PA

ARTICLE

Vote Choices and Valence: Intercepts and Alternate Specifications

Ingrid Mauerer[1] ⬤ and Gerhard Tutz[2]

[1]Faculty of Economics, University of Malaga, Campus El Ejido, 29013 Malaga, Spain; [2]Department of Statistics, LMU Munich, Akademiestraße 1, 80799 München, Germany.

Corresponding author: Ingrid Mauerer; Email: ingridmauerer@uma.es

Abstract
Valence is a crucial concept in studying spatial voting and party competition. The widely adopted approach is to rely on intercepts of vote choice models and to infer, based on their size and direction, how valence affects party strategies in empirical settings. The approach suffers from fundamental statistical flaws. This contribution provides the statistical fundamentals to advance the empirical modeling of valence. It proposes an appropriate modeling approach to interpret intercepts as valences and alternate specifications to parameterize the effects of valence.

Edited by: Jeff Gill

1. Introduction

A key concern of students of spatial voting and party competition is how valence—beyond policy aspects—affects party strategies (for recent reviews, see Adams, Merrill, and Zur 2020; Evrenk 2019; Magyar, Wagner, and Zur 2023). The following approach has been widely applied for decades to study the impact of valence: researchers estimate a vote choice model consisting of choice attributes (spatial proximities), chooser attributes (voter demographics) and intercepts, define the intercepts as valences, and based on the sign and the size of the intercepts, they reach conclusions about how valence influences party vote shares and positional strategies (e.g., Schofield and Sened 2005a, 2006; Zur 2021a,b). Section A of the Supplementary Material contains a not exhaustive list of 32 references, broadly published in top journals and publishing houses and widely cited that adopt this approach.

Mauerer (2020) highlights several difficulties that arise when relying on intercepts as valences, such as their dependence on arbitrary coding decisions of chooser attributes, and recommends studying valence qualities by covariates. However, she does not provide the statistical fundamentals to do so. The present contribution takes up this task and pursues three objectives to advance the empirical modeling of valence.

First, we clarify the interpretation of intercepts by investigating their link to chooser attributes and choice probabilities. We discuss the conventional identifiability restriction and present choice models imposing an identifiability approach that frees researchers from arbitrary coding decisions when interpreting intercepts as valences. Second, we propose an alternate strategy to intercepts to study the impact of valence. We outline a modeling approach and parameterization to incorporate valence as an additional observable utility source. Third, we discuss different specification strategies.

We accomplish these objectives based on identification issues and resulting model properties as well as covariate specification and effect parameterization strategies and illustrate implications for substantive interpretation using national election data.[1] We do not aim to provide a new definition of valence in terms of a substantive meaning or operationalization but investigate the statistical fundamentals.

Next, we outline the background and the analytical challenges involved (see also Mauerer 2020), and lay out our objectives in detail. Section 2 briefly reviews the standard choice model and identification issues to set the methodological ground. Section 3 discusses the impact of coding schemes on interpreting intercepts as valences. Section 4 outlines how to model valence as an observable source of utility. Section 5 closes with concluding remarks.

1.1. Background, Analytical Challenges, and Objectives

The concept of *valence* goes back to Stokes (1963). Instead of competing on policy issues where parties and voters can take different stands, his key argument is that the empirical reality is characterized by competition on competence, performance, trust, handling abilities, success, or lack thereof. Since Stokes' original conceptualization, the literature has gone many different ways to define and incorporate valence in theoretical and empirical models, what Green and Jennings (2017, 550–551) call the "Valence Soup" in reviewing the immense literature that resulted meanwhile:

> "Thus far authors have defined valence as a valence dimension, a party valence score, valence as a candidate's character or strategic advantage, a leader advantage or disadvantage, valence as a strategic advantage, as candidate quality, candidate experience, education or the lack thereof, as party activism, the level of activist support, candidate spending, the reputation of candidates, scandals and corruption (or their absence) in political parties and corruption at the level of candidates."

Many more definitions can be added to this list when inspecting the enormous formal literature (for a review, see Evrenk 2019). Within the spatial voting literature in the tradition of Downs (1957) where spatial proximity is the primary source of voter utility, valence is frequently vaguely defined as a second dimension of competition, a quality of a party (or candidate) that is not policy-related, that marks a difference between the parties, and this difference benefits the parties.

Our contribution operates within the influential literature on probabilistic spatial voting models, where a random aspect that is independent from spatial considerations is incorporated (e.g., Adams 1999; Burden 1997; Coughlin 1992; Enelow and Hinich 1989). This prominent research strand settled on the discrete choice framework (see, e.g., Train 2009) to arrive at empirical spatial vote choice models because the framework comes with several attractive features. In particular, both observed and unobserved choice determinants affect voter utility, and a random utility maximization process is the typical underlying choice rule. It also allows exploring how attributes of choice alternatives (e.g., voter–party spatial proximities) and choosers (i.e., voters) determine vote choices (e.g., Alvarez and Nagler 1998). While choice models can be formulated in many different ways, the conditional logit model (McFadden 1974) has become the standard choice model[2] in the study of spatial voting (e.g., Adams, Merrill III, and Grofman 2005; Kedar 2005; Merrill III and Adams 2001; Stoetzer and Zittlau 2015).

A widely adopted approach to integrating valence in empirical spatial vote choice models relies on the intercepts (see Section A of the Supplementary Material). The intercepts as a measure of valence were introduced by the *Spatial Valence Model of Politics* (e.g., Schofield and Sened 2005b, 2006) where

---

[1]Replication data and code for this paper are available in Mauerer and Tutz (2023b) at https://doi.org/10.7910/DVN/SKWTGS.

[2]The model assumes that the unobserved factors are i.i.d. random variables that follow a maximum extreme value distribution, which yields a closed-form of the log-likelihood and the well-known maximum likelihood estimation can be applied to obtain parameter estimates (for a recent model review in the context of spatial voting, see Mauerer and Tutz 2023a).

they are understood as parties' (candidates') average nonpolicy qualities. A fundamental property of the standard choice model is that the intercepts represent the unobserved utility sources and, consequently, reproduce the vote shares in the data, given a particular set of covariates (Mauerer 2020, 308–310), which has major implications for the numerous works that rely on intercepts to understand how valence affects spatial competition. The typical result is that parties with large vote shares have a large valence, and parties with small vote shares have a small valence or are "valence disadvantaged."

Let us consider recent work to demonstrate that such results are no substantive findings on how valence affects party strategies and that merely a model property produces them. In the vibrant debate on the "collapse of centrist parties" in Europe, Zur (2021a,b) concludes that the decline of centrist party vote shares is the result of the loss of valence of these parties by drawing on the size and the direction of the intercepts.[3]

Take, for example, the 2013 German election data Zur (2021b) analyzes, where the party vote shares are: Major-right CDU: .404, major-left SPD: .302, FDP: .030, Greens: .096, Left: .128, popular-right AfD: .044. The FDP, which has the smallest vote share, is classified as the centrist party and specified as the reference party. First, consider an intercept-only model where the following intercept estimates reproduce the vote shares: CDU: 2.73, SPD: 2.44, FDP: 0, Greens: 1.30, Left: 1.59, AfD: .51. Since the FDP has the smallest vote share, all intercepts, which are relative to the FDP, must be positive. The intercept for the CDU is the largest because the difference in vote shares between the CDU and the FDP is the largest. Thus, the direction and size of the intercepts are directly related to the vote shares.

Next, consider the intercepts, given spatial proximity on the left-right dimension, from which Zur (2021b) derives his conclusions: CDU: 2.79, SPD: 2.22, FDP: 0, Greens: 1.10, Left: 1.78, AfD: 1.53. When the model accounts for spatial proximity, there is some change in the intercepts; however, they still mainly reflect the vote shares. Thus, the vote share decline of centrist parties is explained by their vote share decline, which is not an explanation but results from transforming vote shares into intercepts using statistical models. Model properties cause the result, not valence (dis)advantages or positional efforts. Put differently, all factors determining vote choice that are not specified by covariates end up in the intercepts, and these factors might or might not be related to valence (Mauerer 2020).

Another crucial model feature affects the interpretation of intercepts as valences: the inclusion of chooser attributes, that is, attributes that characterize voters. The influential contribution *A Unified Theory of Party Competition* by Adams *et al.* (2005) introduced them as *nonpolicy motivations*, which are, for example, socioeconomic factors or ties related to religion or class. These variables are of key theoretical importance as they integrate a behavioral perspective in the spatial modeling tradition to better understand centrifugal forces (see already Adams and Merrill III 1999a,b, 2000; Merrill III and Adams 2001). Besides the theoretical importance, such variables add substantial explanatory power, as numerous works in line with this prominent research strand demonstrate for several decades and many polities (e.g., Adams and Merrill III 1999a, 771, 787; 2000, 741). Chooser attributes also enter the empirical applications of Schofield's Spatial Valence Model, sometimes referred to as *socioeconomic valences* (e.g., Schofield and Zakharov 2010, 179) or just *sociodemographics* (e.g., Schofield and Sened 2006), which are not part of the formal but of the empirical model. A key argument for including them in the empirical modeling is again the improvement of model fit.

Mauerer (2020) demonstrates that chooser attributes are directly linked to the intercepts, so their coding determines the intercept values and the information they contain. Our first objective is to investigate further the implications of interpreting intercepts as valences and to outline a parameter identification strategy that matches the definition of average valences in Schofield's Spatial Valence Model, which the conventional identifiability restriction does not. Here, we demonstrate the key points

---

[3]The intercepts are labeled "party-specific coefficients" (e.g., Zur 2021b, 714), but relabeling does not change the major points. We note that all concerns we raise when relying on intercepts to measure valence apply and affect the results presented in Zur (2021a,b) as he claims that his results hold when including socioeconomic voter attributes; see Zur (2021b, 716) and Zur (2021a, 1769).

using the same German vote choice data as in Mauerer (2020). However, as we will lay out, relying on intercepts as a measure of valence still comes with several other drawbacks, such as the assumption that valence aspects are the only choice determinants that remain unobserved or their relative nature, which brings us to our second objective.

We outline a modeling approach to incorporate valence as an additional observable source of voter utility, which is consistent with existing studies within the spatial voting literature that consider valence as a measurable concept (e.g., Adams *et al.* 2011; Buttice and Stone 2012; Franchino and Zucchini 2015; Stone, Maisel, and Maestas 2004; Stone and Simas 2010). We propose to specify valence qualities as attributes that characterize parties (candidates) and present a parameterization that provides one valence effect for each party. We illustrate the benefits of the specification and parametrization strategy drawing on survey questions on candidate character traits in the American National Election Study.

Our third objective is to demonstrate the difference between specifying valence qualities as choice or chooser attributes by revisiting another prominent framework, the *Valence Politics Model of Party Choice* (e.g., Clarke *et al.* 2004, 2009, 2011; Sanders *et al.* 2011; Whiteley *et al.* 2013). Empirical applications of the model take a completely different approach than the dominant one in the spatial literature to quantify the theoretical concept of valence. Here, it is considered as an observable concept that can be measured by survey questions on party leader images or performance evaluations and these variables are specified as chooser attributes. By replicating a typical vote choice model in this research strand that uses the British Election Study, we illustrate the implications for interpretation and model complexity resulting from the chooser-attribute specification.

2. Standard Choice Model and Identifiability Issues

We briefly review the standard choice model and identifiability issues to set the ground. Let $Y_i \in \{1, \ldots, J\}$ contain J alternatives from which decision makers $i \in \{1, \ldots, n\}$ choose. The model incorporates chooser attributes $x_i^T = (x_{i1}, \ldots, x_{iM})$, also referred to as chooser-specific variables, as well as choice attributes $z_{ij}^T = (z_{ij1}, \ldots, z_{ijK})$, known as choice-specific variables, into the utility functions

$$u_{ij} = \beta_{j0} + \sum_{m=1}^{M} x_{im}\beta_{jm} + \sum_{k=1}^{K} z_{ijk}\alpha_k.$$

A logistic response function connects the choice probabilities to the utility functions

$$P(Y_i = j | x_i, z_i) = \frac{\exp(u_{ij})}{\sum\limits_{r=1}^{J} \exp(u_{ir})} = \frac{\exp(\beta_{j0} + x_i^T\beta_j + z_{ij}^T\alpha)}{\sum\limits_{r=1}^{J} \exp(\beta_{r0} + x_i^T\beta_r + z_{ir}^T\alpha)}, j \in \{1, \ldots, J\}, \tag{1}$$

where $\beta_{10}, \ldots, \beta_{J0}$ are alternative-specific intercepts, $\beta_j^T = (\beta_{j1}, \ldots, \beta_{jM})$ are the parameters associated with chooser attributes x_i, and $\alpha^T = (\alpha_1, \ldots, \alpha_K)$ is the coefficient vector related to choice attributes, summarized in $z_i^T = (z_{i1}^T, \ldots, z_{ij}^T)$. Equation (1) gives the model in its general unidentified version. Restrictions or side constraints are required to prevent linear dependency and thus ensure parameter identifiability.

One key restriction refers to the intercepts and chooser-specific covariates x_i, which vary across decision makers but not alternatives. Their invariance across alternatives causes that not all parameters $\beta_j^T = (\beta_{j1}, \ldots, \beta_{jM})$ are identified. The same is true for the intercepts $\beta_{10}, \ldots, \beta_{J0}$. The standard side constraint is to define one alternative as the *reference alternative*. For example, when the first alternative $(Y_i = 1)$ serves as the reference, one sets $\beta_{10} = 0, \beta_1^T = (0, \ldots, 0)$. We will come back to the invariance of chooser attributes across alternatives in Section 4 when we discuss the difference between specifying valence qualities as chooser or choice attributes.

Table 1. (0–1) Coding and effect coding for four-categorical chooser attribute.

(a) (0-1) Coding

$$x_{L(s)} = \begin{cases} 1 & \text{if } L = s \\ \\ 0 & \text{if } L \neq s. \end{cases}$$

(b) Effect Coding

$$x_{L(s)} = \begin{cases} 1 & \text{if } L = s \\ -1 & \text{if } L = S \\ 0 & \text{otherwise.} \end{cases}$$

L	$x_{L(1)}$	$x_{L(2)}$	$x_{L(3)}$
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

L	$x_{L(1)}$	$x_{L(2)}$	$x_{L(3)}$
1	1	0	0
2	0	1	0
3	0	0	1
4	-1	-1	-1

2.1. The Identifiability of Categorical Chooser Attributes

An entirely different form of identifiability must be imposed on categorical chooser attributes that define groups of decision makers. Let $L \in \{1, \ldots, S\}$ denote a chooser attribute with S categories that represent *subpopulations*, socioeconomic groups, or, more generally, attribute levels. The analyst's decision on how such attributes enter the utility functions is crucial for the resulting model properties and parameter interpretation. Since chooser attributes are directly linked to intercepts, the specific identifiability restriction involved has implications for the information the intercepts contain and, therefore, for interpreting intercepts as valences.

The conventional identifiability approach relies on *(0–1) coding*. We will clarify the consequences of the conventional (0–1) coding for interpreting intercepts as valences, which means one has to investigate the link between the identifiability restriction imposed here and the choice probabilities. Another way to deal with categorical chooser attributes is *effect coding*. Under this modeling approach, the dummy variables $x_{L(s)}$ for $s \in \{1, \ldots, S\}$ subpopulations take the values $0, 1, -1$. Table 1 compares (0–1) and effect coding for a four-categorical chooser attribute L. Under both coding schemes, $(S-1)$ dummy variables are sufficient. The last subpopulation is redundant since $L = S$ is implicitly determined by either the vector $(0, \ldots, 0)$ or $(-1, \ldots, -1)$.

Even though the coding schemes yield equivalent models, the interpretation of intercepts strongly depends on the chosen coding. We will present choice models imposing effect coding and discuss its benefits for interpreting intercepts as average valences.

3. Coding Schemes and Intercepts as Valence

The section investigates the link between the coding of categorical chooser attributes and choice probabilities and how it impacts the interpretation of intercepts as valences. We first discuss the difficulty with the conventional (0-1) coding, then present choice models imposing effect coding and outline the resulting model properties. For simplicity, we ignore the choice attributes z_{ij} initially because their coding does not affect the intercepts[4] and illustrate the key points with simplified examples. The section closes by discussing the implications of relying on intercepts to measure valence based on a vote choice model containing choice attributes z_{ij} and chooser attributes x_i.

---

[4]Based on the argument that (0–1) coding for categorical choice attributes z_{ij} causes confounding with the intercepts, effect coding was deemed superior for a long time. Daly, Dekker, and Hess (2016) clarified that intercepts are not affected by the coding of choice attributes and proposed a weighting procedure for effect coding. However, weights have a certain arbitrariness and complicate parameter interpretation and hypothesis testing.

3.1. The (0–1) Coding and the Reference Population

The (0–1) coding approach imposes an identification restriction that involves the need to define a *reference population* not to be confused with the reference alternative among choice alternatives. The analyst selects one subpopulation that serves as a reference to which the parameters are compared. Even though the selection is arbitrary and any subpopulation can be chosen as a reference, the specific choice directly affects the interpretation of intercepts.

The choice probabilities for subpopulation $s \in \{1, \ldots, S\}$ are given by

$$P(Y_i = j | L = s) = \frac{\exp(\beta_{j0} + \beta_{js})}{\sum\limits_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs})}. \tag{2}$$

The choice of a reference population s_0 imposes the restriction $\beta_{js_0} = 0$ for all j. Alternatively, the model can be written with dummy variables $x_{L(s)}$

$$P(Y_i = j | L = s) = \frac{\exp(\beta_{j0} + \sum_{s \in S_0} \beta_{js} x_{L(s)})}{\sum\limits_{r=1}^{J} \exp(\beta_{r0} + \sum_{s \in S_0} \beta_{rs} x_{L(s)})}, \tag{3}$$

where $S_0 = \{1, \ldots, s_0 - 1, s_0 + 1, \ldots, S\}$.

For parameter interpretation, it is helpful to consider the log odds between any two alternatives $j_1, j_2 \in \{1, \ldots, J\}$,

$$\log \left(\frac{P(Y_i = j_1 | L = s)}{P(Y_i = j_2 | L = s)} \right) = \beta_{j_1 0} - \beta_{j_2 0} + \beta_{j_1 s} - \beta_{j_2 s}. \tag{4}$$

When selecting the first choice as the reference alternative ($\beta_{10} = \beta_{1s} = 0$), one obtains

$$\log \left(\frac{P(Y_i = j | L = s)}{P(Y_i = 1 | L = s)} \right) = \beta_{j0} + \beta_{js}, \qquad \frac{P(Y_i = j | L = s)}{P(Y_i = 1 | L = s)} = e^{\beta_{j0}} e^{\beta_{js}}. \tag{5}$$

For the reference population s_0, one obtains

$$\beta_{j0} = \log \left(\frac{P(Y_i = j | L = s_0)}{P(Y_i = 1 | L = s_0)} \right), \qquad e^{\beta_{j0}} = \frac{P(Y_i = j | L = s_0)}{P(Y_i = 1 | L = s_0)}. \tag{6}$$

Thus, there is a direct link between the intercepts and the reference population. The intercepts represent the (log) odds of alternative j compared to alternative 1 *in the reference population*. The crucial point is that the definition of the reference population determines the interpretation of intercepts so that their meaning changes when the arbitrarily selected reference population changes.

For the covariate effects, one obtains

$$\beta_{js} = \log \left(\frac{P(Y_i = j | L = s)}{P(Y_i = 1 | L = s)} \right) - \log \left(\frac{P(Y_i = j | L = s_0)}{P(Y_i = 1 | L = s_0)} \right),$$

$$e^{\beta_{js}} = \frac{P(Y_i = j | L = s) / P(Y_i = 1 | L = s)}{P(Y_i = j | L = s_0) / P(Y_i = 1 | L = s_0)}. \tag{7}$$

Hence, $e^{\beta_{js}}$ give the relative odds (or odds ratios) that compare the odds in subpopulation s to the odds in reference population s_0.

3.1.1. Example

We use the same survey data as in Mauerer (2020) to demonstrate the implications for interpretation.[5] Let Y_i contain vote choices for the German political parties CDU ($Y_i = 1$), SPD ($Y_i = 2$), FDP ($Y_i = 3$),

---

[5]Section B of the Supplementary Material describes the data set.

Table 2. Gender based on (0–1) coding with differing reference populations.

		Reference: male (1 female, 0 male)				Reference: female (1 male, 0 female)		
		β	s.e.	e^{β}		β	s.e.	e^{β}
Intercept CDU	β_{10}	0		1	β_{10}	0		1
Intercept SPD	β_{20}	0.49	0.12	1.64	β_{20}	0.33	0.13	1.39
Intercept FDP	β_{30}	−1.51	0.22	0.22	β_{30}	−2.86	0.42	0.06
Intercept Greens	β_{40}	−0.97	0.18	0.38	β_{40}	−1.13	0.20	0.32
Intercept Left	β_{50}	−1.43	0.21	0.24	β_{50}	−1.22	0.20	0.30
Gender CDU	β_{11}	0		1	β_{11}	0		1
Gender SPD	β_{21}	−0.16	0.18	0.85	β_{21}	0.16	0.18	1.18
Gender FDP	β_{31}	−1.35	0.47	0.26	β_{31}	1.35	0.47	3.87
Gender Greens	β_{41}	−0.16	0.27	0.85	β_{41}	0.16	0.27	1.18
Gender Left	β_{51}	0.21	0.30	1.24	β_{51}	−0.21	0.30	0.81

Note: CDU ($j = 1$) is reference alternative, SPD ($j = 2$), FDP ($j = 3$), Greens ($j = 4$), Left ($j = 5$). *Source*: 1998 German election study. $N = 715$.

Greens ($Y_i = 4$), and Left ($Y_i = 5$). We focus on the dichotomous variable gender $G \in \{1,2\}$ (1 female, 2 male) and estimate the model for different reference populations

$$P(Y_i = j | G = s) = \frac{\exp(\beta_{j0} + \beta_{js})}{\sum\limits_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs})},$$

with the restriction that one of the two gender-specific parameters is set to zero. We specify the party CDU as the reference alternative ($\beta_{10} = \beta_{1s} = 0$). Table 2 reports the estimates. The left part contains the parameters for males as the reference population and the right part for females as the reference population.

To demonstrate that the information in the intercepts depends on the chosen reference population, let us consider the SPD vote ($j = 2$) and focus first on the estimates for males as the reference. The exposed intercept $e^{\beta_{20}} = 1.64$ gives the odds of males voting SPD compared to CDU (see Equation (6)). The product of the exposed intercept and gender estimate $e^{\beta_{20}} e^{\beta_{21}} = 1.64 \times .85 = 1.39$ (see Equation (5)) gives the corresponding odds for females. The gender-specific estimate $e^{\beta_{21}} = .85$ gives the relative odds between the two subpopulations (see Equation (7)).[6] In the reversed coding (females as reference), $e^{\beta_{20}} = 1.39$ gives the effect for females, $e^{\beta_{20}} e^{\beta_{21}} = 1.39 \times 1.18 = 1.64$ the effect for males, and $e^{\beta_{21}} = 1.18 = 1/.85$ the relative odds between the two subpopulations. Thus, even though the arbitrary (0–1) coding leaves the behavioral implications of the model unchanged, the parameters differ when the reference population changes, which has major implications for interpreting intercepts as valences considered in Section 3.3.

---

[6]The probability of the SPD vote in the male population is $P(Y_i = 2 | G = 2) = \exp(\beta_{20})/(1 + \sum_{j=2}^{5} \exp(\beta_{j0})) = .471$, and the one for the CDU is $P(Y_i = 1 | G = 2) = 1/(1 + \sum_{j=2}^{5} \exp(\beta_{j0})) = .287$. Thus, the odds of males voting SPD versus CDU are $e^{\beta_{20}} = .471/.287 = 1.64$. The probability of the SPD vote in the female population is $P(Y_i = 2 | G = 1) = \exp(\beta_{20} + \beta_{21})/(1 + \sum_{j=2}^{5} \exp(\beta_{j0} + \beta_{j1})) = .453$; and the one for the CDU is $P(Y_i = 1 | G = 1) = 1/(1 + \sum_{j=2}^{5} \exp(\beta_{j0} + \beta_{j1})) = .325$. Thus, the odds of females voting SPD versus CDU are $.453/.325 = 1.39$. The relative odds are $e^{\beta_{21}} = (.453/.325)/(.471/.287) = .85$.

3.2. Effect Coding and Average Preferences

The benefit of effect coding is that it removes dependence on a pre-selected arbitrary reference population when interpreting intercepts as valences. It implies an identification restriction such that the resulting parameters relate to average preferences over subpopulations. We first discuss the choice model for S subpopulations and then add covariates.

3.2.1. Choice Model for S Subpopulations

Effect coding imposes for $s \in \{1, \ldots, S\}$ subpopulations the restriction

$$\sum_{s=1}^{S} \beta_{js} = 0 \text{ for all } j.$$

The restriction implicitly uses the *geometric mean* (GM) to average across all S subpopulations. The geometric mean can be considered a natural choice for averaging positive numbers based on product formation and root extraction. The geometric mean across subpopulations given the model holds has the form

$$GM(j) = \left(\prod_{s=1}^{S} P(Y_i = j | L = s) \right)^{1/S} = \left(\prod_{s=1}^{S} \frac{\exp(\beta_{j0} + \beta_{js})}{\sum_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs})} \right)^{1/S}. \tag{8}$$

Since $\prod_{s=1}^{S} e^{\beta_{js}} = 1$, one obtains

$$GM(j) = \gamma \exp(\beta_{j0}), \tag{9}$$

where $\gamma = (\prod_{s=1}^{S} \sum_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs}))^{-1/S}$. $GM(j)$ is the *average preference* (i.e., choice probability) for alternative j, with the geometric mean defining the average. It represents an average across subpopulations, not observations or alternatives. γ is a constant that does not depend on the chooser attribute level s or alternative j.

This allows for a simple interpretation of intercepts, which are given by

$$\beta_{j0} = \log(GM(j)/\gamma), \qquad e^{\beta_{j0}} = GM(j)/\gamma. \tag{10}$$

Thus, $e^{\beta_{j0}}$ represents the preference for alternative j averaged over subpopulations times a constant $(1/\gamma)$. The intercepts indicate whether preferences vary across alternatives when accounting for possible variation across subpopulations. Even when the preferences differ across subpopulations ($\beta_{js} \neq 0$), the average preferences for the alternatives are the same ($GM(1) = \cdots = GM(J) = \gamma$) when the intercepts are zero ($\beta_{10} = \cdots = \beta_{J0} = 0$). Consequently, the intercepts represent average preferences *not explained by subpopulations*. When the first alternative is the reference ($\beta_{10} = \beta_{1s} = 0$), one obtains

$$e^{\beta_{j0}} = \frac{GM(j)}{GM(1)} = \frac{1}{S} \left(\sum_{L=1}^{S} \frac{P(Y_i = j | L = s)}{P(Y_i = 1 | L = s)} \right). \tag{11}$$

The covariate parameters represent *deviations* from the average preferences. Compared to the reference alternative 1, β_{js} give the additive effects on the average log odds and $e^{\beta_{js}}$ the multiplicative effects on the average odds,

$$\beta_{js} = \log \left(\frac{P(Y_i = j | L = s)}{P(Y_i = 1 | L = s)} \right) - \beta_{j0}, \qquad e^{\beta_{js}} = \frac{P(Y_i = j | L = s)}{P(Y_i = 1 | L = s)} / e^{\beta_{j0}}. \tag{12}$$

Table 3. Gender based on effect coding.

		(1 female, −1 male)			(1 male, −1 female)		
		β	s.e.	e^β		β	s.e.	e^β
Intercept CDU	β_{10}	0		1	β_{10}	0		1
Intercept SPD	β_{20}	0.41	0.09	1.51	β_{20}	0.41	0.09	1.51
Intercept FDP	β_{30}	−2.19	0.24	0.11	β_{30}	−2.19	0.24	0.11
Intercept Greens	β_{40}	−1.05	0.13	0.35	β_{40}	−1.05	0.13	0.35
Intercept Left	β_{50}	−1.33	0.15	0.27	β_{50}	−1.33	0.15	0.27
Gender CDU	β_{11}	0		1	β_{11}	0		1
Gender SPD	β_{21}	−0.08	0.09	0.92	β_{21}	0.08	0.09	1.08
Gender FDP	β_{31}	−0.68	0.24	0.51	β_{31}	0.68	0.24	1.97
Gender Greens	β_{41}	−0.08	0.13	0.92	β_{41}	0.08	0.13	1.08
Gender Left	β_{51}	0.11	0.15	1.11	β_{51}	−0.11	0.15	0.90

Note: CDU ($j = 1$) is reference alternative, SPD ($j = 2$), FDP ($j = 3$), Greens ($j = 4$), Left ($j = 5$). *Source*: 1998 German election study. $N = 715$.

The ratio of two parameters gives the relative odds between any two subpopulations $s_1, s_2 \in \{1, \ldots, S\}$

$$\frac{e^{\beta_{js_1}}}{e^{\beta_{js_2}}} = \frac{P(Y_i = j|L = s_1)/P(Y_i = 1|L = s_1)}{P(Y_i = j|L = s_2)/P(Y_i = 1|L = s_2)}. \tag{13}$$

3.2.2. Example
We consider again the variable gender $G \in \{1, 2\}$ (1 female, 2 male) to illustrate the interpretation under effect coding

$$P(Y_i = j|G = s) = \frac{\exp(\beta_{j0} + \beta_{js})}{\sum\limits_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs})},$$

with the restriction $\beta_{j1} + \beta_{j2} = 0$ or $\beta_{j1} = -\beta_{j2}$, respectively. Table 3 shows the estimates based on effect coding for the variable gender in two versions.

Compared to the reference party CDU ($\beta_{10} = \beta_{11} = 0$), the intercepts β_{j0} give the average preferences for party $j \in \{2, 3, 4, 5\}$, averaged over the male and female populations. The intercepts are identical under the two coding versions because the sum of the (log) odds over the two subpopulations is used to calculate them (see Equation (11)). One obtains

$$\beta_{j0} = \frac{1}{2} \left(\log \left(\frac{P(Y_i = j|G = 1)}{P(Y_i = 1|G = 1)} \right) + \log \left(\frac{P(Y_i = j|G = 2)}{P(Y_i = 1|G = 2)} \right) \right),$$

$$e^{\beta_{j0}} = \frac{1}{2} \left(\frac{P(Y_i = j|G = 1)}{P(Y_i = 1|G = 1)} + \frac{P(Y_i = j|G = 2)}{P(Y_i = 1|G = 2)} \right).$$

For example, the exposed SPD intercept $e^{\beta_{20}} = 1.51$ indicates that the average odds of voting SPD is about 1.51 times higher than voting CDU. The gender-specific parameter $e^{\beta_{21}}$ shows how the subpopulations deviate from that average preference (see Equation (12)): $e^{\beta_{21}} = .92$ ($1/e^{\beta_{21}} = 1.08$) suggests that the odds of females (males) voting SPD compared to CDU is .92 (1.08) times the average odds.[7]

---

[7]The (0-1) and effect coding schemes yield identical choice probabilities. We reiterate them to facilitate the illustration of parameter interpretation. The probability of the SPD vote in the male population is .471, and for the CDU vote .287. The probability of the SPD vote in the female population is .453, and for the CDU vote .325. Thus, the average odds of voting

3.2.3. Choice Model for S Subpopulations and Additional Covariates

Next, we add covariates $\boldsymbol{x}_i$ to the utility functions. The restriction for identifiability $\sum_{s=1}^{S} \beta_{js} = 0$ for all j yields the choice probabilities for S subpopulations

$$GM(j, \boldsymbol{x}_i) = \left(\prod_{s=1}^{S} P(Y_i = j | L = s, \boldsymbol{x}_i) \right)^{1/S} = \left(\prod_{s=1}^{S} \frac{\exp(\beta_{j0} + \beta_{js} + \boldsymbol{x}_i^T \boldsymbol{\delta}_j)}{\sum_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs} + \boldsymbol{x}_i^T \boldsymbol{\delta}_r)} \right)^{1/S}$$

$$= \exp(\boldsymbol{x}_i^T \boldsymbol{\delta}_j) \left(\prod_{s=1}^{S} \frac{\exp(\beta_{j0} + \beta_{js})}{\sum_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs} + \boldsymbol{x}_i^T \boldsymbol{\delta}_r)} \right)^{1/S} \tag{14}$$

$$= \gamma(\boldsymbol{x}_i) \exp(\beta_{j0} + \boldsymbol{x}_i^T \boldsymbol{\delta}_j),$$

where $\gamma(\boldsymbol{x}_i) = (\prod_{s=1}^{S} \sum_{r=1}^{J} \exp(\beta_{r0} + \beta_{rs} + \boldsymbol{x}_i^T \boldsymbol{\delta}_r))^{-1/S}$. $GM(j, \boldsymbol{x}_i)$ is the average preference for alternative j given $\boldsymbol{x}_i$, averaged over subpopulations and the average defined by the geometric mean.

For the intercepts, one obtains

$$\beta_{j0} = \log(GM(j, \boldsymbol{0})/\gamma(\boldsymbol{0})), \qquad e^{\beta_{j0}} = GM(j, \boldsymbol{0})/\gamma(\boldsymbol{0}). \tag{15}$$

Thus, $e^{\beta_{j0}}$ is the average preference for alternative j times the constant $1/\gamma(\boldsymbol{0})$, where $\gamma(.)$ is evaluated at $\boldsymbol{0}^T = (0, \ldots, 0)$. However, when the first alternative is the reference, $\beta_{10} = \beta_{1s} = 0, \boldsymbol{\delta}_1^T = (0, \ldots, 0)$, the intercepts give the average odds

$$e^{\beta_{j0}} = \frac{GM(j, \boldsymbol{x}_i)}{GM(1, \boldsymbol{x}_i)} = \frac{1}{S} \left(\sum_{L=1}^{S} \frac{P(Y_i = j | L = s, \boldsymbol{x}_i)}{P(Y_i = 1 | L = s, \boldsymbol{x}_i)} \right). \tag{16}$$

Compared to reference alternative 1, the covariate parameters β_{js} give the additive effects on the average log odds and $e^{\beta_{js}}$ the multiplicative effects on the average odds,

$$\beta_{js} = \log \left(\frac{P(Y_i = j | L = s, \boldsymbol{x}_i)}{P(Y_i = 1 | L = s, \boldsymbol{x}_i)} \right) - \beta_{j0}, \qquad e^{\beta_{js}} = \frac{P(Y_i = j | L = s, \boldsymbol{x}_i)}{P(Y_i = 1 | L = s, \boldsymbol{x}_i)} / e^{\beta_{j0}}, \tag{17}$$

where again $\prod_{s=1}^{S} e^{\beta_{js}} = 1$ holds. The model can also include the term $\boldsymbol{z}_{ij}^T \boldsymbol{\alpha}$, yielding the average $GM(j, \boldsymbol{x}_i, \boldsymbol{z}_{ij})$.

3.3. Intercepts as a Measure of Valence

Next, we demonstrate the benefits of effect coding for interpreting intercepts as average valences as defined in the widely recognized *Spatial Valence Model of Politics* (e.g., Schofield and Sened 2005b, 2006) and discuss the difficulties that remain when equaling the intercepts with valence.

3.3.1. Application: Quantities in Schofield's Spatial Valence Approach

The central quantity is the *valence ranking* where the intercepts are ranked $\beta_{[J]0} \geq \beta_{[J-1]0} \geq \cdots \geq \beta_{[2]0} \geq \beta_{[1]0}$.[8] The lowest valence party $\beta_{[1]0}$ plays a crucial role in quantifying average valences. First, the

---

SPD versus CDU are $\frac{1}{2}((.471/.287) + (.453/.325)) = 1.51 = e^{\beta_{20}}$. The multiplicative effect for females on the average odds is $e^{\beta_{21}} = (.453/.325)/e^{\beta_{20}} = .92$, and the corresponding multiplicative effect for males is $1/e^{\beta_{21}} = (0.471/0.287)/1.51 = 1.08$.

[8]We note that the intercept for the reference alternative is not an actual estimate. We stick to the standard procedure, which includes this intercept in the valence quantities. Our notation differs from Schofield's work, where λ_j denotes the intercepts and p the number of parties.

Table 4. Empirical quantities in the spatial valence approach.

	(0–1) Coding	Effect coding
Model 1: all covariates		
Ranking	$\beta_{10} \geq \beta_{20} \geq \beta_{50} \geq \beta_{30} \geq \beta_{40}$	$\beta_{20} \geq \beta_{10} \geq \beta_{50} \geq \beta_{40} \geq \beta_{30}$
	$0 \geq -0.06 \geq -0.37 \geq -1.73 \geq -2.44$	$0.18 \geq 0 \geq -1.39 \geq -2.34 \geq -3.10$
$\lambda_{av(1)}$	-0.54	-0.89
Λ	1.90	2.21
Model 2: all covariates, gender reversed		
Ranking	$\beta_{10} \geq \beta_{20} \geq \beta_{50} \geq \beta_{40} \geq \beta_{30}$	$\beta_{20} \geq \beta_{10} \geq \beta_{50} \geq \beta_{40} \geq \beta_{30}$
	$0 \geq -0.25 \geq -0.30 \geq -2.94 \geq -3.26$	$0.18 \geq 0 \geq -1.39 \geq -2.34 \geq -3.10$
$\lambda_{av(1)}$	-0.87	-0.89
Λ	2.39	2.21
Model 3: covariate region omitted		
Ranking	$\beta_{20} \geq \beta_{10} \geq \beta_{30} \geq \beta_{50} \geq \beta_{40}$	$\beta_{20} \geq \beta_{10} \geq \beta_{50} \geq \beta_{40} \geq \beta_{30}$
	$0.05 \geq 0 \geq -1.21 \geq -1.22 \geq -1.46$	$0.23 \geq 0 \geq -1.63 \geq -1.93 \geq -2.85$
$\lambda_{av(1)}$	-0.60	-0.83
Λ	0.86	2.02

Note: Vote choice models containing spatial proximities and voter demographics. Section B of the Supplementary Material reports estimation tables. Party numbers: 1 (CDU, reference alternative, vote share: .30), 2 (SPD, vote share: .46), 3 (FDP, vote share: .04), 4 (Greens, vote share: .11), 5 (Left, vote share: .08). $\lambda_{av(1)}$ is the average valence other than lowest ranked party; Λ is the valence difference for lowest valence party (see text). *Source*: 1998 German election study. $N = 715$.

average valence of parties without the lowest ranked is calculated, $\lambda_{av(1)} = \left[1/(J-1)\right] \sum_{j=2}^{J} \beta_{[j]0}$. Then, the valence difference between this average and the lowest valence party, $\Lambda = \lambda_{av(1)} - \beta_{[1]0}$. We stick to the German election data and now consider all covariates: six standard voter demographics $\boldsymbol{x}_i$ (dichotomous variables union membership, working class, Catholic denomination, gender, region, and the quantitative variable age centered around the sample mean) and four spatial proximities $\boldsymbol{z}_{ij}$ (voter–party proximities on the issues of immigration, nuclear energy, European integration, and the Left-Right dimension).[9]

Table 4 compares Schofield's valence quantities for three vote choice models. Model 1 includes all covariates. In Model 2, we reversed the coding for one variable only, gender. Model 3 omits the variable region to demonstrate the dependence of the intercepts on the included covariates. The left part uses (0–1) coding, and the right part effect coding, based on the average $GM(j, \boldsymbol{x}_i, \boldsymbol{z}_{ij})$. The party CDU is the reference alternative. Both coding schemes use identical probabilities when modeling choice behavior but yield different valence quantities because the information in the intercepts depends on the coding.

Under (0–1) coding, the intercepts give the relative preferences of the voter segment $\boldsymbol{x}_i^T = \boldsymbol{0}^T = (0, \ldots, 0)$. Here, arbitrary coding decisions determine the composition of a particular electorate for which valence effects are calculated. Thus, the valence quantities in Model 1 refer to the subpopulation where all voter demographics take the value of 0, that is, average-aged male voters that do not belong to the working class, are no union members, do not have a catholic denomination, and are based in former East Germany. When the coding of only one variable changes, a different electorate is considered, which changes the information in the intercepts and, consequently, all intercept values and valence quantities. Whereas under Model 1, the Greens (β_{40}) result as the lowest ranked party, it is the FDP (β_{30}) under the

---

[9]See Section B of the Supplementary Material for details.

reversed coding of gender in Model 2, yielding different valence quantities; $\lambda_{av(1)}$ reduces from -0.54 to -0.87, and Λ increases from 1.90 to 2.39.

As a result, depending on arbitrary coding decisions, many different subpopulations can be constructed so that the researcher can calculate and present many different valence effects under (0–1) coding. And, the range of possible compositions of particular electorates gets larger the more voter demographics the model contains, which increases the model fit, and the more values these covariates can take.

Under effect coding, the intercepts represent relative average preferences (see Equation (16)). Consequently, the specific coding does not affect the valence quantities, yielding stable results. The SPD (β_{20}) is the highest valence party and the lowest ranked party is the FDP (β_{30}), $\lambda_{av(1)} = -0.89$ and $\Lambda = 2.21$ under different coding versions.

Model 3 demonstrates how the ceteris paribus condition affects the valence quantities. As in any regression-based modeling and independent from the coding schemes, parameter interpretation depends on the included covariates. When omitting the variable region, a different electorate for which valences are calculated is defined, yielding different intercept values and, thus, valence effects. Whereas under effect coding, the ranking remains stable and the average valences and differences only slightly change, the quantities show a huge variation under (0–1) coding. The SPD (β_{20}) instead of the CDU (β_{10}) results as the highest valence party, the Greens (β_{40}) as the lowest valence party, the FDP (β_{30}) is ranked in the middle and Λ changes from 2.39 to 0.86.

If one wants to stick to interpreting intercepts as valences, effect coding is the better option. It better matches the definition of valence as an "*average* perception, among the electorate" (Schofield 2005, 348, italics added). Whereas (0–1) coding does not involve average preferences, effect coding does and the researcher avoids making inferences for one particular subpopulation only. However, two fundamental model properties make relying on intercepts to study valence aspects challenging and questionable.

First, the interpretation is not reference-free. Since the intercept of the reference alternative is set to zero under the standard side constraint to ensure identifiability, the reliance on intercepts only allows an interpretation relative to the chosen reference party (candidate). Second, if covariates contributing to explaining choices are not considered, this information enters the intercepts and increases the amount of unobserved utility sources. Since the intercepts reflect the importance of *all unobserved utility sources*, relying on intercepts implies that only valence qualities remain unobserved and the analyst succeeds in measuring all other choice determinants. If intercepts are assumed to represent valence, it should be made more clear what is implied. Then, given a particular set of covariates, valence comprises all the remaining unexplained (because unobserved or unobservable) factors that determine the choice.

4. Valence as an Observable Source of Utility

This section deals with specification and parameterization issues that arise when valence is considered as an observable source of voter utility. That is, when researchers model valence by covariates suitable for measuring valence qualities. We first outline a modeling strategy that overcomes the drawbacks of the intercepts. We propose a covariate specification that considers valence as a choice attribute and an effect parameterization that removes dependence on a reference alternative and provides one valence effect for each party (candidate). Then, we discuss the difference between specifying valences as choice or chooser attributes.

4.1. Valence as a Choice Attribute

We propose to specify valence qualities as choice-specific variables z_{ij} measuring attributes that characterize parties (candidates) to incorporate valence as an observable source of voter utility. To arrive at a valence effect for each party, we estimate parameters $\alpha_j^T = (\alpha_{11}, \ldots, \alpha_{JK})$ that are specific to each alternative j. In contrast to the standard generic specification in Equation (1), which constrains the parameters $\boldsymbol{\alpha}$ to be the same for all alternatives ($\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_J := \boldsymbol{\alpha}$), the alternative-wise specification

relaxes the assumption that decision makers assign the same weight to a choice attribute independent from which alternative they evaluate. The alternative-wise specification has been considered in the study of spatial voting before for voter–party issue proximities (e.g., Mauerer 2016; Mauerer, Thurner, and Debus 2015). Here, we apply it to the parameterization of valence attributes to study how such qualities contribute to the utility each party provides voters

$$u_{ij} = \beta_{j0} + \sum_{m=1}^{M} x_{im}\beta_{jm} + \sum_{k=1}^{K} z_{ijk}\alpha_{jk} = \beta_{j0} + \boldsymbol{x}_i^T\boldsymbol{\beta}_j + \boldsymbol{z}_{ij}^T\boldsymbol{\alpha}_j.$$

The parameters $\boldsymbol{\alpha}_j$ do not depend on a reference alternative. Thus, the effects are the same on all odds such that the relative odds remain the same independent of a reference alternative.

4.1.1. Application: Valence Qualities as Candidate Character Traits

We draw on the 2016 U.S. presidential election study and model the choice between the Democratic nominee Hillary Clinton and the Republican opponent Donald Trump to demonstrate the benefits of the specification strategy in studying the impact of valence qualities. We operationalize valence qualities by survey questions on candidate personality traits, thereby tapping into the dimension of valence as a character quality (Adams *et al.* 2011; Stone *et al.* 2004; Stone and Simas 2010). Voters assessed the candidates on six traits (strong leadership, really cares, knowledgeable, honest, speaks mind, and even-tempered). We generated an additive index for each candidate as an overall character assessment.[10]

Table 5 reports three models. Model 1 includes four spatial proximities with generic parameters for simplicity and typical voter demographics. Model 2 adds character traits in generic specification. Model 3 specifies the character traits with alternative-wise parameters. Likelihood Ratio tests ($\chi^2(1) = 256.77$) indicate that character traits are highly significant and considerably improve model fit. Model 3 with alternative-wise parameters for character traits fits significantly better than Model 2 with one generic parameter ($\chi^2(1) = 4.29$). While in Model 1 all spatial proximities are highly significant, they greatly lose explanatory power, their effects are much weaker, and only two proximities (Liberal-Conservative and Defense) remain significant when including character traits whose effects are dominant in Models 2 and 3. The alternative-wise estimates in Model 3 suggest that character traits have a larger impact on the preference for the Democratic than the Republican candidate, ceteris paribus.

The empirical application demonstrates that operationalizing valence qualities by survey questions on candidate character traits and specifying them as choice attributes $\boldsymbol{z}_{ij}$ with alternative-wise parameters $\boldsymbol{\alpha}_j$ is a promising alternate strategy to intercepts to study how valence aspects affect vote choices. Including valence qualities as additional observable utility sources considerably improves the model performance so that less relevant information enters the intercepts. We note the ceteris paribus condition also applies here. The parameters reflect the association between the character traits and the dependent variable, given spatial proximities and voter demographics.

4.2. Valence as Chooser Attributes

The *Valence Politics Model of Party Choice* is the main competing approach to the spatial voting framework. Empirical applications also apply vote choice models containing spatial proximities and voter demographics. Instead of defining the intercepts as valence, this research strand usually considers multiple valence qualities, measured by survey questions on party leader images, performance evaluations, or problem-solving capacities and specified as chooser-specific variables (e.g., Clarke *et al.* 2009; Sanders *et al.* 2011; Whiteley *et al.* 2013). We revisit this modeling approach by demonstrating the difference between specifying valence qualities as chooser or choice attributes.

---

[10]The traits are measured on five-point scales from "not well at all" to "extremely well." Section C of the Supplementary Material provides details on the empirical application and operationalization.

Table 5. Vote choice models with valence qualities as candidate character traits.

	Model 1		Model 2		Model 3	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
Voter demographics						
Republican intercept	−2.26	0.41	−0.66	0.50	−2.71	1.16
Republican age	−0.00	0.01	0.01	0.01	0.01	0.01
Republican female	−0.07	0.12	0.20	0.18	0.18	0.18
Republican Black	−1.27	0.33	−0.52	0.33	−0.69	0.36
Republican Latino	−0.52	0.26	−0.42	0.37	−0.40	0.36
Spatial proximities						
Lib-Cons	0.73	0.07	0.58	0.12	0.59	0.12
Spending	0.46	0.07	0.17	0.11	0.18	0.11
Defense	0.60	0.07	0.24	0.10	0.25	0.10
Health care	0.19	0.05	−0.04	0.07	−0.03	0.07
Valence						
Candidate character traits			2.61	0.26		
Candidate character traits: Republican					2.22	0.31
Candidate character traits: Democrat					2.94	0.31
Log-Lik.	−243.40		−115.01		−112.86	

Note: Democratic Candidate is the reference alternative. Categorical voter attributes in effect coding, age centered around the sample mean.
Source: 2016 American National Election Study. $N = 1,847$.

The specification of valence as chooser attributes x_i, which vary across choosers i but not alternatives j, requires J variables $x_i^{(j)}$ and $J \times (J-1)$ parameters $\beta_j^{(j)}$ for $p = 1$ continuous or binary valence quality. Considering multiple valence features quickly leads to complex models and parameter inflation. For example, when analyzing $J = 3$ alternatives and considering three valence features where two are continuous and one is four-categorical, one has to deal with 15 variables and 30 parameters: 3×2 variables with $3 \times (3-1) \times 2$ parameters for the two continuous valence features and 3×3 variables with $3 \times (3-1) \times (4-1)$ parameters for the four-categorical valence feature. Moreover, the resulting parameters depend on a reference alternative, and only a subset of the parameters is of direct interest in studying the impact of valence.

For parameter interpretation, it is helpful to consider the log odds between any two alternatives $j_1, j_2 \in \{1, \dots, J\}$,

$$\log\left(\frac{P(Y_i = j_1 | x_i^{(j)})}{P(Y_i = j_2 | x_i^{(j)})} \right) = (\beta_{j_1 0} - \beta_{j_2 0}) + x_i^{(j)} (\boldsymbol{\beta}_{j_1}^{(j)} - \boldsymbol{\beta}_{j_2}^{(j)}). \tag{18}$$

Compared to reference alternative 1, $\beta_{10} = 0$, $\boldsymbol{\beta}_1^{(j)} = (0, \dots, 0)$, the parameters $e^{\boldsymbol{\beta}_j^{(j)}}$ give the relative odds when $x_i^{(j)}$ increases by one unit

$$e^{\boldsymbol{\beta}_j^{(j)}} = \frac{P(Y_i = j | x_i^{(j)} + 1)/P(Y_i = 1 | x_i^{(j)} + 1)}{P(Y_i = j | x_i^{(j)})/P(Y_i = 1 | x_i^{(j)})}. \tag{19}$$

Thus, the valence effects depend on the reference alternative, which makes their interpretation demanding.

The specification of valence as choice attributes z_{ij}, which take different values for each alternative j, is much more parsimonious and one can estimate J parameters α_j that are reference-free and of direct interest to study valence effects. For $p = 1$ continuous or binary valence quality, only one variable z_{ij} is necessary and J parameters α_j can be estimated. When analyzing $J = 3$ alternatives and three valence features (two continuous and one four-categorical), one only has to deal with 5 variables and 17 parameters: 2 variables with 3×2 parameters for the two continuous valence features and 3 variables with $3 \times (4 - 1)$ parameters for the four-categorical valence feature.

The interpretation of valence effects is alternative-specific and independent from the reference alternative:

$$e^{\alpha_j} = \frac{P(Y_i = j | z_{ij} + 1)/P(Y_i = l | z_{ij} + 1)}{P(Y_i = j | z_i)/P(Y_i = l | z_i)}. \tag{20}$$

The parameters e^{α_j} give the relative odds when z_{ij} increases by one unit. Next, we provide an empirical example to demonstrate the difference between both specifications.

4.2.1. Application: Valence Qualities as Party Leader Images

We draw on a simplified version of the model in Sanders *et al.* (2011) and consider voting for the three major British parties Labour (Lab, $j = 1$), Conservatives (Cons, $j = 2$), and the Liberal Democrats (LD, $j = 3$) in the 2010 British election. Valence is operationalized by several features, such as party leader images, assessments of party competence in different areas, or judgments about what party can best handle the most important issue facing Britain today. We focus on party leader images[11] and control for spatial proximities and voter demographics.[12] Table 6 only reports the parameters for party leader images (Section D of the Supplementary Material contains full estimation tables). The upper part gives the estimates for party leader images as chooser attributes based on different reference alternatives, and the lower part for party leader images as a choice attribute with alternative-wise effects.

When specifying party leader images as chooser attributes, three variables, one for each party $(x_i^{(1)}, x_i^{(2)}, x_i^{(3)})$, are necessary. Each variable $x_i^{(j)}$ is associated with two identified parameters $\beta_j^{(j)}$, yielding the utility functions $u_{ij} = \beta_{j0} + x_i^{(1)}\beta_j^{(1)} + x_i^{(2)}\beta_j^{(2)} + x_i^{(3)}\beta_j^{(3)}$. Thus, one obtains six parameters to present valence effects, which are interpreted relative to a reference alternative, for example, to Labour, by setting $\beta_{10} = \beta_1^{(1)} = \beta_1^{(2)} = \beta_1^{(3)} = 0$. Take the Labour leader image $x_i^{(1)}$. When Labour is the reference, both identified parameters are of direct interest to evaluate the party's valence effect. The parameter $\beta_2^{(1)} = -0.72$ gives the difference to the Conservatives and $\beta_3^{(1)} = -0.49$ the one to the Liberal Democrats, suggesting that an increase in Labour's valence harms the Conservative vote more than the Liberal Democrats vote. When one selects the Conservatives or the Liberal Democrats as the reference, only one parameter in each case is of direct interest ($\beta_1^{(1)} = 0.72$ when the Conservative vote is the reference and $\beta_1^{(1)} = 0.49$ when the Liberal Democrat vote is the reference) because the remaining parameter gives the difference between Liberal Democrats and Conservatives, which is only of indirect interest when evaluating the valence of Labour.

When we are interested in how the Conservative leader image $x_i^{(2)}$ impacts voting, the model with the Conservatives as reference contains the relevant information: $\beta_1^{(2)} = -0.90$ and $\beta_3^{(2)} = -1.13$, suggesting that an increase in valence for Conservative has a larger negative impact on the Liberal Democrats than Labour. The same applies to the Liberal Democrats leader image $x_i^{(3)}$. Thus, the researcher must estimate

---

[11]Survey question: "Using a scale that runs from 0 to 10, where 0 means strongly dislike and 10 means strongly like, how do you feel about [name of party leader]?"

[12]The spatial proximities are on the issues of crime and taxes. The voter demographics are the dichotomous variables union membership, working class, gender, homeowner (in effect coding), and the quantitative variables income (standardized annual household income) and age (centered around the sample mean). See Section D of the Supplementary Material for details.

Table 6. Vote choice model with valence qualities as party leader images.

	(a) Valence as chooser attributes $x_i^{(j)}$ with different reference parties							
	Reference			Reference			Reference	
	Lab ($j=1$)			Cons ($j=2$)			LD ($j=3$)	
		β	s.e.		β	s.e.		β	s.e.
Lab leader image $x_i^{(1)}$	$\beta_1^{(1)}$	0		$\beta_1^{(1)}$	0.72	0.06	$\beta_1^{(1)}$	0.49	0.05
	$\beta_2^{(1)}$	−0.72	0.06	$\beta_2^{(1)}$	0		$\beta_2^{(1)}$	−0.22	0.05
	$\beta_3^{(1)}$	−0.49	0.05	$\beta_3^{(1)}$	0.22	0.05	$\beta_3^{(1)}$	0	
Cons leader image $x_i^{(2)}$	$\beta_1^{(2)}$	0		$\beta_1^{(2)}$	−0.90	0.09	$\beta_1^{(2)}$	0.23	0.06
	$\beta_2^{(2)}$	0.90	0.09	$\beta_2^{(2)}$	0		$\beta_2^{(2)}$	1.13	0.09
	$\beta_3^{(2)}$	−0.23	0.06	$\beta_3^{(2)}$	−1.13	0.09	$\beta_3^{(2)}$	0	
LD leader image $x_i^{(3)}$	$\beta_1^{(3)}$	0		$\beta_1^{(3)}$	0.21	0.08	$\beta_1^{(3)}$	−0.73	0.07
	$\beta_2^{(3)}$	−0.21	0.08	$\beta_2^{(3)}$	0		$\beta_2^{(3)}$	−0.94	0.08
	$\beta_3^{(3)}$	0.73	0.07	$\beta_3^{(3)}$	0.94	0.08	$\beta_3^{(3)}$	0	
	(b) Valence as choice attribute z_{ij}								
	α_j	s.e.							
Lab leader image	0.58	0.05							
Cons leader image	0.92	0.06							
LD leader image	0.71	0.05							

Note: Vote choice models containing spatial proximities and voter demographics. Section D of the Supplementary Material reports full estimation tables. *Source*: 2010 British Election Study. $N = 1,262$.

the model with different reference alternatives to detect all relevant valence effects. But these valence effects are not reference-free and always only allow a relative interpretation.

Under the proposed approach, which specifies valence qualities as a choice attribute z_{ij} with alternative-wise parameters α_j, the resulting utility functions are $u_{ij} = \beta_{0j} + z_{ij}\alpha_j$. One obtains one parameter for each party ($\alpha_1, \alpha_2, \alpha_3$) that contains the relevant information. The alternative-wise parameters indicate that party leader images have the largest impact on the preference for the Conservatives and the smallest one for Labour, ceteris paribus, which is hardly seen when using the chooser-attribute approach.

5. Concluding Remarks

This contribution provides the statistical fundamentals to advance the empirical modeling of valence, a crucial concept in the study of public choice. We outline the effect coding scheme for chooser attributes that facilitates the interpretation of intercepts as valence because it frees researchers from making inferences for a specific reference population only and, therefore, matches the definition of average valences as introduced by Schofield's widely applied Spatial Valence Model. However, relying on intercepts still comes with severe drawbacks that are independent of the coding schemes. The most critical point is probably that of approaching valence as an immeasurable concept. Defining the intercepts as valences implies that all unobserved choice-determining factors equal valence aspects. Consequently, when researchers want to stick to intercepts as valences, they should aim to capture as many non-valence-related factors by covariates to keep unobserved utility sources low and provide model fit measures to evaluate that. Then, effect coding presents a solution when the data do not contain suitable variables on valence qualities.

We also propose a covariate specification and effect parameterization strategy to incorporate valence aspects as an additional observable source of voter utility and, therefore, to overcome the drawbacks of the intercepts as valences and discuss different specification strategies. Our proposed modeling approach requires variables that are able to measure the theoretical concept of valence. Our empirical applications, where we operationalize valence by candidate character traits and party leader images (measured by like–dislike scores), are promising and yield insightful results. We hope this contribution inspires researchers to capture valence qualities through observable variables to keep the unobserved variable effect low, which is one major goal of empirical modeling.

Future research should focus on what variables are best to operationalize valence qualities and carefully consider them already in the data collection. For example, the literature on affective polarization is not in agreement about whether party leader like–dislike scores capture the general affect toward party leaders that might not be related to their qualities (e.g., Reiljan *et al.* 2023).

References

Adams, J. 1999. "Multiparty Spatial Competition with Probabilistic Voting." *Public Choice* 99 (3): 259–274.

Adams, J., and S. Merrill, III. 1999a. "Modeling Party Strategies and Policy Representation in Multiparty Elections: Why Are Strategies So Extreme?" *American Journal of Political Science* 43 (3): 765–791.

Adams, J., and S. Merrill, III. 1999b. "Party Policy Equilibrium for Alternative Spatial Voting Models: An Application to the Norwegian Storting." *European Journal of Political Research* 36 (2): 235–255.

Adams, J., and S. Merrill, III. 2000. "Spatial Models of Candidate Competition and the 1988 French Presidential Election: Are Presidential Candidates Vote-Maximizers?" *The Journal of Politics* 62 (3): 729–756.

Adams, J., S. Merrill, III, and B. Grofman. 2005. *A Unified Theory of Party Competition: A Cross-National Analysis Integrating Spatial and Behavioral Factors*. New York: Cambridge University Press.

Adams, J., S. Merrill, III, E. N. Simas, and W. J. Stone. 2011. "When Candidates Value Good Character: A Spatial Model with Applications to Congressional Elections." *The Journal of Politics* 73 (1): 17–30.

Adams, J., S. Merrill, and R. Zur. 2020. "The Spatial Voting Model." In *The SAGE Handbook of Research Methods in Political Science and International Relations*, edited by L. Curini and R. Franzese, 205–223. London: Sage.

Alvarez, R. M., and J. Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42 (1): 55–96.

Burden, B. C. 1997. "Deterministic and Probabilistic Voting Models." *American Journal of Political Science* 41 (4): 1150–1169.

Buttice, M. K., and W. J. Stone. 2012. "Candidates Matter: Policy and Quality Differences in Congressional Elections." *The Journal of Politics* 74 (3): 870–887.

Clarke, H. D., D. Sanders, M. C. Stewart, and P. Whiteley. 2004. *Political Choice in Britain*. Oxford: Oxford University Press.

Clarke, H. D., D. Sanders, M. C. Stewart, and P. Whiteley. 2009. *Performance Politics and the British Voter*. New York: Cambridge University Press.

Clarke, H. D., D. Sanders, M. C. Stewart, and P. Whiteley. 2011. "Valence Politics and Electoral Choice in Britain, 2010." *Journal of Elections, Public Opinion and Parties* 21 (2): 237–253.

Coughlin, P. J. 1992. *Probabilistic Voting Theory*. New York: Cambridge University Press.

Daly, A., T. Dekker, and S. Hess. 2016. "Dummy Coding vs Effects Coding for Categorical Variables in Choice Models: Clarifications and Extensions." *Journal of Choice Modelling* 21: 36–41.

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

Enelow, J. M., and M. J. Hinich. 1989. "A General Probabilistic Spatial Theory of Elections." *Public Choice* 61 (2): 101–113.

Evrenk, H. 2019. "Valence Politics." In *The Oxford Handbook of Public Choice*, edited by R. D. Congleton, B. Grofman, and S. Voigt, 266–291. New York: Oxford University Press.

Franchino, F., and F. Zucchini. 2015. "Voting in a Multi-Dimensional Space: A Conjoint Analysis Employing Valence and Ideology Attributes of Candidates." *Political Science Research and Methods* 3 (2): 221–241.

Green, J., and W. Jennings. 2017. "Valence." In *The SAGE Handbook of Electoral Behavior*, edited by K. Arzheimer, J. Evans, and M. S. Lewis-Beck, 538–560. Los Angeles: Sage.

Kedar, O. 2005. "How Diffusion of Power in Parliaments Affects Voter Choice." *Political Analysis* 13 (4): 410–429.

Magyar, A., S. Wagner, and R. Zur. 2023. "Party Strategies: Valence versus Position." In *The Routledge Handbook of Political Parties*, edited by R. D. Congleton, B. Grofman, and S. Voigt, 199–210. London: Routledge.

Mauerer, I. 2016. "A Party-Varying Model of Issue Voting. A Cross-National Study." PhD Thesis, University of Munich (LMU).

Mauerer, I. 2020. "The Neglected Role and Variability of Party Intercepts in the Spatial Valence Approach." *Political Analysis* 28 (3): 303–317.

Mauerer, I., P. W. Thurner, and M. Debus. 2015. "Under Which Conditions Do Parties Attract Voters' Reactions to Issues? Party-Varying Issue Voting in German Elections 1987–2009." *West European Politics* 38 (6): 1251–1273.

Mauerer, I., and G. Tutz. 2023a. "Heterogeneity in General Multinomial Choice Models." *Statistical Methods & Applications* 23: 129–148.

Mauerer, I., and G. Tutz. 2023b. "Replication Data for: Vote Choices and Valence: Intercepts and Alternate Specifications." Harvard Dataverse, V1. https://doi.org/10.7910/DVN/SKWTGS

McFadden, D. L. 1974. "Conditional Logit Analysis of Qualitative Choice Behaviour." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–142. New York: Academic Press.

Merrill, S. , III, and J. Adams. 2001. "Computing Nash Equilibria in Probabilistic, Multiparty Spatial Models with Nonpolicy Components." *Political Analysis* 9 (4): 347–361.

Reiljan, A., D. Garzia, F. Ferreira da Silva, and A. H. Trechsel. 2023. "Patterns of Affective Polarization toward Parties and Leaders across the Democratic World." *American Political Science Review*: 1–17. https://doi.org/10.1017/S0003055423000485

Sanders, D., H. D. Clarke, M. C. Stewart, and P. Whiteley. 2011. "Downs, Stokes and the Dynamics of Electoral Choice." *British Journal of Political Science* 41 (2): 287–314.

Schofield, N. 2005. "A Valence Model of Political Competition in Britain: 1992–1997." *Electoral Studies* 24 (3): 347–370.

Schofield, N., and I. Sened. 2005a. "Modeling the Interaction of Parties, Activists and Voters: Why Is the Political Center So Empty?" *European Journal of Political Research* 44 (3): 355–390.

Schofield, N., and I. Sened. 2005b. "Multiparty Competition in Israel, 1988–96." *British Journal of Political Science* 35 (4): 635–663.

Schofield, N., and I. Sened. 2006. *Multiparty Democracy: Elections and Legislative Politics*. Political Economy of Institutions and Decisions. New York: Cambridge University Press.

Schofield, N., and A. Zakharov. 2010. "A Stochastic Model of the 2007 Russian Duma Election." *Public Choice* 142 (1): 177–194.

Stoetzer, L. F., and S. Zittlau. 2015. "Multidimensional Spatial Voting with Non-Separable Preferences." *Political Analysis* 23 (3): 415–428.

Stokes, D. E. 1963. "Spatial Models of Party Competition." *The American Political Science Review* 57 (2): 368–377.

Stone, W., and E. N. Simas. 2010. "Candidate Valence and Ideological Positions in U.S. House Elections." *American Journal of Political Science* 54 (2): 371–388.

Stone, W. J., L. S. Maisel, and C. D. Maestas. 2004. "Quality Counts: Extending the Strategic Politician Model of Incumbent Deterrence." *American Journal of Political Science* 48 (3): 479–495.

Train, K. E. 2009. *Discrete Choice Methods with Simulation*, 2nd Edn. New York: Cambridge University Press.

Whiteley, P., H. D. Clarke, D. Sanders, and M. C. Stewart. 2013. *Affluence, Austerity and Electoral Change in Britain*. Cambridge: Cambridge University Press.

Zur, R. 2021a. "The Multidimensional Disadvantages of Centrist Parties in Western Europe." *Political Behavior* 43 (4): 1755–1777.

Zur, R. 2021b. "Stuck in the Middle: Ideology, Valence and the Electoral Failures of Centrist Parties." *British Journal of Political Science* 51 (2): 706–723.