)]\}}{B}$$

And one fails to reject the null hypothesis whenever $\widehat{ASL}$ is larger than standard levels of significance; I consider a 5 percent significance level, like for the other tests.

E Robustness checks

In this appendix, I provide the tables of results of the two robustness checks described in Subsections 6.1 and 6.2.

In Section 4, I observe that missing observations in both the Discrete choice and SUR samples primarily stem from the self-placement question. Respondents who answered "Haven't thought much about it" or "Don't know", along with those who considered themselves as "Moderate/middle of the road" (i.e., $\widetilde{a}_i = 4$), were prompted to categorize themselves via a follow-up question. Table E1 utilizes the responses to this follow-up question to expand the sizes of the Discrete choice and SUR samples. The coding assumptions, as described in Subsection 6.1, involve various considerations. The bootstrap methods employed are similar to those explained in Appendix D. The results are discussed in Subsection 6.1.

In Section 4, I also emphasize the importance of considering more than just two stimuli when constructing the policy space. But a policy space can still be constructed by considering only two stimuli. The drawback is that the resulting locations are fully determined by Assumptions (i) and (ii), as explained in Remark C2 of Appendix C. In other words, the reported positions of these stimuli by the respondents become irrelevant in determining their actual location. Despite of this drawback, some readers may be interested in the outcomes obtained when using only the reported positions of Obama and McCain to construct the policy space. Table E2 presents these findings.

Before to comment them, observe that Table E2 provides confidence intervals based on bootstrap percentiles for θ and γ. It also provides the achieved significance levels of the bootstrap single-crossing tests. As mentioned in Subsection 6.2, the general bootstrap algorithm differs from the one described in Appendix D. The reason is that the reported positions of Obama and McCain by the respondents become irrelevant to determine their location in a first step: $x_o = -\sqrt{\frac{1}{2}}$ and $x_m = \sqrt{\frac{1}{2}}$ whatever the sample. Given that there is no uncertainty in x_o and x_m (nor in a_i, $i = 1, \ldots, N$), there is no two-step estimation problem. Recall that in Appendix D, if we consider the Discrete choice sample, the bootstrap methods were based on the sample $\mathbb{W} = (\omega_1, \omega_2, \ldots, \omega_N)'$, where $\omega_i = (y_i, \widetilde{x}_{i,o}, \widetilde{x}_{i,m}, \widetilde{x}_{i,dem}, \widetilde{x}_{i,rep}, \widetilde{a}_i)$. For each bootstrap sample $\mathbb{W}^*$, we estimated first the actual locations x_o^* and x_m^* and the bliss point a_i^* of each respondent i using the Aldrich-McKelvey method, and then the conditional logit to obtain $\widehat{\theta}^*$ and $\widehat{\gamma}^*$. In Table E2, the bootstrap

Table E1: Discrete choice and SUR estimates for the 2008 Presidential election using the follow-up question to increase the sample sizes

	(A) Discrete choice model (Stated choice)	(B) SUR model (Thermometer scores)
Obama's valence-advantage		
$\widehat{\theta}$	0.386	0.337
	(0.036)	(0.053)
	$[0.316, 0.458]^{ii.}$	[0.258, 0.428]
Exponent parameter		
$\widehat{\gamma}$	0.846	0.333
	(0.088)	(0.056)
	$(-\infty, 0.988]^{iii.}$	$(-\infty, 0.423]$
Scale parameters		
$\widehat{\sigma}$	0.409	
	(0.031)	
$\widehat{c}$		83.876
		(5.542)
$\widehat{\beta}$		40.921
		(5.658)
N	1721	1895
Log-likelihood	-740.987	-17388.000
Bootstrap single-crossing test, with $B = 999$ replications		
$\widehat{ASL}$	0.041	0.000

Notes: i. Standard errors are in parentheses.
ii. The intervals in brackets below the estimated standard errors of $\widehat{\theta}$ correspond to the percentile 95 percent confidence intervals for the parameters. These intervals are equal-tailed: they are the distance between the lower 0.025 and upper 0.025 quantiles of $B = 999$ bootstrap estimates of the parameter of interest. For instance, the interval [0.316, 0.458] below the estimated standard error of $\widehat{\theta}$ in the case of the discrete choice model corresponds to the percentile 95 percent confidence interval $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$ for θ, where $\widehat{\theta}^*_{0.025} = 0.316$ is the lower 0.025 and $\widehat{\theta}^*_{0.975} = 0.458$ the upper 0.025 quantiles of the $B = 999$ bootstrap estimates $\widehat{\theta}^*(b)$, $b = 1, \ldots, B$.
iii. The intervals below the estimated standard errors of $\widehat{\gamma}$ correspond to the percentile 95 percent confidence intervals based on a lower one-tailed alternative test. For instance, the interval $(-\infty, 0.988]$ below the estimated standard error of $\widehat{\gamma}$ in the case of the discrete choice model corresponds to the percentile 95 percent confidence interval $(-\infty, \widehat{\gamma}^*_{0.95}]$ for γ, where $\widehat{\gamma}^*_{0.95} = 0.988$ is the upper 0.05 quantiles of the $B = 999$ bootstrap estimates $\widehat{\gamma}^*(b)$, $b = 1, \ldots, B$.
iv. Bootstrap single-crossing test ($\widehat{ASL}$) provides the achieved significance level of the test H_0 : "the single-crossing property holds" versus H_A : "the single-crossing property does not hold". Given that there are $B = 999$ bootstrap samples, there are B sets of parameter estimates $[x^*_{ob}, x^*_{mb}, \widehat{\theta}^*_b, \widehat{\gamma}^*_b]$, $b = 1, \ldots, B$. I check for each set of parameter estimates if the single-crossing property holds. The proportion of bootstrap samples for which the single-crossing property holds is the estimate of the achieved significance level:

$$\widehat{ASL} = 1 - \frac{\sharp\{b = 1, \ldots, B \ ; \ [(|x^*_{ob} - x^*_{mb}|^{\widehat{\gamma}^*_b} > \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b < 1) \text{ or } (|x^*_{ob} - x^*_{mb}|^{\widehat{\gamma}^*_b} = \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b = 1)]\}}{B}$$

The method is discussed in Appendix D.

methods are based on the sample $\mathbb{H} = (\eta_1, \eta_2, \ldots, \eta_N)'$, where $\eta_i = (y_i, x_o, x_m, a_i)$, with $x_o = -\sqrt{\frac{1}{2}}$ and $x_m = \sqrt{\frac{1}{2}}$. A bootstrap sample $\mathbb{H}^*$ of size N is obtained by sampling from $\eta_1, \eta_2, \ldots, \eta_N$ with replacement. For each bootstrap sample $\mathbb{H}^*$, I directly estimate the conditional logit. Given that there are B bootstrap samples, this leads to B estimates for $\widehat{\theta}_b^*$ and $\widehat{\gamma}_b$, $b = 1, \ldots, B$. I then use the empirical distribution of these bootstrap estimates to build the percentile confidence intervals.[12] Concerning, the single-crossing hypothesis, I count the proportion of bootstrap samples for which the single-crossing property holds.

Regarding the results, Obama is again the candidate with a valence-advantage. I obtain $\widehat{\theta} = 0.327$ with the discrete choice model(versus 0.269 in Table 2), and $\widehat{\theta} = 0.440$ with the SUR model (versus 0.227 in Table 2). In the case of the discrete choice model, the null hypothesis $\mathrm{H}_0: \ \theta = 0$ is rejected whether we use the t-ratio $\left(\frac{0.327}{0.042} = 7.738\right)$, which is larger than the critical values of the standard normal distribution for conventional levels of significance, or the 95 percent confidence interval based on bootstrap percentiles [0.248, 0.415]. The same conclusion is reached with the SUR model, as evident in Column (B) of Table E2.

Regarding the exponent parameter γ, its estimate is $\widehat{\gamma} = 0.666$ with the discrete choice model (compared to 0.818 in Table 2), and $\widehat{\gamma} = 0.392$ with the SUR model (versus 0.371 in Table 2). Once again, all the one-sided tests for $\mathrm{H}_0: \ \gamma \geq 1$ versus $\mathrm{H}_A: \ \gamma < 1$ reject the null hypothesis.

Finally, the single-crossing property is rejected at the 5 percent significance level in both the discrete choice model and the SUR model. The equation $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ based on the discrete choice estimates has again two solutions in a, $a^* \simeq 0.217$ and $a^{**} \simeq 23.663$, and a^{**} falls outside the observed data-range [-7.778, 6.364]. This issue,

[12] Given that there is no randomness in x_o, x_m, and a_i, $i = 1, \ldots, N$, there is no two-step estimation problem, so one may ask why I provide these percentile confidence intervals. Two main reasons explain this choice. First, it permits to have a homogeneous presentation of the different tables of results. Second, they could have lead to different conclusions compared to the simple t-tests.

again, does not occur with the SUR model: $\widehat{u}(a, x_o, \widehat{\theta}, \widehat{\gamma}) = \widehat{u}(a, x_m, \widehat{\gamma})$ has two solutions in a, $a^* \simeq 0.424$ and $a^{**} \simeq 1.556$, and both of them fall within the observed data-range [-7.778, 7.778]. So voters whose bliss points are to the right of 1.556 are closer to McCain, but their estimated utility is higher for Obama. Again, a reader may naturally ask if these respondents do rank Obama above McCain by providing a higher thermometer score to Obama. 78 respondents have their bliss points to the right of 1.556. Out of these 78 respondents, 28 rate Obama higher than McCain and 12 rate them equally.

F The impact of an increase in the valence-advantage parameter

In Subsection 6.3, I highlight that conditional to the fact that the single-crossing property is not satisfied in the theoretical policy space, the length of the interval (a^*, a^{**}) diminishes as the valence-advantage parameter θ increases. This appendix provides a graphical explanation to this claim.

The presentation follows Figure 1 in Section 1, in particular Panel (C) given that $\gamma < 1$ and the single-crossing property is not satisfied. Recall that in Figure 1, w.l.o.g., $x_1 > x_2$, i.e., candidate 1 locates on the right of candidate 2, and $\theta = \theta_1 - \theta_2 > 0$, i.e., candidate 1 has a valence-advantage over candidate 2. Let's consider an increase in θ_1, the valence associated to candidate 1, to θ_1', so $\theta' = \theta_1' - \theta_2 > \theta$.

The increase in θ_1 shifts $u(a, x_1, \theta_1, \gamma)$ upward in Figure F1. Initially, the set of voters who prefer the disadvantaged candidate in terms of valence is the open interval (a^*, a^{**}), where a^* and a^{**} are the solutions to the equation $u(a, x_1, \theta_1, \gamma) = u(a, x_2, \theta_2, \gamma)$. After the increase in θ_1, the set of voters who prefer the disadvantaged candidate in terms of valence is the open interval $(a^{'*}, a^{'**})$, where $a^{'*}$ and $a^{'**}$ are the solutions to the equation $u(a, x_1, \theta_1', \gamma) = u(a, x_2, \theta_2, \gamma)$. As shown in Figure F1, the lower bound a^* increases to

Table E2: Discrete choice and SUR estimates for the 2008 Presidential election using a policy space constructed with only two stimuli

	(A) Discrete choice model (Stated choice)	(B) SUR model (Thermometer scores)
Obama's valence-advantage		
$\widehat{\theta}$	0.327	0.440
	(0.042)	(0.059)
	[0.248, 0.415]	[0.330, 0.562]
Exponent parameter		
$\widehat{\gamma}$	0.666	0.392
	(0.053)	(0.034)
	$(-\infty, 0.740]^{iii.}$	$(-\infty, 0.439]$
Scale parameters		
$\widehat{\sigma}$	0.503	
	(0.036)	
$\widehat{c}$		70.367
		(1.233)
$\widehat{\beta}$		21.755
		(1.266)
N	1278	1398
Log-likelihood	-554.179	-12861.500
Bootstrap single-crossing test, with $B = 999$ replications		
$\widehat{ASL}$	0.000	0.000

Notes: i. Standard errors are in parentheses.
ii. The intervals in brackets below the estimated standard errors of $\widehat{\theta}$ correspond to the percentile 95 percent confidence intervals for the parameters. These intervals are equal-tailed: they are the distance between the lower 0.025 and upper 0.025 quantiles of $B = 999$ bootstrap estimates of the parameter of interest. For instance, the interval [0.248, 0.415] below the estimated standard error of $\widehat{\theta}$ in the case of the discrete choice model corresponds to the percentile 95 percent confidence interval $[\widehat{\theta}^*_{0.025}, \widehat{\theta}^*_{0.975}]$ for θ, where $\widehat{\theta}^*_{0.025} = 0.248$ is the lower 0.025 and $\widehat{\theta}^*_{0.975} = 0.415$ the upper 0.025 quantiles of the $B = 999$ bootstrap estimates $\widehat{\theta}^*(b)$, $b = 1, \ldots, B$.
iii. The intervals below the estimated standard errors of $\widehat{\gamma}$ correspond to the percentile 95 percent confidence intervals based on a lower one-tailed alternative test. For instance, the interval $(-\infty, 0.740]$ below the estimated standard error of $\widehat{\gamma}$ in the case of the discrete choice model corresponds to the percentile 95 percent confidence interval $(-\infty, \widehat{\gamma}^*_{0.95}]$ for γ, where $\widehat{\gamma}^*_{0.95} = 0.740$ is the upper 0.05 quantiles of the $B = 999$ bootstrap estimates $\widehat{\gamma}^*(b)$, $b = 1, \ldots, B$.
iv. Bootstrap single-crossing test ($\widehat{ASL}$) provides the achieved significance level of the test H_0 : "the single-crossing property holds" versus H_A : "the single-crossing property does not hold". The method is different than the one presented in Appendix D: here, the locations of Obama $\left(x_o = -\sqrt{\frac{1}{2}}\right)$ and McCain $\left(x_m = \sqrt{\frac{1}{2}}\right)$ are fully determined by Assumptions (i) and (ii) (see Remark C2 in Appendix C), so there is no uncertainty in x_o and x_m. For each bootstrap sample, I directly estimate the conditional logit or the SUR to obtain $\widehat{\theta}^*$ and $\widehat{\gamma}^*$. Given that there are $B = 999$ bootstrap samples, there are B sets of parameter estimates $[\widehat{\theta}^*_b, \widehat{\gamma}^*_b]$, $b = 1, \ldots, B$. I check for each set of parameter estimates if the single-crossing property holds. The proportion of bootstrap samples for which the single-crossing property holds is the estimate of the achieved significance level:

$$\widehat{ASL} = 1 - \frac{\sharp\{b = 1, \ldots, B \; ; \; [(|x_o - x_m|^{\widehat{\gamma}^*_b} > \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b < 1) \text{ or } (|x_o - x_m|^{\widehat{\gamma}^*_b} = \widehat{\theta}^*_b \text{ and } \widehat{\gamma}^*_b = 1)]\}}{B}$$

.

a'^{*}, while the upper bound a^{**} decreases to a'^{**}, so the length of the interval (a^*, a^{**}) diminishes.

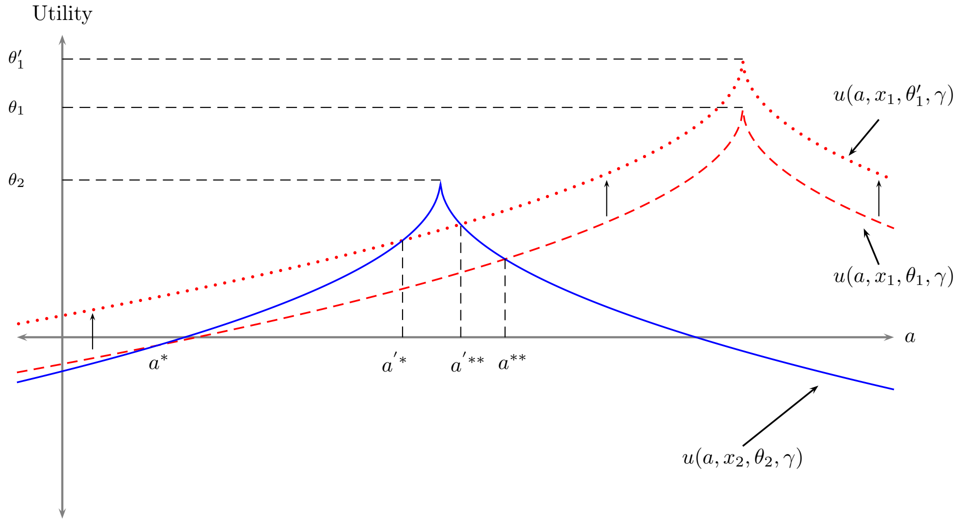


Figure F1: The effects of an increase in the valence-advantage of candidate 1 over candidate 2 when the single-crossing property is not satisfied (parameters: $\gamma = 0.5$, $x_1 = 1$, $\theta_1 = 1.9$, $\theta_1' = 2.3$, $x_2 = -1$, $\theta_2 = 1.3$)