# PA

# Using Conjoint Experiments to Analyze Election Outcomes: The Essential Role of the Average Marginal Component Effect

## Kirk Bansak[1], Jens Hainmueller[2], Daniel J. Hopkins[3] and Teppei Yamamoto[4]

[1] Assistant Professor, Department of Political Science, University of California, Berkeley, 210 Social Sciences Building, Berkeley, CA 94720-1950, USA. E-mail: kbansak@berkeley.edu
[2] Professor, Department of Political Science, Stanford University, 616 Serra Street, Encina Hall West, Room 100, Stanford, CA 94305-6044, USA. E-mail: jhain@stanford.edu
[3] Professor, Department of Political Science, Perelman Center for Political Science and Economics, University of Pennsylvania, 133 S. 36th Street, Philadelphia, PA 19104, USA. E-mail: danhop@sas.upenn.edu
[4] Associate Professor, Department of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. E-mail: teppei@mit.edu, URL: http://web.mit.edu/teppei/www

## Abstract

Political scientists have increasingly deployed conjoint survey experiments to understand multidimensional choices in various settings. In this paper, we show that the average marginal component effect (AMCE) constitutes an aggregation of individual-level preferences that is meaningful both theoretically and empirically. First, extending previous results to allow for arbitrary randomization distributions, we show how the AMCE represents a summary of voters' multidimensional preferences that combines directionality and intensity according to a probabilistic generalization of the Borda rule. We demonstrate why incorporating both the directionality and intensity of multi-attribute preferences is essential for analyzing real-world elections, in which *ceteris paribus* comparisons almost never occur. Second, and in further empirical support of this point, we show how this aggregation translates directly into a primary quantity of interest to election scholars: the effect of a change in an attribute on a candidate's or party's expected vote share. These properties hold irrespective of the heterogeneity, strength, or interactivity of voters' preferences and regardless of how votes are aggregated into seats. Finally, we propose, formalize, and evaluate the feasibility of using conjoint data to estimate alternative quantities of interest to electoral studies, including the effect of an attribute on the probability of winning.

*Keywords:* survey experiments, elections, conjoint, Average Marginal Component Effect, voter preferences

## 1 Introduction

Elections are a defining feature of representative democracies (Schumpeter 1950), and extensive research within political science has sought to understand voting behavior (Achen *et al.* 2017; Campbell *et al.* 1960; Key 1966). Frequently, these research questions take the form of how a specific attribute of a party or candidate influences the share of votes they win in an election. For example, how much will a candidate's vote share change if her gender is female instead of male? That critical question is the topic of Schwarz and Coppock (2022). Meanwhile, Auerbach and Thachil (2018) investigate the impact of co-ethnicity and education on choices for slum presidents in the slums of two Indian cities. In both cases, researchers sought to quantify how changes in a candidate's multidimensional profile would influence citizens' choice of candidates.

Those studies are two examples among many. How voters choose between candidates or parties who vary on multiple dimensions is a central question of political behavior. In recent years, conjoint experiments have emerged as a tool for answering such questions. With a carefully designed conjoint experiment, election scholars can study voters' multidimensional preferences by unbiasedly estimating the causal effects of multiple candidate attributes on

hypothetical vote choices. At the core of this approach is a causal quantity of interest, the average marginal component effect (AMCE), which represents how much the probability of choosing a candidate would change on average if one candidate attribute switched levels (Hainmueller, Hopkins, and Yamamoto 2014). The introduction of this approach sparked numerous applications, many focused on electoral politics (see Bansak *et al.* 2021b for a review). It has also prompted the development of statistical tools (De la Cuesta, Egami, and Imai 2021; Egami and Imai 2019; Hanretty, Lauderdale, and Vivyan 2020).

There are many situations important to social scientists which seem ripe for analysis through conjoint designs, as they ask individuals to rank or choose between bundles comprised of multiple attributes. Voting behavior certainly has these elements, but so do choices about which immigrants to admit, which policies to adopt, and various other topics (e.g., Adida, Lo, and Platas 2019; Bansak, Bechtel, and Margalit 2021; Mummolo and Nall 2017). To date, however, the empirical adoption of conjoint designs has outpaced theoretical discussions of what quantities conjoint designs can—and cannot—recover. As a result, some scholars have critiqued practices for analyzing and interpreting conjoint experiments (e.g., Abramson, Koçak, and Magazinnik 2021; Ganter 2020; Leeper, Hobolt, and Tilley 2020).

This paper's central contribution is to demonstrate the utility of the AMCE as a tool for analyzing elections and answering key research questions. To do so, it couples formal analyses of the AMCE and possible alternatives with a review of recent empirical studies of voting. As we explain, analyses using the AMCE have a straightforward, meaningful interpretation that is consistent with the most common quantity of interest in recent studies—changes in vote shares. Given the explosion of research using conjoint designs to study elections—and given recent critiques of the AMCE—this demonstration is critical both in consolidating existing knowledge generated from conjoint designs and in planning future research. We also explore alternative electoral quantities of interest that can and cannot necessarily be reliably estimated from conjoint data, highlighting both opportunities and limits in what can be recovered from the conjoint experimental design.

Specifically, we illuminate the AMCE's formal and conceptual underpinnings to show its central role in the conjoint analysis of elections. In unpacking the AMCE, we clarify how it should and should not be interpreted, highlighting how it aggregates individual-level preferences into a quantity of interest that is essential for studying election outcomes. As initially shown by Abramson *et al.* (2021) and generalized here, the AMCE captures both the direction and intensity of preferences (also see Ganter 2020). This property of incorporating both directionality and intensity is inherent in the definition of averaging, and so it also holds true for numerous other common estimands, most notably the average treatment effect (ATE). As we show, this property is crucial for explaining choices or outcomes in multidimensional settings. That is, it is precisely this property that endows the AMCE with a straightforward, politically meaningful interpretation as an attribute's average causal effect on a candidate's or party's expected *vote share*. As will be detailed, this expectation is taken with respect to a target election distribution of interest, which incorporates the distribution of both voters and candidate/party attributes. Importantly, this equivalence between AMCEs and effects on vote shares holds regardless of the structure of voter preferences.

Furthermore, through a literature review of 82 articles in four electoral politics journals, we demonstrate that vote shares and their individual-level analogs are by far the most common quantities of interest in empirical electoral research. The AMCE thus provides a fitting tool for researchers using conjoints to study the effects of candidate or party attributes on vote shares. In addition to identifying a core quantity of interest involving the central outcome in election research, AMCEs are also easily estimated without arbitrary functional form assumptions.

Certainly, the fact that the AMCE recovers a key quantity of interest to election scholars does not mean that it is the only appropriate estimand in conjoint election analyses. Thus, in the second half of this paper, we employ the same framework Hainmueller *et al.* (2014) developed

for the AMCE to examine how paired-profile, forced-choice conjoint data could be used to study other quantities of interest. In doing so, we highlight the important broader point that research designs and quantities of interest should go hand in hand. As a corollary of this point, one cannot and should not expect the AMCE to provide answers to all research questions, nor data from a conjoint design to allow for recovery of any quantity of interest. We demonstrate these points by defining and distinguishing between two main alternative quantities of interest, one of which can be feasibly recovered using typically sized conjoint data, whereas the other is likely best suited to an alternative study design. The first involves the effect of an attribute on a candidate's probability of winning an election. The second involves the fraction of voters preferring a specific attribute.

Our analysis uncovers several challenges, especially for the second alternative estimand. First, due to the nonlinearity in majority rule, estimation of either quantity using conjoint data requires a model-based approximation of a high-dimensional conditional expectation function. This contrasts with the AMCE, which can be estimated without such modeling assumptions via a design-based approach motivated purely by randomization. Second, and relatedly, the two estimands differ substantially in their difficulty of estimation using typical conjoint data, with estimation of the second being essentially infeasible. The reason for this disparity lies in precisely how majority rule is applied in each case, as discussed below and in the Supplementary Material. We find initial evidence of feasibility for the probability of winning through our exploration of possible estimators and simulations, although we note that additional tailoring of the conjoint design specifically to this quantity of interest would further improve the estimation. In contrast, our analysis also reveals the infeasibility of using typically sized conjoint data to estimate the second estimand, the fraction of voters preferring an attribute, due to the sparsity of individual-level data. Hence, alternative non-conjoint designs and data collection methods would be necessary for this quantity of interest.

Third, we also highlight how this second alternative is less informative about attributes' importance for voting behavior in multi-attribute contexts. A simple example (which we revisit below) shows why. Consider an attribute on which voters hold largely homogeneous preferences but that is trivial from the standpoint of vote choice, such as candidates' handedness (right-handed vs. left-handed). The vast majority of voters are right-handed, so assuming right-handed voters prefer right-handed candidates, the fraction of voters preferring this attribute *all else equal* is very large. However, this overwhelming preference does little to influence the election outcome, since actual candidates differ across other, more important dimensions. The AMCE would reflect that, as it would be near zero.

Finally, we provide practical guidance for applied researchers employing conjoint experiments and suggest possible paths for future research. Overall, this paper contributes to the growing methodological literature on conjoint experiments by connecting the most commonly used causal estimand—the AMCE—to a foundational theory of individual preferences and showing its interpretability as a key quantity of interest to electoral scholars. The paper also highlights other quantities of interest that can (or cannot) be effectively investigated using conjoint data, thereby reducing confusion among applied users.

## 2 Formal and Conceptual Underpinnings of the AMCE

To unpack the AMCE's formal and conceptual underpinnings, we first present a general framework for analyzing voter preferences in multi-attribute elections, in which candidates are characterized by multiple observed attributes. We then use the framework to show how the AMCE relates to individual preferences, highlighting the important role of relative preference intensity. A key implication is that the AMCE identifies a central quantity of interest in electoral research: an attribute's average effect on expected vote shares.[1]

---

1 As will be explained in greater detail later, the expectation in the AMCE is taken with respect to a target election distribution of interest, which incorporates the distribution of both voters and candidate/party attributes. More generally, the same

**Table 1.** Candidate preference rankings for three voter types.

| Profile | Attribute | | | Profile rank: | | |
|---|---|---|---|---|---|---|
| number | A | B | C | Type 1 | Type 2 | Type 3 |
| 1 | 1 | 1 | 1 | 1 | 2 | 6 |
| 2 | 1 | 1 | 0 | 2 | 6 | 2 |
| 3 | 1 | 0 | 1 | 3 | 4 | 8 |
| 4 | 1 | 0 | 0 | 4 | 8 | 4 |
| 5 | 0 | 1 | 1 | 5 | 1 | 5 |
| 6 | 0 | 1 | 0 | 6 | 5 | 1 |
| 7 | 0 | 0 | 1 | 7 | 3 | 7 |
| 8 | 0 | 0 | 0 | 8 | 7 | 3 |

*Note*: This table shows the preference ranking for three voter types over candidate profiles defined by binary attributes A, B, and C.

## 2.1 Formalizing Preferences in Multi-Attribute Elections

Consider a paired-profile, forced-choice conjoint experiment, where each respondent $i \in \{1, \ldots, N\}$ completes $K$ tasks in which the respondent casts hypothetical votes between two candidates varying across $L$ attributes. Each of the $L$ attributes takes on $D_l$ discrete levels, respectively, such that $l \in \{1, \ldots, L\}$. One can view this design as a simulation of a two-candidate election in which citizens vote for one of two candidates varying across $L$ observed attributes.

As an illustration, consider a toy example in which candidates are characterized by three binary attributes (i.e., $L = 3, D_1 = D_2 = D_3 = 2$). We label these attributes *A*, *B*, and *C*, respectively, and denote their levels by 0 and 1, such that $A \in \{0, 1\}, B \in \{0, 1\}$, and $C \in \{0, 1\}$. These values then fully characterize the election's candidates. We use $[abc]$ to denote a candidate (or conjoint profile) whose values on these attributes are such that $A = a$, $B = b$, and $C = c$. There are $2^3 = 8$ possible unique candidates, that is, [000], [001], [010], [011], [100], [101], [110], and [111]. More generally, there are $\prod_{l=1}^{L} D_l$ possible unique candidates.

Given a choice where alternatives are characterized by multiple attributes, a natural formalization of individual preferences is to consider a preference ordering over the full set of possible unique attribute combinations. Namely, we define each voter's preferences to be binary relations over the set of possible unique candidates. To simplify exposition, we assume that each voter has a strict preference ordering over all $\prod_{l=1}^{L} D_l$ unique candidates. For example, consider the "Type 1" voter in Table 1. In the table, the eight possible candidates (defined in columns 2–4) are ordered from top to bottom according to the Type 1 voter's preference ranking (column 5). This preference can also be represented using standard decision-theoretic notation, such that $[111] > [110] > [101] > [100] > [011] > [010] > [001] > [000]$.

## 2.2 Defining the AMCE

Since elections are a means of preference aggregation, electoral researchers may ask how one can learn about collective decisions from individual preferences expressed through conjoint experiments. It is fruitful to begin with several desirable criteria for this objective. First, it would be valuable to have an aggregate measure that captures the multidimensionality of the typical electoral choice, in which voters choose between candidates differing across many dimensions simultaneously. Second, the measure should map onto a meaningful empirical phenomenon of interest, such that electoral researchers can make causal or predictive inferences about elections

framework can be employed to show that the AMCE identifies an attribute's average effect on choice probabilities in multi-attribute decision-making problems beyond elections.

using it. Finally, the measure should be empirically tractable, in the sense that researchers can use observed data from conjoint experiments to estimate it with sufficient precision and ideally without strong modeling assumptions.

The AMCE is one quantity of interest in electoral studies that meets all three criteria. First, the AMCE aggregates preference orderings over all possible profiles in a systematic manner that accounts for the multidimensional nature of the electoral decision problem by incorporating the directionality *and* intensity of preferences. Second, the AMCE directly represents the causal effect of a particular attribute on a candidate's expected vote share, which a literature review reveals as the most prominent quantity of interest for electoral scholars. Third, identification and unbiased estimation of the AMCE can proceed under a limited set of assumptions and via straightforward nonparametric methods (Hainmueller *et al.* 2014). The remainder of this section highlights each feature.

To define the AMCE under the current setup, let $Y_i([abc],[a'b'c']) \in \{0,1\}$ denote voter $i$'s potential outcome given a paired forced-choice contest between profiles $[abc]$ and $[a'b'c']$. The potential outcome equals 1 if respondent $i$ would choose the first candidate (i.e., $[abc]$) given the choice task, which would occur if and only if $[abc] > [a'b'c']$ for that respondent, and $Y_i([abc],[a'b'c']) = 0$ otherwise (i.e., if $[a'b'c'] > [abc]$). Then, the AMCE for attribute $A$ is defined as the expected difference between the potential outcomes for all paired contests where the first candidate's attribute $A$ equals 1 and the potential outcomes for all contests where the first candidate's attribute $A$ equals 0, given a known, prespecified distribution of the other attributes. So, without loss of generality, the AMCE for attribute $A$ is

$$AMCE_A \equiv \mathbb{E}\big[Y_i([1BC],[A'B'C']) - Y_i([0BC],[A'B'C'])\big], \tag{1}$$

where the expectation is defined over the joint distribution of candidate attributes from which all attributes other than $A$ for the first candidate (i.e., $B, C, A', B'$, and $C'$) are drawn, *and* the sampling distribution for the $N$ respondents from the target population of voters. The AMCEs for attributes $B$ and $C$ are defined analogously, and the definition extends to conjoint designs with more attributes, attributes with more than two levels, non-forced-choice outcomes, and tasks with more than two profiles, with appropriate notational changes.

A few remarks are useful to highlight in particular what the AMCE does *not* represent, and thereby avoid confusion among researchers. First, note that the relevant contrast for the AMCE is between attribute $A = 1$ for the first profile and $A = 0$ for the *same profile*, both against another profile randomly drawn from the prespecified distribution. Suppose attribute $A$ is gender, such that $A = 1$ means a female candidate and $A = 0$ means a male candidate. Then, $AMCE_A$ compares the probability of a female candidate profile chosen against another randomly generated profile (whether male or female) to the probability of a male profile chosen against a similarly generated profile. That is, the AMCE asks how much better or worse a randomly selected candidate would fare if gender switches from male to female. Specifically, the AMCE is *not* the probability of a female candidate being chosen against a randomly generated male candidate. This difference has been a point of confusion in some applied work.

Second, the AMCE aggregates individual preferences with respect to two dimensions: across attributes and voters. Specifically, the AMCE employs *averaging* of individual preferences both across the distributions of possible candidates and voters in the target population. This is in contrast to other means of preference aggregation. For instance, Abramson *et al.* (2021) show that as a result of the way in which the AMCE aggregates preferences, the AMCE does not reflect a simple majority preference (i.e., a positive AMCE does not necessarily mean a majority of voters prefer the attribute level in question). This is similar to how a positive sign of a standard ATE also does not necessarily mean that a majority of units have positive individual-level treatment effects.

However, while Abramson *et al.* (2021) conclude from this that the AMCE is largely uninformative with respect to questions of interest to political scientists, in the following sections, we highlight how the AMCE's properties actually make it an invaluable tool for studying voter choice and election outcomes.

## 2.3 The AMCE and Preference Intensity

The first of the AMCE's desirable properties for studying voter choice and election outcomes is that it captures the multidimensionality of the conjoint choice task. It does so by incorporating both the direction and intensity of preferences about individual attributes through aggregating the probabilities of profiles winning pairwise comparisons.

*2.3.1 The Simplest Case: The Independent Uniform Attribute Distribution.* In the simplest setup with attributes distributed independently and uniformly, the AMCE's aggregation of winning probabilities reduces to the simple averaging of the profile *ranks*. Continuing with the three-attribute example, let $r_i(a, b, c) \in \{1, \dots, 8\}$ represent the rank of profile $[abc]$ for voter $i$. Then, consider a voter's average rank for the profiles that contain a particular attribute level such that the average rank of $A = a$ for voter $i$ is defined as $S_i^A(a) \equiv \frac{1}{4} \sum_{b \in \{0,1\}} \sum_{c \in \{0,1\}} r_i(a, b, c)$. Comparing a voter's average ranks with respect to different levels of an attribute (e.g., $S_i^A(1)$ vs. $S_i^A(0)$) captures the directionality and intensity of her preferences with respect to the attribute.

For example, consider the Type 1 voter in Table 1. Intuitively, this voter strongly favors $A = 1$ to $A = 0$ because profiles containing $A = 1$ are more highly ranked than profiles containing $A = 0$ irrespective of other attributes. For attribute $B$, the voter favors profiles with $B = 1$ to those with $B = 0$, but only if the profiles are not better in terms of $A$. As for $C$, the voter generally likes profiles with $C = 1$ better than those with $C = 0$, but $C$'s value only influences the final ranking when the profiles are tied on other attributes. Thus, we can summarize these preferences as an intense preference for $A = 1$ over $A = 0$, a moderate preference for $B = 1$ over $B = 0$, and a mild preference for $C = 1$ over $C = 0$.

Considering the average profile ranks across different attribute levels captures these intuitions. For illustration, Table 2 provides the Type 1 voter's average ranks. The average rank for a Type 1 voter $i$ of $A = 1$, $S_i^A(1)$, is equal to 2.5, whereas the average rank of $A = 0$ is 6.5. This implies that the voter prefers $A = 1$ to $A = 0$. Similarly, $S_i^B(1) = 3.5$ and $S_i^B(0) = 5.5$, implying $B = 1$ is preferred to $B = 0$. Likewise, $S_i^C(1) = 4$ and $S_i^C(0) = 5$, so that $C = 1$ is preferred to $C = 0$. The relative values of the rank means provide a natural metric for the *intensity* of the voter's preferences: for attributes $A$, $B$, and $C$, the rank means are 2.5 versus 6.5 (intense preference), 3.5 versus 5.5 (moderate preference), and 4 versus 5 (mild preference), respectively. Incorporating these differences in preference intensity is key for capturing the attributes' importance for the resulting vote choices in contests between multidimensional profiles.

The AMCE can, in fact, be shown to be directly related to these average rankings. Using a difference between the average ranks as a measure of the extent to which a voter prefers a particular attribute level over the other (e.g., $S_i^A(1) - S_i^A(0)$), one can further quantify the aggregate preference for $A = 1$ over $A = 0$ across all voters by taking the average value of $S_i^A(1) - S_i^A(0)$ across $i \in \{1, \dots, N\}$, which we denote by $\bar{S}^A(1) - \bar{S}^A(0)$. Under a uniform joint distribution of all the attributes, the AMCE for $A = 1$ relative to $A = 0$ is proportional to $\bar{S}^A(1) - \bar{S}^A(0)$, such that $AMCE_A \propto \bar{S}^A(1) - \bar{S}^A(0)$ as defined in Equation (1). Seen this way, it is clear that the AMCE represents an aggregation of individual preferences that explicitly accounts for intensity: $S_i^A(1) - S_i^A(0)$ represents an individual voter's relative preference intensity for $A = 1$ over $A = 0$, and $\bar{S}^A(1) - \bar{S}^A(0)$ averages this across voters.

---

*Kirk Bansak et al.* | Political Analysis

**Table 2.** Average ranks by attribute.

| Attribute | Value | Average rank: Type 1 | Type 2 | Type 3 |
|---|---|---|---|---|
| A | 0 | 6.5 | 4.0 | 4.0 |
| A | 1 | 2.5 | 5.0 | 5.0 |
| B | 0 | 5.5 | 5.5 | 5.5 |
| B | 1 | 3.5 | 3.5 | 3.5 |
| C | 0 | 5.0 | 6.5 | 2.5 |
| C | 1 | 4.0 | 2.5 | 6.5 |

*Notes:* This table shows the average ranks for candidate profiles with and without a given attribute for three voter types. Type 1 voters have an intense preference for A, moderate preference for B, and mild preference for C. Type 2 voters have a mild preference for not A, moderate preference for B, and intense preference for C. Type 3 voters have a mild preference for not A, moderate preference for B, and intense preference for not C.

2.3.2 *The General Case: Arbitrary Attribute Distributions.* The above result largely reproduces a main result in Abramson *et al.* (2021), who prove the proportionality between the AMCE and differences in Borda scores in the special case that attributes are uniformly and independently distributed. Here, we further expand upon this by contributing the proof of a more general result. Namely, we show that the AMCE of an attribute is proportional to what we call the *expected Borda score* (EBS) of that attribute under an arbitrary, potentially nonuniform and nonindependent randomization distribution of attributes, provided that profiles are independently drawn within pairs. While the Borda score of a profile equals the number of times the profile is chosen against all possible alternative profiles, the EBS represents the *expected* number of times a profile is chosen against $Q$ independently randomly drawn profiles, thereby generalizing the Borda score to choice settings that are probabilistic rather than deterministic.

The generalization to an arbitrary randomization distribution is important for several reasons. First, theoretically, our probabilistic generalization builds a bridge from the original deterministic, choice-theoretic result of Abramson *et al.* (2021) to the statistical analysis of randomized experiments and treatment effects, the latter of which has been the predominant framework for analyzing conjoint survey designs in the recent literature (e.g., De la Cuesta *et al.* 2021; Hainmueller *et al.* 2014; Leeper *et al.* 2020). Second, recent methodological research has formally highlighted the importance of carefully choosing the randomization distribution of attributes and made proposals (and recommendations on the process) for deviating from uniform, independent distributions (De la Cuesta *et al.* 2021; Ganter 2020). Third, partially prompted by these recommendations, a number of applied researchers adopt conjoint designs that utilize nonuniform, nonindependent attribute distributions for the purpose of realism and external validity (e.g., Huff and Kertzer 2018; Leeper and Robison 2020).

To consider the general case of a paired conjoint with an arbitrary attribute distribution, we denote the $j$th profile in respondent $i$'s $k$th task by $X_{ijk} \equiv [X_{ijkl}, \tilde{X}_{ijk}] \in \mathcal{X}, j \in \{1, 2\}, k \in \{1, \ldots, K\}, l \in \{1, \ldots, L\}$, where $X_{ijkl}$ represents the attribute of interest $l$ and $\tilde{X}_{ijk}$ the collection of the remaining $L - 1$ attributes. Then, the general potential outcome with respect to profile pair $[x_1, x_2]$ for respondent $i$ can be defined as $Y_i(x_1, x_2) = Y_i(x, \tilde{x}_1, x_2)$, where $x_1 = [x, \tilde{x}_1]$, representing the respondent's choice in a conjoint task comparing $x_1$ against $x_2$. Under the standard consistency assumption, the observed outcome for respondent $i$'s $k$th task can be written as $Y_{ik} = Y_i(X_{i1k}, X_{i2k})$.

Next, we introduce preference relations for the profiles. Let $Q \equiv |\mathcal{X}|$, the number of unique profiles. For each respondent $i$, we index elements of $\mathcal{X}$ in the order of their preference, such that

$\mathcal{X} = \{x_{(1i)}, \ldots, x_{(mi)}, \ldots, x_{(Qi)}\}$, where $x_{(mi)} > x_{(m'i)}$ if and only if $m < m'$. Then, assume the potential outcome to be a deterministic reflection of the respondent's preference ordering,[2] such that

$$
Y_i(x_{(mi)}, x_{(m'i)}) = \begin{cases} 1, & \text{if } x_{(mi)} > x_{(m'i)} \iff m < m', \\ 0, & \text{if } x_{(mi)} \leq x_{(m'i)} \iff m \geq m'. \end{cases} \tag{2}
$$

Under this setup, the AMCE for the population of $N$ respondents with respect to the $l$th attribute can be written as

$$
AMCE_l(t_1, t_0; p) \equiv \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_p \left[ Y_i(t_1, \tilde{X}_{i1k}, X_{i2k}) - Y_i(t_0, \tilde{X}_{i1k}, X_{i2k}) \right]
$$

$$
= \frac{1}{N} \sum_{i=1}^{N} \sum_{\tilde{x}_1 \in \tilde{\mathcal{X}}} \sum_{x_2 \in \mathcal{X}} \{Y_i(t_1, \tilde{x}_1, x_2) - Y_i(t_0, \tilde{x}_1, x_2)\} p(\tilde{x}_1, x_2), \tag{3}
$$

where $\tilde{\mathcal{X}}$ is the support of $\tilde{X}_{ijk}$ and $p(\tilde{x}_1, x_2) = \Pr(\tilde{X}_{i1k} = \tilde{x}_1, X_{i2k} = x_2)$, a known joint distribution of $\tilde{X}_{i1k}$ and $X_{i2k}$.[3]

As discussed initially by Hainmueller *et al.* (2014) and more extensively by De la Cuesta *et al.* (2021), the AMCE is defined with respect to the attribute distribution $p$. In theory, $p$ can be set to any target distribution of interest, for example, the real-world distribution of the attributes.[4] In practice, it is common to set $p(\tilde{x}_1, x_2)$ to the randomization distribution of the attributes actually used in the experiment, which we denote by $p^*(\tilde{x}_1, x_2)$. This allows the AMCE to be identified by the observed difference in means with respect to $X_{ijkl} = t_1$ versus $X_{ijkl} = t_0$. Although this is perhaps the most commonly used specification in applied research, it is *not* required for the results in the rest of this section to hold. In fact, the only assumption we need about the target distribution is the following independence assumption about profiles.

**Assumption 1** (Independently Generated Profiles)**.**

$$
\tilde{X}_{i1k} \perp\!\!\!\perp X_{i2k}, \quad \text{such that} \quad p(\tilde{x}_1, x_2) = p(\tilde{x}_1) p(x_2).
$$

That is, we assume that profiles in each paired comparison are independently drawn from each other. (Note that attributes within each profile are allowed to be arbitrarily correlated.) Assumption 1 is satisfied in a vast majority of conjoint applications we are aware of, where researchers independently draw profiles for their actual experiment and use the actual randomization distribution $p^*$ for their (implicit) definition of the AMCE (i.e., $p = p^*$).[5]

---

2 This assumption might be contradicted by observed data under a couple of scenarios. First, note that Equation (2) assumes that all respondents choose the first profile in a trivial choice task between two identical profiles, which may be violated by actual observed responses. If this occurs, one can simply interchange the profile index between the two profiles to resolve the technical violation. The subsequent analysis can then proceed as usual unless the no profile-order effect assumption (Hainmueller *et al.* 2014) is violated. Second, observed choices may violate the basic properties of preference relations, such as symmetry and transitivity. In practice, neither of these violations is likely to be a problem in applied settings, since the probability of drawing profiles meeting these conditions (e.g., two identical profiles in the same task, multiple instances of an identical comparison switching sides for the same respondent, etc.) is typically negligibly small.

3 Here, we assume the $N$ respondents in the actual data constitute the population of interest. However, one can also consider the respondents to be a random sample from the infinite population of interest (as in Hainmueller *et al.* 2014). In that case, the sample average over $i \in \{1, \ldots, N\}$ in Equation (3) should be replaced with the expectation over the respondent sampling distribution, and the subsequent derivation goes through without other modifications. In Section 2.4, we refer to this population distribution of interest as the *target voter distribution* in the context of an election and denote it by $\mathcal{V}$.

4 In Section 2.4, we define this as the *target attribute distribution* of an election of interest and denote it by $\mathcal{A}$.

5 One somewhat common scenario where this assumption is violated is when profile pairs are generated via rejection sampling where exactly tied comparisons are excluded. However, the between-profile dependence introduced by such designs is typically negligibly minor since exact ties occur with very small probabilities.

---

To generalize the proportionality between the AMCE and the Borda score for any attribute distribution satisfying Assumption 1, we introduce the following quantity.

DEFINITION 1 (Expected Borda Score) $g_i(x_{(mi)}; p) \equiv Q \Pr(X_{i2k} \prec x_{(mi)})$, where $\Pr(\cdot)$ is defined with respect to the target attribute distribution $p$.

In words, the EBS of profile $x_{(mi)}$ represents the expected number of times the profile is chosen by respondent $i$ against $Q$ profiles randomly drawn from the target profile distribution $p$. It is straightforward to show that the original Borda score, $b_i(x_{(mi)}) \equiv \sum_{m' \neq m} \mathbf{1}\{x_{(mi)} > x_{(m'i)}\} = Q - m$, is a special case of $g_i(x_{(mi)}; p)$ when $\Pr(X_{i2k} = x_{(m)}) = 1/Q \; \forall \; m \in \mathcal{X}$ (i.e., the discrete uniform). This is because

$$g_i(x_{(mi)}; p) = Q \sum_{m' \in \{1,\dots,M\}, m' > m} \frac{1}{Q} = Q - m = b_i(x_{(mi)}).$$

Now, we are ready to state our main proposition for this section.

**Proposition 1.** *Suppose that the target attribute distribution for the AMCE satisfies Assumption 1. Then, the difference between the EBS of attribute $t_1$ and of attribute $t_0$ is proportional to the AMCE of $t_1$ against $t_0$.*

A proof is provided in Section A of the Supplementary Material. Proposition 1 implies that the AMCE can generally be seen as an aggregation of individual preferences according to a probabilistic generalization of the Borda rule. That is, the AMCE is a summary of voters' multidimensional preferences that incorporates both ordering and intensity.

2.3.3  *Importance of Preference Intensity for Analyzing Real-World Elections.*  Why is the quantification of preference intensity, in addition to binary preference relations, essential in the context of studying voter choice and election outcomes? The answer lies in the multidimensionality of the problem. In real-world elections where votes are cast for candidates characterized by multiple attributes, candidates in any particular matchup are likely to differ across multiple attributes. In such multidimensional choice settings where *ceteris paribus* comparisons almost never occur, the intensity of preferences plays a crucial role in determining voters' selections.[6] As an example, consider the handedness example introduced earlier—that is, assume that a voter would *all else equal* prefer a candidate who shares the same handedness as she does. Because the vast majority of people are right-handed, there would be a pronounced *ceteris paribus* majority preference for right-handedness over left-handedness. Indeed, given the overwhelming extent to which the world is right-handed, the size of this majority preference for right-handedness (i.e., the fraction of voters preferring this attribute all else equal) might even exceed that for any other attributes evaluated (e.g., age, previous experience, and policy positions).

This result, of course, obscures our understanding of real-world voter choice, in which candidates differ across many different attributes and voters need to choose candidates based not on their *ceteris paribus* preferences with respect to individual attributes but rather the balance of their preference intensity across all attributes. If one considers voters' preference intensity via the average rank framework above (or its generalization), voters' preference for a right-handed candidate would be trivial, as the average rank of right-handed candidates (or their EBS) would be only slightly above left-handed candidates'. This reflects real-world voting behavior: voters would ignore the handedness information when presented with multidimensional candidate profiles and make their choices as a function of the attributes they deemed relevant. By taking preference

---

6  Here, by *ceteris paribus*, we mean a contrast between two candidates who are exactly the same on all attributes except for one. For more on *ceteris paribus* comparisons, see the discussion in Section D.2 of the Supplementary Material.

intensity into account, the AMCE captures this real-world behavior and its implications for voter choice and election outcomes; in this example, the AMCE for right-handedness would be near zero.

## 2.4 The AMCE as the Effect on Vote Shares

The AMCE's second desirable property is that it represents a quantity of broad interest to empirical elections scholars: the average causal effect of an attribute on *vote shares*, with the expectation taken with respect to a target election distribution as defined below. Specifically, in a forced-choice conjoint experiment, the AMCE equals the expected difference in the choice probability of a candidate with a treatment attribute level (e.g., gender = female) and that of a candidate with the baseline level of the attribute (e.g., male) in an election with the same number of candidates (i.e., two in a typical paired conjoint). Importantly, this property holds regardless of the structure of individual voters' preferences. AMCEs identify vote shares irrespective of whether the intensity of voters' preferences about individual attributes is homogeneous or heterogeneous, or whether there are interactions between candidate attributes in shaping voter preferences. The property also holds independently of the electoral formulae used for aggregating votes into seats, making the AMCE a useful quantity for both majoritarian and proportional representation elections.

Taken literally, the AMCE is only well defined in the context of a conjoint experiment. That is, estimating $AMCE_A$ corresponds to asking the following question: if we randomly draw a female candidate and her opponent from the set of possible candidates, how much more likely is the female candidate to win the paired conjoint task, compared to a male candidate randomly drawn in the same manner on average? This quantity is of interest to many applied researchers, since the conjoint choice tasks themselves can be a robust and reliable measure of attitudes and opinions (e.g., Bansak *et al.* 2021a; Jenke *et al.* 2021). Nonetheless, a crucial question for many elections scholars is whether the AMCE is also informative about elections and the aggregation of individual preferences into outcomes through such elections. Does the AMCE map onto any electoral quantity of interest?

Here, we show that the AMCE equals a quantity summarizing the causal effect of a candidate attribute on *vote shares* in an election matching the specifications of the conjoint. By vote share, we mean the percentage of votes cast for a candidate in an election. An attribute's AMCE in a conjoint experiment resembling an election can be interpreted as the average causal effect of the attribute on the vote share of a randomly selected candidate with that attribute as opposed to the baseline level of the attribute. Thus, the AMCE is interpretable in terms that are directly relevant for the study of elections.

To make our point formally, we define a *target election* to be represented by a pair $\langle \mathcal{A}, \mathcal{V} \rangle$, where $\mathcal{A}$ and $\mathcal{V}$ refer to the *target attribute distribution* and the *target voter distribution*, respectively. The attribute distribution $\mathcal{A}$ is a probability measure on the combinations of candidate attributes, whereas the voter distribution $\mathcal{V}$ is a probability measure on individual preferences over the attribute combinations in $\mathcal{A}$'s support. For example, consider Table 1, which represents the toy example of an election with candidates with three binary attributes and three voter types. The attribute distribution is a probability mass function over all possible attribute combinations or profiles (i.e., rows in the table's left half). For instance, it could be a uniform categorical distribution over the eight possible profiles. The voter distribution, in turn, is a probability mass function over the three voter types (i.e., columns in the table's right half), for example, Pr(Type 1) = .3, Pr(Type 2) = .4, and Pr(Type 3) = .3. Note that "target" in these definitions indicates that these distributions usually correspond to populations of voters and candidates that are of interest to the researcher, such as those resembling candidates and voters in a real-world election.

Now, consider a conjoint experiment on a representative sample of respondents randomly drawn from the target voter distribution $\mathcal{V}$. Furthermore, suppose profiles are randomly generated according to the target attribute distribution $\mathcal{A}$. It follows that the AMCE of each attribute

under the design can be interpreted as the attribute's average effect on candidate vote shares in the target election $\langle \mathcal{A}, \mathcal{V} \rangle$. The general result is stated in the following proposition.

> **Proposition 2** (Identification of the Expected Difference in Vote Shares). *Consider a J-profile conjoint experiment in which respondents are a simple random sample of size N drawn from $\mathcal{V}$. Then, for any N, J, $\mathcal{A}$, and $\mathcal{V}$, the AMCE for attribute A = a (versus the baseline level A = $a_0$) given the randomization distribution $\mathcal{A}$ identifies the difference in the expected vote share of a candidate with A = a and one with A = $a_0$ in the target election $\langle \mathcal{A}, \mathcal{V} \rangle$ with J candidates.*

The proposition follows trivially from the definitions of the AMCE and the target election, noting that the expected value of the conjoint potential outcome for a profile set (e.g., $\mathbb{E}[Y_i([abc], [a'b'c'])]$) equals the proportion of the votes cast for the first candidate in the corresponding target election. (A formal proof is omitted.) Of note, Proposition 2 holds more generally for designs with $J > 2$ profiles per choice task. This means that the AMCE allows researchers to use a *J*-profile conjoint design to study vote shares in *J*-candidate single-vote elections.

Proposition 2 implies that election scholars can use appropriately designed conjoint survey experiments to predict candidates' vote shares in elections and interpret the resulting AMCEs as the causal effects of candidate attributes on predicted vote shares. For example, an AMCE of −0.02 for a male candidate versus a female candidate indicates that male gender has an average causal effect of −2 percentage points on candidates' vote shares in elections resembling the experiment's design: on average, a randomly selected male candidate would earn 2 fewer percentage points of the total vote share than a female.

But how common are vote shares as a quantity of interest in empirical election research? We conducted a literature review including all articles on voting in four journals that commonly publish studies on voting behavior between 2015 and 2019.[7] Of the 82 articles reviewed, 87% include either aggregate vote shares or their individual-level analogs as one key outcome. The small minority of studies that do not feature outcomes related to vote shares instead have outcomes such as the probability of a candidate or party winning. Thus, not only does the AMCE recover a politically meaningful quantity, but also it recovers a quantity that has been the primary quantity of interest even in most non-conjoint studies in recent years.

## 3 Alternative Quantities of Interest

The AMCE has desirable properties and a straightforward vote share interpretation. However, there are certainly other election-related quantities that may be of interest to researchers implementing paired-profile forced-choice conjoint designs. For example, Ganter (2020) has recently proposed an alternative estimand, the "average component preference" (ACP), designed to explore patterns of preferences from forced-choice conjoint data. Similar to the AMCE, the ACP also reflects patterns of preferences in terms of both directionality and intensity.[8] In other applications, however, researchers may wish to disentangle preference direction and intensity.

---

7 Section B of the Supplementary Material describes our procedure and sample.
8 Indeed, Ganter (2020) defines preferences directly in terms of the averages of the potential outcomes for profile pairs involving an attribute contrast of interest: The ACP for $A = a$ versus $A = a'$ (where $a \neq a'$) can be written using our notation as $\mathbb{E}[Y_i([aBC], [a'B'C'])] - 0.5$, which Ganter defines as "respondents' preferences" themselves (p. 3). Importantly, the expectation $\mathbb{E}$ here is taken with respect to both the target attribute and profile distributions $\mathcal{A}$ and $\mathcal{V}$ combined, implying that the quantity represents the aggregation of individual preferences incorporating both directionality and intensity. In that sense, Ganter's (implicit) formalization of individual preferences is similar to the AMCE, even though the author emphasizes differences more than similarities. A related quantity of interest is the marginal mean (MM) proposed by Leeper *et al.* (2020), which is simply one of the two terms comprising an AMCE ($\mathbb{E}[Y_i([aBC], [A'B'C'])]$). The MM also incorporates both directionality and intensity of preferences about a given attribute level of interest because of its averaging. In any event, it is worth reiterating that the AMCE has an additional important feature that it directly maps onto a key election choice-related quantity of interest, as previously described.

---

In general, estimators of the AMCE are not appropriate for estimating alternative quantities, such as the effect of attributes on the probability of winning. Indeed, electoral systems research has long recognized that vote shares do not linearly translate into seat shares except under pure proportional representation (e.g., Taagepera and Shugart 1989). Here and in the Supplementary Material, we define alternative quantities of interest from conjoint experiments that are potentially useful for analyzing elections and that map onto intuitively meaningful election-related concepts.[9] We focus on two types of quantities, related to the probability of winning and the fraction of voters preferring an attribute. The Supplementary Material also sketches possible estimation approaches using model-based procedures, which are potentially promising approaches to investigate various electoral quantities, but leaves more detailed technical discussion for future research. As we demonstrate, one of our alternative quantities can be feasibly estimated using typically sized conjoint data, whereas the other cannot.[10] These results highlight the important broader point that researchers should craft their experimental designs and data collection methods with the ultimate quantity(ies) of interest in mind.

### 3.1 Probability of Winning

In a paired forced-choice conjoint design simulating a two-candidate election, a natural quantity of interest is the *probability of winning*, or the probability that one candidate will win a majority of votes. To formalize this quantity, recall that respondent $i$ chooses candidate $[abc]$ over candidate $[a'b'c']$ if and only if $Y_i([abc],[a'b'c']) = 1$. Candidate $[abc]$ therefore wins a majority vote against candidate $[a'b'c']$ if and only if

$$\mathbb{E}_{\mathcal{V}}[Y_i([abc],[a'b'c'])] > 0.5, \qquad (4)$$

where the expectation $\mathbb{E}_{\mathcal{V}}$ is defined over the target voter distribution $\mathcal{V}$. In words, candidate $[abc]$ wins a majority vote against candidate $[a'b'c']$ if more than half of respondents drawn from the target voter distribution would choose $[abc]$ over $[a'b'c']$, or equivalently if $[abc] > [a'b'c']$ for more than half of respondents.

Equation (4) constitutes a building block for various possible quantities of interest we term probabilities of winning. Let $M([ABC],[A'B'C']) \equiv \mathbf{1}\{\mathbb{E}_{\mathcal{V}}[Y_i([ABC],[A'B'C'])] > 0.5\}$, a binary random variable representing whether profile $[ABC]$ wins a majority of votes against $[A'B'C']$. For example, suppose that the researcher is interested in how likely a candidate with attributes $A = a$, $B = b$, and $C = c$ is to win a majority against another candidate randomly drawn from the target population. This probability is

$$\mathbb{E}_{\mathcal{A}}[M([abc],[A'B'C'])], \qquad (5)$$

where the expectation $\mathbb{E}_{\mathcal{A}}$ is taken with respect to the target attribute distribution $\mathcal{A}$, which the second candidate's attributes $A'$, $B'$, and $C'$ are drawn from. Alternatively, the researcher might be interested in a particular attribute (e.g., $A = a$) and how likely a candidate with that attribute is to win under majority rule. This alternative quantity can be defined as

$$\mathbb{E}_{\mathcal{A}}[M([aBC],[A'B'C'])], \qquad (6)$$

where the expectation now averages over the first candidate's attributes other than $A$ and the second candidate's attributes. Yet another quantity of interest is how often a candidate with

---

---

attribute $A = a$ will win against a candidate with attribute $A = a'$. This quantity is

$$\mathbb{E}_{\mathcal{A}}\left[M([aBC], [a'B'C'])\right], \tag{7}$$

where the expectation is now defined with respect to the distribution of $B$, $C$, $B'$, and $C'$.

The choice between different conceptions of the probability of winning depends on researchers' substantive question. Researchers may be interested in a real-world politician and ask about the likelihood a similar candidate would win a majority (Equation (5)). Alternatively, researchers might ask how likely a female candidate is to win a majority against a randomly drawn candidate (Equation (6)). Regardless of the estimand chosen, it is imperative to clarify one's substantive question of interest and map it to a well-defined estimand in terms of potential outcomes.

Inference about the probability of winning proves more challenging than about AMCEs. This is due to the nonlinearity built into majority rule (or, more generally, into the electoral formula which translates votes into seats) and the resulting high-dimensional estimation problem. To see the challenge, consider estimating the probability of winning for a female candidate against a male candidate (Equation (7)) where $A = a$ ($a'$) represents a female (male) candidate. Without additional assumptions about the potential outcomes' functional form, we can obtain a sample analog of Equation (7) as follows: calculate the female candidate's vote share for each of the $W$ possible unique contests between females and males, determine whether the female candidate wins the majority in each, and calculate the average of the resulting indicators over the $W$ contests.

Although this nonparametric plug-in estimator is consistent for Equation (7) as the numbers of respondents ($N$) and tasks ($K$) grow infinitely for a fixed number of attributes ($L$), a practical difficulty is that $W$ is very large compared to the sample size ($NK$) in typical conjoints, making the data too sparse for this inferential problem. For example, with eight binary attributes, there are $W = 2^{(8-1)\times 2} - 1 = 16,383$ possible unique contests between female and male candidates. With 1,000 respondents each completing 20 tasks, we can only expect to have slightly more than one observation for each possible pairwise comparison. Thus, the fully nonparametric estimator is impractical in all but the simplest experiments.

More promising would be a model-based approach which explicitly models the majority indicator $M([ABC], [A'B'C'])$ as a function of the attributes, which can then be used to estimate any of the probability of winning quantities of interest defined above by averaging the estimated $M$ over the distribution of the attributes corresponding to the target estimand. In Section C of the Supplementary Material, we sketch one such procedure and perform simulations that provide baseline evidence for the feasibility of using conjoint data to estimate quantities of interest related to the probability of winning. The results highlight how probabilities of winning, while substantially more difficult to estimate than AMCEs, are still promising quantities of interest to evaluate using conjoint data. In addition, some minor tailoring of the conjoint design specifically for this quantity of interest could be undertaken to further improve the estimation. For instance, if one were interested in Equation (7) with respect to one particular attribute, fixing $a$ for the first profile and $a'$ for the second profile (while randomizing the other attributes) would improve the precision of the subsequent estimates. More research in this area is valuable, and it relates to the broader point that experimental design should be tailored for each causal quantity of interest.

### 3.2 Fraction of Voters Preferring an Attribute

Another set of possible quantities of interest pertains to the fraction of voters preferring attribute $A = a$ over $A = a'$ and whether that fraction constitutes a majority. To define this quantity meaningfully, we first must define preferences over individual attributes (as opposed to profiles as a whole), which had not previously been necessary. Drawing on Section 2's definition of

preferences, we say that *a voter prefers attribute $A = a$ to $A = a'$ if and only if the average rank for $a$ is less than the average rank for $a'$*.[11] It is then straightforward that, assuming $\mathcal{A}$ to be uniform over the set of all possible attribute combinations, voter $i$ prefers attribute $A = a$ over $A = a'$ if and only if $\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])] > 0.5$, which follows from the fact that $S_i^A(a) < S_i^A(a')$ iff $\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])] > 0.5$.[12]

Based on this definition, we can define the fraction of voters preferring $A = a$ over $A = a'$ as

$$\mathbb{E}_{\mathcal{V}}\left[\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])] > 0.5\}\right]. \tag{8}$$

Note that this quantity does *not* equal the probability of winning defined in Equation (7) since the order of the two expectations is reversed. Instead, the quantity amounts to first classifying all voters into those (for example) preferring female versus male candidates and then calculating the proportion of female preferers.

The distinction between the two alternative quantities—the probability of winning and the fraction of voters preferring an attribute—is subtle but important. Equation (8) does not generally equal the probability of a female candidate winning a majority election against a male candidate, which may be of more interest to scholars focused on electoral outcomes. In contrast, the fraction of voters preferring attribute $A = a$ over $A = a'$ may be less informative about the importance of that attribute for actual voting behavior and outcomes in multi-attribute contexts. As our handedness example above illustrated, the vast majority of voters may prefer a right-handed candidate, but this attribute would almost never determine any voter's actual choice, making its AMCE or effect on the probability of winning nearly zero.

In Section D of the Supplementary Material, we extend our discussion of this quantity to consider more restricted versions of Equation (8) focusing on the fraction of voters preferring attribute $A = a$ over $A = a'$ *holding all other attributes equal*, a potential quantity of interest noted by Abramson *et al.* (2021). Among those, we propose Equation (8) as the definition if researchers remain interested in analyzing the fraction of voters who prefer a particular attribute regardless of its relative importance against other attributes. Nonetheless, estimating this quantity presents far greater challenges than estimating the probabilities of winning, since it requires explicitly analyzing voters' preferences at the individual level.

That is, one first needs to estimate the equation's inner expectation term, $\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])]$, which equals the probability of choosing a profile containing $A = a$ versus another profile containing $A = a'$ for a specific voter $i$. Unfortunately, only a handful of observations per voter will be available to estimate that inner expectation for any particular comparison $a$ versus $a'$ in typical conjoint experiments. This is in contrast to the probability of winning quantities of interest, in which the inner expectation is taken with respect to all voters, and hence can be modeled and estimated on the basis of an entire dataset. Poor performance in estimating the fraction-preferring quantity is therefore likely, as the application of the indicator function to a noisy input will result in severe misclassification, and misclassification is always negatively correlated with the true value. Indeed, Section D of the Supplementary Material highlights this problem based on the same simulations used for assessing the probability of winning estimands. For researchers interested in a fraction-preferring quantity, alternative research designs are likely warranted.

---

11 More generally, denote a profile by a length-$L$ vector $p = [d_1, \ldots, d_L]$ such that $d_l \in \{1, \ldots, D_l\}$. Let $r(p) \in \{1, \ldots, R\}$ denote the rank of profile $p$, where $R = \prod_{l=1}^{L} D_l$. Define the average rank for the $l$th attribute $d_l = f$ as $S^l(f) \equiv \frac{D_l}{R} \sum_{d_1=1}^{D_1} \cdots \sum_{d_{l-1}=1}^{D_{l-1}} \sum_{d_{l+1}=1}^{D_{l+1}} \cdots \sum_{d_L=1}^{D_L} r([d_1, \ldots, f, \ldots, d_L])$. Then, $f \gtrsim f'$ for attribute $l$ iff $S^l(f) \leq S^l(f')$.

12 This definition also applies to the general case of arbitrarily distributed attributes: a voter prefers $a$ to $a'$ if and only if $\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])] > 0.5$, that is, if she would choose a profile including $a$ more often than one including $a'$ on average. The equivalence to the average ranks definition, however, no longer holds when $\mathcal{A}$ is nonuniform.

---

### 3.3 Revisiting the AMCE

The above discussion of alternative quantities of interest returns us to the AMCE's third desirable property: empirical tractability. As Hainmueller *et al.* (2014) detail, by virtue of attribute randomization, the AMCE can be nonparametrically identified via a simple difference in means, much like a standard experiment with a single treatment. Because the AMCE is a purely linear function of the potential outcomes, it does not require a functional form assumption as the probability of winning or the fraction of voters preferring an attribute does.

This discussion should not be novel to those familiar with causal inference and the ATE as a causal estimand. All of the quantities of interest discussed so far can be viewed as causal quantities, in that they involve counterfactual comparisons between possible combinations of attributes or treatment components. When making inferences about a causal quantity, one faces the problem of identifying counterfactual comparisons not directly observed in the data. As is well known, treatment randomization solves this problem for common causal estimands such as the ATE. Less well known, however, is that randomization solves the identification problem only for a certain class of causal estimands. Fortunately, this class of estimands includes causal effects such as the ATE. However, it excludes others, such as the *median* treatment effect, or the effect of the treatment on an individual unit at the median of the individual-level treatment effect distribution.

This is analogous to the relationship between the AMCE and alternative aggregations of treatment effects. Whereas the AMCE is nonparametrically identified by the observed difference in means due to random assignment, quantities involving nonlinear mappings such as the probability of winning require additional assumptions and/or complicated modeling techniques. This helps explain why recent empirical conjoint applications have gravitated toward AMCEs.

Scholars in various fields have focused on the ATE because it can be identified with minimal assumptions and provides a useful, interpretable summary of causal effects. In that regard, the fact that the AMCE combines both preference directionality and intensity is a feature, not a bug. If a small number of people always support a candidate with a specific attribute *a*, they may overwhelm the majority of respondents who slightly prefer its inverse *a'*. This is true of the ATE, too; if a small number of lives are saved by taking a medication, that may overwhelm the temporary, negative side effects that a larger number of people experience on any measure of long-term health. Combining directionality and intensity is fitting in many political applications: in many cases, a minority of people with intense preferences over a certain attribute can drive its electoral significance. Moreover, this is not merely a rhetorical point because, as illustrated above, the AMCE identifies the difference in expected vote shares.

## 4 Practical Recommendations

While our analysis demonstrates that the AMCE recovers a meaningful quantity of interest for elections scholars, it also uncovers nuances in the interpretation of the AMCE and raises cautions against possible misinterpretations. Here, we provide guidance on what type of language applied researchers can use to summarize empirical findings based on AMCEs.

There are at least two straightforward ways to describe AMCE estimates. First, consider the generic case in which respondents choose between profiles in a forced-choice design. Here, the AMCE can be described as the effect on the probability of choosing a profile when an attribute changes values for that profile. So one might say: "Changing the age of the candidate from young to old increases the probability of choosing the candidate profile by $\delta$ percentage points."[13]

---

13  Note this applies generally to nonelectoral contexts, such as choice between products. The formulation of the effect as a choice probability also works similarly for choices between more than two profiles.

Second, consider elections, in which the conjoint involves a choice between candidates. Here, the AMCE can also be interpreted as the effect on the candidate's expected vote share when an attribute changes values. For example, one could state: "Changing the age of the candidate from young to old increases the expected vote share of the candidate by $\delta$ percentage points." Thus, AMCEs in electoral conjoints allow applied researchers to make succinct empirical statements about core quantities of interest.

Certainly, the AMCE involves nuances which researchers should master. In particular, again using the candidate age example, the difference in the expected vote share specifically refers to the vote share difference that any young versus old candidate would obtain on average against an opponent randomly drawn from the attributes' randomization distribution (see Section 2.2). Moreover, the usual caveats about interpreting survey experiments apply: one must exercise caution when the goal is extrapolating empirical findings from survey experiments to outcomes in actual elections (but see Auerbach and Thachil 2018; Hainmueller, Hangartner, and Yamamoto 2015).

Additionally, researchers should remember that the AMCE averages the effect of an attribute over two distributions: the attribute distribution and the respondent distribution. This has two important implications. The first is that the sampling strategy and experimental design should reflect the target distributions (i.e., $\mathcal{A}$ and $\mathcal{V}$ defined in Section 2.4) which researchers seek to make inferences about (De la Cuesta *et al.* 2021; Hainmueller *et al.* 2014).[14] The second is an opportunity: researchers can actually investigate the heterogeneity over which the AMCE averages by analyzing conditional AMCEs (Hainmueller *et al.* 2014). Most commonly, researchers focus on conditional AMCEs as defined by particular subsets or characteristics of respondents (see Bansak *et al.* 2021b and Leeper *et al.* 2020 for design advice). In addition, a similar though less common form of this investigation involves analyzing AMCEs conditional upon a restricted subset of profiles—for instance, AMCEs for one particular attribute while holding another attribute fixed at specific values (see Bansak, Hainmueller, and Hangartner 2016 for an example, as well as Egami and Imai 2019 for the related task of explicitly estimating interactions between attributes). If the researcher worries that an overall AMCE may conceal important patterns of heterogeneity, she can use AMCEs estimated for subsets of the data to investigate that possibility empirically.

Lastly, as our discussion has shown, researchers should not interpret the AMCE estimates as referring to the fraction of respondents who prefer a specific attribute or the probability of winning the election. In other words, using again our candidate age example, researchers should not infer from a positive AMCE estimate that (1) the majority of voters necessarily prefer old candidates to young candidates, (2) that the proportion of voters who prefer old candidates is $\delta$ higher than the proportion who prefer young candidates, or (3) that changing the age of the candidate from young to old increases the probability that the candidate would win an election by $\delta$ percentage points. The AMCE is not (and never was) designed to estimate these alternative quantities of interest.

Of course, researchers can also consider using conjoint data to investigate quantities of interest other than the AMCE, such as those involving the relationship between attributes and the probability of a candidate winning. This can be done via the estimands introduced in Section 3.1 and the estimation procedures presented in the Supplementary Material, although additional research is valuable. For researchers interested in the fraction of voters preferring a particular attribute (Section 3.2), our recommendation is to reconsider whether the quantity really captures their substantive query, since it represents preferences about an attribute regardless of its relative

---

14  Note that De la Cuesta *et al.* (2021) also discuss how to use model-based approaches to estimate the AMCE when the target attribute distribution of interest is different from the attribute randomization distribution used in the conjoint design.

importance against other attributes. Indeed, research designs other than conjoint experiments (e.g., a direct survey question on one's preference for the attribute itself) may be better suited for such inquiries.

## 5  Conclusion

We employed a general framework for analyzing voter preferences in electoral conjoint experiments with multiple candidate attributes to study the AMCE's formal and conceptual properties. If voters have preference rankings over the set of multi-attribute profiles and vote for their preferred profiles, the AMCE recovers a core quantity of interest to election scholars: the effects of candidate attributes on expected vote shares in elections mirroring the conjoint design. This crucial AMCE property holds regardless of the structure of voter preferences or the electoral formulae mapping votes to seats. Additionally, we explored other possible quantities of interest in conjoint experiments and discussed possible estimation strategies. We also provided practical guidance on interpreting AMCEs and analyzing conjoint data.

Our study has several implications. First, our results highlight the essential role of the AMCE for analyzing elections using conjoint experiments. AMCEs—under general conditions—identify the effects of changes in attributes on candidates' expected vote shares. As our literature review showed, vote shares are the central outcome of interest for much of the elections literature. The bottom line is simple: if one is interested in candidate or party attributes' effects on vote shares, the AMCE is a fitting tool. Not only do AMCEs identify the effects on vote shares under general conditions, they are also easy to estimate and not reliant on arbitrary functional form assumptions.

Second, by going beyond AMCEs, we highlight that conjoint experiments can be informative about other, less widely used causal quantities. Specifically, we have defined several estimands that capture the relationship between attributes and the probability of winning and sketched estimation procedures. This revealed that it is important to precisely define what is being compared when considering relative probabilities of winning. In addition, estimation of such quantities requires additional modeling assumptions beyond those guaranteed by randomization, although our research provides preliminary evidence that this is feasible and can be done reliably. We contrasted this set of quantities with another alternative, the fraction of voters preferring a specific attribute, which cannot be feasibly investigated using typical conjoint data, and may also be less informative for election scholars focused on voter choice and electoral outcomes.

Third, our analysis of the AMCE and alternative quantities of interest addresses concerns raised by other scholars who suggest that AMCEs are not informative on questions of interest to political scientists. On the contrary, our findings demonstrate that AMCEs are central for scholarship on elections. Certainly, it is correct that AMCEs do not disaggregate preference directionality from intensity and hence will not necessarily correspond to the fraction of voters preferring an attribute $A = a$ over $A = a'$, as highlighted by Abramson $et\ al.$ (2021). Yet, the fraction of voters preferring a specific attribute has not to date been of significant interest to empirical election scholars, perhaps because it does not account for the multi-attribute nature of elections. Put differently, just because many voters might prefer a specific attribute in isolation does not mean that this attribute will have much effect on vote shares or the probability of winning since voting might be mainly driven by more important attributes. Conversely, the AMCE addresses what often interests election scholars by revealing how an attribute affects vote shares averaging across candidates with many possible combinations of other attributes.

Finally, our study points to fruitful avenues for future research. We have proposed procedures for estimating alternative quantities of interest related to candidates'/parties' probability of winning elections/seats, which may be starting points for future inquiries. With additional modeling

along with carefully tailored designs, such approaches may extract further insights from conjoint data.

## Acknowledgment

## Supplementary Material

For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2022.16.

## Data Availability Statement

Replication materials are available in Bansak *et al.* (2022).

## References

Abramson, S. F., K. Koçak, and A. Magazinnik. 2021. "What Do We Learn about Voter Preferences from Conjoint Experiments?" *American Journal of Political Science* (forthcoming).

Achen, C., L. Bartels, C. H. Achen, and L. M. Bartels. 2017. *Democracy for Realists*. Princeton, NJ: Princeton University Press.

Adida, C. L., A. Lo, and M. R. Platas. 2019. "Americans Preferred Syrian Refugees Who Are Female, English-Speaking, and Christian on the Eve of Donald Trump's Election." *PLoS One* 14 (10): e0222504.

Auerbach, A. M., and T. Thachil. 2018. "How Clients Select Brokers." *American Political Science Review* 112 (4): 775–791.

Bansak, K., M. M. Bechtel, and Y. Margalit. 2021. "Why Austerity? The Mass Politics of a Contested Policy." *American Political Science Review* 115 (2): 486–505.

Bansak, K., J. Hainmueller, and D. Hangartner. 2016. "How Economic, Humanitarian, and Religious Concerns Shape European Attitudes Toward Asylum Seekers." *Science* 354 (6309): 217–222.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto. 2021a. "Beyond the Breaking Point?" *Political Science Research and Methods* 9 (1): 53–71.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto. 2021b. "Conjoint Survey Experiments." In *Advances in Experimental Political Science*, edited by J. N. Druckman and D. P. Green. Cambridge, MA: Cambridge University Press.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto. 2022. "Replication Materials for: Using Conjoint Experiments to Analyze Election Outcomes: The Essential Role of the Average Marginal Component Effect (AMCE)." https://doi.org/10.7910/DVN/NBRBHO, Harvard Dataverse, V1.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes. 1960. *The American Voter*. New York: Wiley.

De la Cuesta, B., N. Egami, and K. Imai. 2021. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution." *Political Analysis* 30: 19–45.

Egami, N., and K. Imai. 2019. "Causal Interaction in Factorial Experiments." *Journal of the American Statistical Association* 114 (526): 529–540.

Ganter, F. 2020. "Identification of Preferences in Forced-Choice Conjoint Experiments: Reassessing the Quantity of Interest." Working Paper, Columbia University.

Hainmueller, J., D. Hangartner, and T. Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments Against Real-World Behavior." *Proceedings of the National Academy of Sciences* 112 (8): 2395–2400.

Hainmueller, J., D. J. Hopkins, and T. Yamamoto. 2014. "Causal Inference in Conjoint Analysis." *Political Analysis* 22 (1): 1–30.

Hanretty, C., B. E. Lauderdale, and N. Vivyan. 2020. "A Choice-Based Measure of Issue Importance in the Electorate." *American Journal of Political Science* 64: 519–535.

Huff, C., and J. D. Kertzer. 2018. "How the Public Defines Terrorism." *American Journal of Political Science* 62 (1): 55–71.

Jenke, L., K. Bansak, J. Hainmueller, and D. Hangartner. 2021. "Using Eye-Tracking to Understand Decision-Making in Conjoint Experiments." *Political Analysis* 29: 75–101.

Key, V. 1966. *The Responsible Electorate: Rationality in Presidential Voting, 1936–1960*. Cambridge, MA: Belknap Press.

Leeper, T. J., S. B. Hobolt, and J. Tilley. 2020. "Measuring Subgroup Preferences in Conjoint Experiments." *Political Analysis* 28: 207–221.

Leeper, T. J., and J. Robison. 2020. "More Important, but for What Exactly? The Insignificant Role of Subjective Issue Importance in Vote Decisions." *Political Behavior* 42 (1): 239–259.

Mummolo, J., and C. Nall. 2017. "Why Partisans Do not Sort: The Constraints on Political Segregation." *Journal of Politics* 79 (1): 45–59.

Schumpeter, J. A. 1950. *Capitalism, Socialism and Democracy*. London: Routledge.

Schwarz, S., and A. Coppock. 2022. "What Have We Learned about Gender from Candidate Choice Experiments? A Meta-Analysis of 67 Factorial Survey Experiments." *Journal of Politics* 84 (2): 655–668.

Taagepera, R., and M. S. Shugart. 1989. *Seats and Votes*. New Haven, CT: Yale University Press.