Using Conjoint Experiments to Analyze Elections: The Essential Role of the Average Marginal Component Effect (AMCE)*

Kirk Bansak[†] Jens Hainmueller[‡] Daniel J. Hopkins[§] Teppei Yamamoto[¶]

First draft: April 30, 2020
This draft: April 30, 2020

Abstract

Political scientists have increasingly deployed conjoint survey experiments to understand multi-dimensional choices in various settings. We begin with a general framework for analyzing voter preferences in multi-attribute elections using conjoints. With this framework, we demonstrate that the Average Marginal Component Effect (AMCE) is well-defined in terms of individual preferences and represents a central quantity of interest to empirical scholars of elections: the effect of a change in an attribute on a candidate or party's expected vote share. This property holds irrespective of the heterogeneity, strength, or interactivity of voters' preferences and regardless of how votes are aggregated into seats. Overall, our results indicate the essential role of AMCEs for understanding elections, a conclusion buttressed by a corresponding literature review. We also provide practical advice on interpreting AMCEs and discuss how conjoint data can be used to estimate other quantities of interest to electoral studies.

*The authors thank Max Spohn for extensive research assistance and Avidit Acharya, Alex Coppock, Justin Grimmer, Kosuke Imai, Ben Lauderdale, Thomas Leeper, and Yiqing Xu for useful comments.

[†]Assistant Professor, Department of Political Science, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, United States. E-mail: kbansak@ucsd.edu

[‡]Professor, Department of Political Science, 616 Serra Street Encina Hall West, Room 100, Stanford, CA 94305-6044. E-mail: jhain@stanford.edu

[§]Professor, Department of Political Science, University of Pennsylvania, Perelman Center for Political Science and Economics, 133 S. 36th Street, Philadelphia PA, 19104. E-mail: danhop@sas.upenn.edu

[¶]Associate Professor, Department of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. Email: teppei@mit.edu, URL: http://web.mit.edu/teppei/www

1 Introduction

Over the past several years, conjoint survey experiments have been widely used in political science to study voter preferences in elections.[1] 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 without invoking strong modeling assumptions (Hainmueller, Hopkins and Yamamoto, 2014). 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 of the candidate's attributes were switched from one level to another (Hainmueller, Hopkins and Yamamoto, 2014). The introduction of this approach sparked a number of conjoint applications, many focused on electoral politics (see Bansak et al., forthcoming, for a review). It has also prompted the development of statistical tools (Egami and Imai, 2019; Leeper, Hobolt and Tilley, forthcoming; de la Cuesta, Egami and Imai, 2019; Hanretty, Lauderdale and Vivyan, 2020).

There are many situations of interest to social and political scientists which seem ripe for analysis through conjoint designs, as they present individuals with the opportunity to rank or choose between bundles comprised of multiple attributes. Voting behavior certainly has these basic elements, but so, too, do choices about which immigrants to admit, which policy packages to adopt, and various other research topics.[2] To date, however, the empirical adoption of conjoint designs has outpaced theoretical discussions of precisely what quantities conjoint designs can—and cannot—recover. As a result, some scholars have critiqued common practices employed for analyzing and interpreting conjoint experiments (e.g. Leeper, Hobolt and Tilley, forthcoming; Abramson, Koçak and Magazinnik, 2019).

In this paper, we illuminate the theoretical and conceptual microfoundations underpinning the AMCE and compare the AMCE with other possible quantities of interest that are also applicable

[1]See, for example, Loewen, Rubenson and Spirling (2012); Franchino and Zucchini (2014); Abrajano, Elmendorf and Quinn (2015); Carnes and Lupu (2016); Horiuchi, Smith and Yamamoto (2018); Kirkland and Coppock (2018); Auerbach and Thachil (2018); Matsuo and Lee (2018); Crowder-Meyer et al. (2018); Teele, Kalla and Rosenbluth (2018); Goggin, Henderson and Theodoridis (2019); Arnesen, Duell and Johannesson (2019); Ono and Burden (2019); Ryan and Ehlinger (2019)

[2]See, for example, Gampfer, Bernauer and Kachi (2014); Bansak, Hainmueller and Hangartner (2016); Bechtel, Genovese and Scheve (2016); Mummolo and Nall (2016); Schachter (2016); Wright, Levy and Citrin (2016); Adida, Lo and Platas (2017); Stokes and Warshaw (2017); Flores and Schachter (2018); Hankinson (2018); Auer et al. (2019); Clayton, Ferwerda and Horiuchi (2019)

to paired-profile, forced-choice conjoint designs and can be defined under a common framework. In unpacking the AMCE, we clarify how it should and should not be interpreted, and we highlight how it aggregates individual-level preferences into a central quantity of interest for electoral scholars. We show that when applied to elections, the AMCE has a straightforward, politically meaningful interpretation as the average causal effect of an attribute on a candidate's or party's expected *vote share*. Importantly, this equivalence between AMCEs and effects on vote shares holds regardless of the structure of voter preferences. Through a literature review of 82 articles in four leading electoral politics journals, we demonstrate that vote shares and their individual-level analogs are indeed far and away the most common quantities of interest in empirical electoral research.

In sum, the AMCE provides a fitting tool for researchers interested in using conjoints to study the effects of candidate or party attributes on vote shares. AMCEs not only identify a key quantity of interest that involves the central outcome in the literature on elections and voting, they are also easy to estimate and do not rely on arbitrary functional form assumptions. To be sure, for the results to translate meaningfully to real-world elections, scholars still need to ensure that their conjoint designs are well crafted by paying close attention to the selection of the sample, the inclusion of relevant attributes, the realism of the set-up, the number of attributes and tasks, and the randomization distribution of the attributes (see Bansak et al., forthcoming; Hainmueller, Hopkins and Yamamoto, 2014; de la Cuesta, Egami and Imai, 2019, for discussion of these issues).

The fact that the AMCE recovers a key quantity of interest to election scholars does not mean it is the only appropriate estimand in conjoint analyses of elections. Thus, we also examine how paired-profile forced-choice conjoint data could potentially be used for studying other electoral quantities of interest. Specifically, we distinguish between two main alternatives that are distinct from the effects on vote shares. The first is the effect of an attribute on a candidate's probability of winning an election. The second is the fraction of voters who prefer a specific attribute, the focus of some previous work (Abramson, Koçak and Magazinnik, 2019). We define these alternative quantities of interest under the same framework used to define the AMCE, thereby formalizing precisely how they are distinct both from each other and the AMCE.

This analysis produces two main insights. First, effects on the probability of winning are a meaningful quantity of interest to study the outcomes of elections between multi-attribute candidates. Yet, due to the nonlinearity built into the majority rule, estimating such effects requires a model-based approach to approximate a high-dimensional conditional expectation function. This is in stark contrast to the AMCE, which can be estimated without such modeling assumptions via a design-based approach motivated purely by the randomization. We provide sketches of possible estimation procedures for the probability of winning that may be promising for future research. Second, the quantity representing the fraction of voters who prefer an attribute is associated with fundamental problems both in terms of interpretation and estimation. It is not directly informative about a given attribute's effect on the outcome of elections between multi-attribute candidates and is substantially more challenging to estimate than even the probability of winning.

We conclude by providing practical guidance for applied researchers who seek to employ conjoint experiments to understand how people make choices among bundles of attributes and suggesting possible paths for future research. Overall, this paper contributes to the growing methodological literature on conjoint survey experiments by grounding the most commonly used causal estimand—the AMCE—in a foundational theory of individual preferences and showing its interpretability as a key quantity of interest to electoral scholars.

2 Microfounding the AMCE

To provide a microfoundation for the AMCE, we first present a general framework for analyzing voter preferences in multi-attribute elections, where candidates are characterized by multiple observed attributes.[3] 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 of this approach is that the AMCE identifies a central quantity of interest in electoral research: the average effect of an attribute on expected vote shares.

2.1 Formalizing Preferences in Multi-Attribute Elections

To begin, consider a paired-profile forced-choice conjoint experiment, where each respondent $i \in \{1, ..., N\}$ completes a series of K tasks in which the respondent makes a hypothetical vote choice between two candidates varying in terms of L attributes. Each of the L attributes takes on D_l

[3]Although we focus here on voters selecting between candidates defined by different attributes, the same general framework can be applied to any other multi-dimensional choice, such as consumers selecting among multi-dimensional products.

Table 1: Candidate Preference Rankings for Three Types of Voters

Profile Number	Attribute A	Attribute B	Attribute C	Profile Rank: Type 1	Profile Rank: Type 2	Profile Rank: 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

Shows the preference ranking for three types of voters over profiles of candidates defined by three binary attributes A, B, and C.

discrete levels, respectively, such that $l \in \{1, ..., L\}$. One can view this design as a simulation of a two-candidate election in which a citizen votes for one of two candidates varying in terms of L observed attributes.

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

In a choice setting where alternatives are characterized by multiple attributes, a natural way to formalize individual preferences is to consider a preference ordering over the full set of possible unique attribute combinations. Namely, we define individual preferences to be binary relations over the set of possible unique candidates for each voter. To simplify exposition, we assume that each voter has a strict preference ordering over all $\prod_{l=1}^{L} D_l$ possible unique candidates. For example, consider the voter represented as "Type 1" in Table 1. In the Table, the 8 possible candidate profiles (defined in the second to fourth columns from the left) are ordered from top to bottom according to the Type 1 voter's preference ranking (the third column from the right). This preference can also be represented using the standard decision-theoretic notation, such that $[111] \succ [110] \succ [101] \succ [100] \succ [011] \succ [010] \succ [001] \succ [000]$.

2.2 Defining the AMCE

Since elections are means of preference aggregation, a question for electoral researchers is how one can learn about collective decisions from individual preferences expressed through conjoint experiments. That is, how can we aggregate individual conjoint responses into a useful quantity of interest for elections scholars? In thinking about this aggregation problem, it is fruitful to begin with several desirable criteria for any such aggregate preference measure. First, the quantity of interest should capture the multidimensionality of the typical electoral choice task, in which individual voters must choose between candidates differing across many dimensions simultaneously. Second, the quantity should map onto a meaningful empirical phenomenon of interest, such that electoral researchers can make causal or predictive inferences about elections based on the quantity. Third and finally, the quantity should be empirically tractable, in the sense that researchers can use observed data from actual conjoint experiments to estimate the quantity with sufficient statistical precision and ideally without strong modeling assumptions.

The Average Marginal Component Effect (AMCE) is one quantity of interest for evaluating an aggregate relationship between attributes and preferences, and, as will be shown in detail, the AMCE meets all three criteria. First, the AMCE aggregates preference orderings over the full set of possible profiles in a systematic manner that takes into account the multidimensional nature of the electoral decision problem by incorporating not only the directionality but also the intensity of preferences. Second, the AMCE directly represents the causal effect of a particular attribute on a candidate's expected vote share, which is revealed via a review of the literature to be the most prominent causal quantity of interest for electoral scholars. Third and finally, identification and unbiased estimation of the AMCE can proceed under a limited set of assumptions and via straightforward nonparametric methods, as has already been shown in previous work (Hainmueller, Hopkins and Yamamoto, 2014). We highlight each of these features in the rest of this section.

To define the AMCE under the current set-up, let $Y_i([abc], [a'b'c']) \in \{0, 1\}$ denote the potential outcome for voter i given a paired forced-choice contest between profiles $[abc]$ and $[a'b'c']$. The potential outcome would take on a value of 1 if respondent i would choose the first candidate (i.e. $[abc]$) given the choice task, which would occur if and only if $[abc] \succ [a'b'c']$ for that respondent. In contrast, $Y_i([abc], [a'b'c']) = 0$ if respondent i chooses the second candidate (i.e. $[a'b'c']$) in that contest, which would be the case if and only if $[a'b'c'] \succ [abc]$ for that respondent. Then, the AMCE for attribute A is defined as the expected difference between the potential outcomes for all paired contests where attribute A for the first candidate equals 1 and the potential outcomes for all contests where A equals 0 for the first candidate, given a known, pre-specified distribution of the other attributes. That is, without loss of generality, the AMCE for attribute A is given by the following expression:

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

where the expectation is defined over both the joint distribution of the candidate attributes from which all the attributes other than A for the first candidate (i.e., B, C, A', B' and C') are drawn, as well as 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 for conjoint designs with larger numbers of attributes, attributes with more than two levels, non-forced-choice outcomes, and tasks with more than two profiles, with appropriate changes in the notation.

A few remarks are in order to illuminate the nature of the definition in equation (1). First, note that the AMCE aggregates individual preferences with respect to two dimensions: across attributes and across voters. Specifically, the AMCE employs *averaging* (i.e., expectation, mean) of individual preferences both across the set of possible candidates and across voters in the target population. This is in contrast to other means of preference aggregation often studied in the classical social choice literature, such as the simple majority rule. As discussed later in this section, the double averaging turns out to imply several desirable properties of the AMCE in terms of substantive relevance and empirical tractability.

Second, 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 pre-specified distribution. That is, suppose attribute A is gender, such that $A = 1$ means the candidate is female and $A = 0$ means the candidate is male. 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 random profile. In other words, the AMCE asks how much better or worse a randomly selected candidate in the election would fare if the gender switches from male to female. In particular, 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.

2.3 The AMCE and Preference Intensity

The first of the AMCE's desirable properties is that it can capture the multidimensionality of the choice task in conjoint experiments. As it turns out, it does so by way of incorporating both the direction and intensity of preferences about individual attributes through averaging the *ranks* of profiles. Continuing with the three-attribute example, let $r_i(a, b, c) \in \{1, ..., 8\}$ represent the rank of the profile $[abc]$ for a 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 not only the directionality but also the intensity of her preferences with respect to the attribute.

For example, consider the Type 1 voter represented in Table 1. Intuitively, it is apparent that this voter strongly favors $A = 1$ to $A = 0$ because profiles containing $A = 1$ are more highly ranked than any profile containing $A = 0$ no matter what the other attributes are. For attribute B, the voter favors profiles with $B = 1$ to those with $B = 0$, but only in so far as the profiles are not better in terms of A. Finally, as for C, the voter generally likes profiles with $C = 1$ better than those with $C = 0$, but the value of C only matters for the final ranking when the profiles are tied in terms of all other attributes. Thus, we can summarize these preferences as an intense preference for $A = 1$ to $A = 0$, a moderate preference for $B = 1$ to $B = 0$, and only a mild preference for $C = 1$ to $C = 0$.

Considering the average profile ranks with respect to different attribute levels, as proposed above, captures these intuitions accurately. For illustration, the average ranks for the Type 1 voter are provided in Table 2. The average rank for a Type 1 voter i with respect to $A = 1$, $S_i^A(1)$, is equal to 2.5, while the average rank with respect to $A = 0$ is 6.5. This implies that the voter

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

Shows the average ranks for profiles of candidates with and without a given attribute for the three types of voters. Type 1 voters have a intense preference for A, a moderate preference for B, and a mild preference for C. Type 2 voters have mild preference for not A, a moderate preference for B, and a intense preference for C. Type 3 voters have mild preference for not A, a moderate preference for B, and a intense preference for not C.

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 each of the attributes: for attributes A, B and C, the rank means are 2.5 vs. 6.5 (intense preference), 3.5 vs. 5.5 (moderate preference), and 4 vs. 5 (mild preference), respectively. As we explain below, incorporating these differences in the intensity of the preferences over attributes is key for capturing the importance of the attributes for the resulting vote choices in contests between multi-dimensional profiles.

The AMCE is, in fact, 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 level of the attribute over the other level (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 $i \in \{1, ..., N\}$, which we denote by $\bar{S}^A(1) - \bar{S}^A(0)$. As shown by Abramson, Koçak and Magazinnik (2019), 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 in this way, it is clear how the AMCE represents an aggregation of individual preferences that explicitly takes intensity into account, as $S_i^A(1) - S_i^A(0)$ represents an individual voter's relative intensity of preference for $A = 1$ over

$A = 0$, and $\bar{S}^A(1) - \bar{S}^A(0)$ averages this over all voters.

Why is the quantification of preference intensity, in addition to binary preference relations, important or useful? After all, a cursory extrapolation from classical social choice theory might lead one to believe that relative intensity of individual preference should not play a role in determining collective choice outcomes. Such reasoning, however, turns out to be misleading when one takes the multidimensionality of preferences into consideration. In real-world elections where votes are cast for candidates characterized by more than one attribute, candidates in any particular match-up 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.

By way of an example, consider its implications for an attribute on which voters may hold largely homogeneous views but that is trivial from the practical standpoint of voter choice, such as candidates' handedness (i.e. right-handed vs. left-handed) as one of several candidate attributes. For the sake of argument, assume that a voter would *all else equal* rather choose a candidate who shares the same handedness as her- or himself. Because the vast majority of people are right-handed, there would then 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, we might then even expect the size of this majority preference for right-handedness (i.e. the fraction of voters who prefer this attribute all else equal) to exceed that for any other attributes in the evaluation, such as age, previous experience, policy positions, etc.

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 were to consider voters' attribute preference intensity according to the average rank framework above, voters' preference for a right-handed candidate would be trivially mild, as the average rank of right-handed candidates would be only slightly higher than that of left-handed candidates. This reflects real-world voting behavior: it goes without saying that in the real world, voters would ignore the handedness information when presented with the multidimensional candidate profiles and make their choices as a function of the attributes they deemed to actually be relevant (such as party affiliation, policy positions, etc.).

By taking preference intensity into account, the AMCE captures this real-world behavior, and in this example, the AMCE for right-handedness would be near zero.

Indeed, the importance of preference intensity for election outcomes has long been recognized in the large literature on probabilistic voting models in political science and political economy. Such voting models are based on the idea that vote decisions reflect uncertainty and are therefore probabilistic, rather than deterministic (see e.g. Lindbeck and Weibull (1987); Coughlin (1992); Enelow and Hinich (1989)). Typically, voter decisions are modeled as the sum of two utility components: A systematic component that reflects the utility that voters derive from observed candidate attributes (e.g. platforms or candidate characteristics) and a random utility shock in the evaluation of candidates that reflects residual uncertainty in preferences. In comparing candidates, voters back the candidate whose overall utility is higher.[4] A tenet of probabilistic voting models is that all voters have some influence on the election outcome and not just the median voter. In fact, a canonical result is that the aggregate voting outcome (i.e. net vote share) is driven by the mean (deterministic part of) utility of voters, and not the median utility. Under standard regularity conditions, the expected vote share of a candidate reflects the sum of utilities that voters derive from the candidate's observed attributes. In particular, if there are many voters of each preference type, then expected vote shares reflect both the number of voters who prefer a candidate with certain attributes as well as the intensities of each voter type's preferences over the attributes. As we show below, the AMCE represents the effect of a candidate attribute on the expected vote share. One interpretation, then, is that the AMCE reflects the change in the average voter utility that results from changing a candidate attribute.

2.4 The AMCE as the Effect on Vote Shares

The second desirable property of the AMCE for electoral research is that it represents a quantity that is of broad interest to empirical elections scholars: the average causal effect of an attribute on *vote shares*. 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 same attribute (e.g. male) in an

[4]This is similar to the random utility framework often used to motivate discrete choice models, such as multinomial or conditional logits (see e.g. Train (2009); Schofield (2007)).

election with the same number of candidates (i.e. two in a typical paired conjoint experiment). 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 of interest for both majoritarian and proportional representation (PR) elections.

Taken literally, the AMCE is only "well-defined" in the context of a conjoint survey 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 candidate profiles, how much more likely is the female candidate to win the paired forced-choice conjoint task, compared to a male candidate randomly drawn in the same manner on average? (To reiterate the point made in Section 2.2, we are *not* asking the question of how likely a female candidate is to win against a male candidate.) This quantity is arguably of interest to many applied researchers of political behavior in and of itself, since the conjoint choice tasks themselves can be robust and reliable measure of attitudes and opinions (e.g. Bansak et al., 2019; Jenke et al., Forthcoming). Nonetheless, a crucial question for many scholars of elections is whether the AMCE is also informative about elections and about the aggregation of individual preferences implemented through such elections. That is, is the AMCE informative about voter preferences in a way that maps onto electoral quantities of interest?

Here, we show that the AMCE also equals a quantity summarizing the causal effect of a candidate attribute on *vote shares* in an election matching the specifications of the conjoint experiment. By vote share, we simply mean the percentage of votes cast for a candidate in an election. The AMCE of an attribute in a conjoint experiment appropriately designed to resemble 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 same attribute) in the election. Thus, the AMCE is interpretable in terms that are directly relevant for the study of elections.

To make our point more 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 a collection of individual preferences over the attribute combinations in the support of $\mathcal{A}$. For example, consider Table 1, which represents the toy example of an election with candidates with three binary attributes and three types of voters. The attribute distribution is a probability mass function over the eight possible attribute combinations or profiles (i.e. rows in the left half of the table). 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 types of voters (i.e. columns in the right half of the table), for example $\Pr(\text{Type 1}) = .3, \Pr(\text{Type 2}) = .4$, and $\Pr(\text{Type 3}) = .3$. Note that the word "target" in these definitions indicates that these distributions usually correspond to some 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 survey experiment on a representative sample of respondents randomly drawn from the target voter distribution $\mathcal{V}$. Furthermore, suppose that profiles are randomly generated according to the target attribute distribution $\mathcal{A}$. Then it follows that the AMCE of each attribute under the design can be interpreted as the average effect of that attribute on vote shares for candidates in the target election $\langle \mathcal{A}, \mathcal{V} \rangle$. The result is more precisely and more generally stated in the following proposition.

Proposition 1 (Identification of the Expected Difference in Vote Shares with the AMCE) *Consider a J-profile conjoint experiment in which respondents are a simple random sample of size N drawn from $\mathcal{V}$. Then, 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 a candidate 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 therefore omitted.)

Proposition 1 implies that scholars of elections can use appropriately designed conjoint survey experiments to predict vote shares of candidates in elections and interpret the resulting AMCE estimates as the causal effects of candidate attributes on predicted vote shares. For example, an

AMCE of 0.2 for a male candidate versus a female candidate indicates that gender has an average causal effect of 20 percentage points on candidates' vote shares in an election that resembles the design of the conjoint experiment: on average, a randomly selected female candidate in the election would earn 20 points more of the total vote share if her gender were male. The AMCE is thus informative about voter preferences expressed through votes in elections.

But how common are vote shares as a quantity of interest in empirical research on elections? To answer that question, we conducted a literature review including all articles on voting in four prominent journals which commonly publish studies on voting behavior between 2015 and 2019.[5] We find that of the sample of 82 articles that we reviewed,[6] 87% include either aggregate vote shares or their individual-level analogs as one key outcome.[7] 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.[8] Thus, not only does the AMCE recover a politically and electorally meaningful quantity, it recovers a quantity that has been the primary quantity of interest even in most non-conjoint studies in recent years.

3 Beyond AMCEs: Alternative Quantities of Interest

As we showed in the previous section, the AMCE has desirable properties as a quantity of interest for electoral research, both in terms of theoretical interpretability and empirical tractability. In particular, the AMCE has a straightforward interpretation as a causal effect on the expected vote share. But there are of course a range of election-related outcomes that may be of interest to researchers implementing paired-profile forced-choice conjoint designs. For instance, they may

[5]The journals include *The American Political Science Review*, *The American Journal of Political Science*, *Electoral Studies*, and *Political Behavior*.

[6]Our initial parameters identified 279 published articles, and 111 of those included an estimate of the effects of candidate or party characteristics on some electorally relevant outcome. We next removed those articles which used a conjoint design, as one of the primary goals of this paper is precisely to clarify the implicit quantity of interest in electoral conjoint experiments. We also removed articles which did not evaluate vote choice specifically (most commonly because their outcome was instead candidate evaluations), leaving us with 82 articles.

[7]We classified the articles' principal outcomes as being measured in terms of several (possibly overlapping) categories. We group articles whose primary outcome is aggregate vote shares with those that consider their individual-level analog, changes in the individual-level probability of voter support for a party or candidate. We then separately identify articles whose primary outcome is the probability of a candidate/party victory or the number of seats won in a legislature.

[8]Also, a small fraction of articles do consider the mapping of individual preferences under different aggregation rules alongside an analysis of vote shares (e.g. Tsai, 2017).

wish to estimate the effect of an attribute on a candidate or party's probability of winning an election.

In this section, we show how researchers may use model-based estimation procedures to make inferences about quantities of interest beyond the AMCE from conjoint experiments. Specifically, we employ the same framework used to analyze the AMCE to define a number of other quantities that are of potential use for electoral researchers.[9] We examine how informative conjoint experiments can be with respect to those alternative quantities of interest, highlighting how broadly applicable conjoint data are for studying elections and voting. We provide sketches of possible estimation procedures, but more detailed technical discussion is beyond the scope of this paper and left for future research. We view these as potentially promising approaches for using conjoint data to investigate and estimate various electoral quantities of interest. Importantly, however, the additional challenges of these procedures relative to estimating AMCEs highlights the unique tractability of estimating the AMCE and the change in vote share that it represents.

In general, estimators designed for the AMCE are not appropriate for estimating alternative quantities, such as the probability of winning. This is not at all surprising, since the AMCE and these quantities are different estimands, both mathematically and substantively. Indeed, it has been well known in the longstanding literature on electoral systems that vote shares do not linearly translate to, for instance, seat shares except under purely proportional representation rules (e.g., Taagepera and Shugart, 1989). However, the AMCE's incongruity with alternative quantities of interest is a basis upon which use of the AMCE has been critiqued (e.g. Abramson, Koçak and Magazinnik, 2019). Although such critiques may provide a useful reminder that different mechanisms of aggregating preferences can produce different results, it is not fruitful to ask whether the AMCE is informative about alternative quantities. A more productive question, which we tackle in the rest of this section, is whether there exist other estimation strategies that can be used to make valid inferences about these quantities based on conjoint data.

[9]Some scholars have recently developed estimation strategies tailored for quantities other than the AMCE from conjoint data, such as interaction effects (Egami and Imai, 2019) and issue importance (Hanretty, Lauderdale and Vivyan, 2020). Here, we focus on quantities that are particularly relevant in the context of elections.

3.1 Probability of Winning

In a paired-profile forced-choice conjoint design simulating a two-candidate election, a natural quantity of interest is the *probability of winning*, or the probability that a particular candidate will win a majority of the votes against another candidate. To formalize this quantity using our framework, 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:

$$\mathbb{E}_{\mathcal{V}}[Y_i([abc],[a'b'c'])] \ > \ 0.5, \tag{2}$$

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 the respondents drawn from the target voter distribution would choose $[abc]$ over $[a'b'c']$ in a conjoint task, or equivalently if $[abc] \succ [a'b'c']$ for more than half of the respondents.

Equation (2) constitutes a building block for various possible quantities of interest that we can call 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 the vote against profile $[A'B'C']$. For example, suppose that the researcher is interested in how likely the candidate with attributes $A = a$, $B = b$ and $C = c$ is to win a majority vote against another candidate randomly drawn from the target population of candidates. This probability can be written as,

$$\mathbb{E}_{\mathcal{A}}\left[M([abc],[A'B'C'])\right], \tag{3}$$

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

$$\mathbb{E}_{\mathcal{A}}\left[M([aBC],[A'B'C'])\right], \tag{4}$$

where the expectation now averages over the first candidate's attributes other than A as well as the second candidate's attributes. Yet another possible quantity of interest is how often a candidate with attribute $A = a$ will win against a candidate with attribute $A = a'$. This probability can also be expressed in terms of equation (2), such that

$$\mathbb{E}_{\mathcal{A}}\left[M([aBC],[a'B'C'])\right], \tag{5}$$

where the expectation is now defined with respect to the distribution of B, C, B' and C'.[10]

The choice between different conceptions of the probability of winning depends on the researcher's substantive question. For example, the researcher might be interested in a particular real-world politician and ask how likely a candidate like her is to win an electoral majority if she were to run in an election. Equation (3) is an appropriate quantity of interest for this researcher. Alternatively, the researcher might want to learn how likely a female candidate is to win a majority vote, either against a candidate randomly drawn from the target population of candidates (equation (4) will be appropriate) or against a male candidate drawn from the population (equation (5) will be appropriate). Regardless of the choice of estimand, it is key to clarify one's substantive question of interest and explicitly map it to an estimand that is well defined in terms of the potential outcomes and the underlying individual attribute preferences.

As it turns out, inference about the probability of winning is much more challenging than inference about the AMCEs. This is due to the nonlinearity built in majority rule (or, more generally, in the electoral formula used to translate votes into seats) and the resulting high dimensionality of the estimation problem. To see the challenge, consider the problem of estimating the probability of winning for a female candidate against a male candidate, i.e., equation (5) where $A = a$ (a') represents the candidate's gender being female (male). Without any additional assumptions about the functional form of the potential outcomes, we can obtain a sample analog of equation (5) by the following procedure: calculate the vote share for a female candidate for each of the Q possible unique contests between a female and a male, determine whether the female candidate wins the majority in each unique contest, and finally calculate the average of the resulting binary majority indicators over the Q contests.

[10]In conjoint designs with more than two profiles in each task, we can define analogous quantities representing seat shares in multiparty plurality elections as natural extensions of these probabilities. For example, the probability of winning greater than some proportion t of the vote share in a J-way single-vote election is a general case of the probability of winning, and the estimation procedure described below can be adapted to accommodate this more general case by simply including any number J of profiles (of candidates or parties) in the modeling of f and replacing 0.5 with any threshold t of interest in the modeling of M.

Although this non-parametric plug-in estimator is consistent for equation (4) as the numbers of respondents (N) and tasks (K) grow infinitely for a fixed number of attributes (L), a practical difficulty is that Q is very large compared to the sample size (NK) in a typical conjoint experiment, making the data too sparse for the inferential problem at hand. For example, with eight binary attributes, there are $Q = 2^{(8-1)\times 2} - 1 = 16,383$ possible unique contests between a female candidate and a male candidate. This means that, with 1,000 respondents each completing 20 tasks, we can only expect to have slightly more than one observation to estimate a majority winner for each possible pairwise comparison. Thus, the fully nonparametric estimator is impractical in all but the simplest kinds of conjoint experiments.

More promising, instead, would be a model-based approach which explicitly models the majority indicator $M([ABC],[A'B'C'])$ as a function of the attributes. Here, we provide a sketch of one potential approach, leaving a fuller development for future work. We begin by noting that $\mathbb{E}_{\mathcal{V}}[Y_i([abc],[a'b'c'])] = \mathbb{E}_{\mathcal{V}}[Y_i \mid A = a, B = b, C = c, A' = a', B' = b', C' = c'] = \Pr(Y_i = 1 \mid A = a, B = b, C = c, A' = a', B' = b', C' = c')$ for any (a, b, c, a', b', c') in the support of $\mathcal{A}$ when the attributes are randomly assigned. Then, a model-based approach to this problem would begin by modeling the following, which we will denote as f for shorthand:

$$f(A, B, C, A', B', C') \equiv \Pr(Y_i = 1 | A, B, C, A', B', C').$$

This is a classical discrete choice problem in which the size of the choice set equals two (and hence it easily generalizes to forced choice tasks with more than two profiles), and we can employ a standard modeling strategy for discrete choice outcomes, such as the conditional logit model (McFadden, 1974).[11] This is akin to the approach to conjoint survey data traditionally used in marketing research (e.g., McFadden, 1986; Ben-Akiva, McFadden and Train, 2019).

Given the increased dimensionality of including the attributes from both profiles in the function, as well as modeling their interactions, it could be useful to additionally employ methods from statistical learning theory that have been developed to improve predictive performance in the face of potentially high-dimensional feature spaces. For instance, shrinkage penalties could be layered on top of generalized linear models (GLMs) and their multinomial extensions to model f using an

[11] For paired conjoints, we can also fit a model equivalent to the conditional logit via a binary logit regression of Y_i on the differences of the attributes (i.e., $A - A'$, $B - B'$, etc.)

elastic net regularized regression framework (e.g., Reid and Tibshirani, 2014).[12] Alternatively, f could be modeled using quasi-parametric learning approaches in place of GLMs, such as random forests, boosted trees, or neural nets (e.g., Prinzie and Van den Poel, 2008). Best practices in supervised learning theory (e.g. model training via cross-validation) would be vital, and researchers could allow both theory and cross-validation performance to guide their selection of a final model.

Once we obtain a high-performing predictive model f, it is straightforward to construct an estimate for the probability of winning of interest. First, given an estimated model $\hat{f}$, one can estimate the vote share for any profile $[abc]$ over any other profile $[a'b'c']$ using $\hat{f}(a, b, c, a', b', c')$. The majority classifier can then be obtained as $\hat{M}([ABC], [A'B'C']) = \mathbf{1}\{\hat{f}(A, B, C, A', B', C') > 0.5\}$, which will allow one to predict whether or not $[abc]$ would win over $[a'b'c']$ in a majority vote by the target population of voters, $\mathcal{V}$. Finally, one can estimate the expectation of M by averaging $\hat{M}$ over the distribution of the attributes corresponding to the target probability of winning. This final step is straightforward since the averaging is with respect to a known sub-distribution of the overall attribute distribution $\mathcal{A}$. To estimate the probability of a female candidate winning against a male candidate (i.e., equation (5)), for example, the following estimator can be used:

$$\sum_{b,c,b',c'} \Pr([ABC] = [abc], [A'B'C'] = [a'b'c'] | A = a, A' = a') \cdot \hat{M}([abc], [a'b'c']), \tag{6}$$

where the sum is taken over the possible values of B, C, B', and C' under the target attribute distribution $\mathcal{A}$, conditional on $A = a$ and $A' = a'$.

The procedure we have outlined so far represents a potentially viable approach to estimating the probability of winning with conjoint data. Unlike the estimation of the AMCE, however, the procedure involves the complex problem of modeling a high-dimensional response function, and thus care must be taken. In particular, validation of the final model is paramount. We provide remarks on the details of the validation procedure in Appendix A.1.

[12] For such penalized regressions, recent methods for estimating interactions between conjoint attributes (Egami and Imai, 2019) may also prove useful for improving predictive performance.

3.2 Fraction of Voters Preferring an Attribute

Another possible quantity of interest pertains to the fraction of voters who prefer attribute $A = a$ over $A = a'$. To construct a meaningful definition of this quantity of interest, we first need to define preferences over individual attributes (as opposed to profiles as a whole), which has not been necessary to this point. Drawing on our definition of preferences in Section 2.1, 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'*.[13] It is then easy to see 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 $\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])] > 0.5$.

Based on this definition, we can define as another possible quantity of interest the fraction of voters who prefer attribute $A = a$ over $A = a'$:

$$\mathbb{E}_{\mathcal{V}}\left[\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])] > 0.5\}\right]. \tag{7}$$

Note that this quantity does *not* equal the probability of winning defined in equation (5) since the order of the two expectations is reversed. Instead, the quantity amounts to first classifying all voters into those who (for example) prefer a female candidate and those who prefer a male candidate, and then calculating the proportion of the female preferers.

The distinction between the two quantities—the probability of winning and the fraction of voters preferring an attribute—is subtle but important, since equation (7) 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 election scholars. In fact, it is unlikely that the fraction of voters who prefer attribute $A = a$ over $A = a'$ is of much relevance to empirical scholars of elections because it tells us little about the importance of that attribute for actual observed voting behavior in multi-attribute contexts. As our handedness example above illustrated, it may well be that the vast majority of voters prefer a right-handed candidate, but this attribute would almost never be

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

relevant for determining the voting decision of any voter (so the AMCEs of that attribute or its effect on the probability of winning would be close to zero).

This arguably limited relevance applies also, and perhaps especially so, to a restricted version of this quantity of interest that has been proposed in other work. Specifically, Abramson, Koçak and Magazinnik (2019) propose a definition of individual attribute preference for attribute $A = a$ over $A = a'$ as $\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'BC])] > 0.5\}$, thereby considering only *ceteris paribus* comparisons (i.e. B and C are equal across the two profiles). Under this definition, the fraction of voters who prefer $A = a$ over $A = a'$ is simplified to

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

This version of the definition differs from that provided in equation (7) in that it is a function of preference relations only between pairs of profiles that are identical on all but one attribute. For example, consider the profile represented as [111] in the case of three binary attributes. This version only allows the profile to be compared against three out of the other seven possible profiles, i.e., [011], [101] and [110], which are each identical to the original profile in all but one of the three attributes. Preferences over other profiles—[100], [010], [001] and [000]—are ignored.[14]

In other words, this restricted definition of individual attribute preferences leaves a large number of profile pairs incomparable, which in our view makes it unsuitable for the analysis of conjoint survey experiments, where the goal often is precisely to analyze preferences about profiles that vary across multiple attributes simultaneously. To see the gravity of the problem, again consider the example of a paired forced-choice conjoint experiment with three binary attributes. Assuming the uniform independent randomization of the attributes (and disregarding the exact ties), the probability that a randomly generated pair results in a *ceteris paribus* comparison in which all attributes are equal save one is $3/7 \simeq .43$. That is, the expected proportion of conjoint tasks that provide *any* information about respondents' preferences per this restricted definition is only 43%, with the remaining 57% of the data contributing no information. Moreover, for a given

[14] The limitation of focusing on *ceteris paribus* comparisons is not readily apparent in the framework Abramson, Koçak and Magazinnik (2019) initially use for the proof of its main results, since the framework rules out any interaction between attributes by construction. Under the no-interaction assumption, if $\exists b, c$ such that $[1bc] \succ [0bc]$ then $[1b'c'] \succ [0b'c']$ for any $b' \in \{0,1\}$ and $c' \in \{0,1\}$, making consideration of all but one *ceteris paribus* comparison per attribute redundant. Although analytically convenient, this no-interaction assumption is unrealistically restrictive as a framework for voter preferences and therefore of limited utility to empirical scholars of elections.

individual attribute, only one out of seven comparisons ($\simeq 14\%$) is considered to contain information about respondents' preferences. The signal-to-noise ratio continues to decline rapidly as the number of attributes increases to more realistic settings, rendering most of the actual choice data "uninformative" by definition. With ten binary attributes, for example, only 10 out of 1023 pairs ($\leq 1\%$) are *ceteris paribus* and thus contain any information about respondents' preferences per this restricted definition. In contrast, all possible comparisons contribute some useful information about respondents' preferences according to the proposed definition in equation (7).

Defining preferences based exclusively on *ceteris paribus* comparisons is not only problematic for comparing profiles themselves, but also for understanding individual attribute preferences. To illustrate, consider a voter who chooses a male white Democratic candidate (e.g., [000]) over both a female white Republican candidate ([101]) and a male black Republican candidate ([011]). According to the restricted definition of individual attribute preference, these two choice outcomes contain *no* information about the voter's preference between a Democratic candidate and a Republican candidate, since neither is a *ceteris paribus* comparison with respect to party affiliation. In the real world, virtually no elections are about *ceteris paribus* contests between candidates; no two candidates or parties running for public office differ in just one way. Hence, based on the restricted definition of individual attribute preferences proposed by Abramson, Koçak and Magazinnik (2019), individual vote choices in almost all actual elections reveal no information about the voters' preferences about the candidates' attributes such as partisanship, race, or gender—an unacceptable starting point for most scholars of elections.

For these reasons, if researchers remain interested in analyzing the fraction of voters who prefer a particular attribute, we propose the quantity of interest defined by equation (7) over the restricted version defined by equation (8). Nonetheless, estimating the fraction of voters who prefer an attribute, whether defined by equation (7) or (8), presents even greater challenges than estimating the probabilities of winning, since it requires explicitly incorporating the heterogeneity of preferences among the voters in the analysis. That is, one would first need a good predictive model for the inner expectation term of the equation, $\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'B'C'])]$, which equals the average vote share of a profile containing $A = a$ versus another profile containing $A = a'$ for a specific voter i. Except in the unlikely event of the target population of voters being perfectly homogeneous or of no interactions between attributes, such a model would require (probably numerous) parameters representing how the effect of each individual attribute varied as a function of observed respondent characteristics (e.g., coefficients on interaction terms), in addition to the cross-attribute interaction terms already required for modelling the potential outcomes. The modeling exercise thus involves an even higher-dimensional prediction problem than the case of the probability of winning. In fact, if the researcher is truly interested in the fraction of voters who prefer candidates with a particular attribute (e.g. female) over another (e.g. male), a much simpler strategy might be to forgo a conjoint design altogether and instead ask directly whether respondents prefer a female candidate or a male candidate without explicitly mentioning the other attributes. This would obviously ignore the important multi-attribute nature of elections.

3.3 Revisiting the AMCE

The above discussion of alternative quantities of interest brings us back to the third desirable property of the AMCE: its empirical tractability. As detailed in Hainmueller, Hopkins and Yamamoto (2014), by virtue of the randomization of attributes, the AMCE can be nonparametrically identified via a simple difference in means with respect to the attribute of interest, much like a standard experiment with a single treatment. This is possible because the AMCE is a linear function of the potential outcomes, unlike the probability of winning or the fraction of voters preferring an attribute, which both involve a non-linear mapping.

Indeed, this discussion should be familiar to those well versed in causal inference methodology and the Average Treatment Effect (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 (Hainmueller, Hopkins and Yamamoto, 2014). When making statistical inferences about a causal quantity, one must face the fundamental problem of causal inference (Holland, 1986), or the problem of identifying counterfactual comparisons never directly observed in the data. As is well known, randomization of the treatment solves this problem for common causal estimands, such as the ATE, allowing for valid inference without further modeling or distributional assumptions. Less well known, however, is the fact that randomization solves the identification problem only for a certain class of causal estimands. Fortunately, this class of estimands includes some useful causal effects, such as the ATE, but excludes others, such as the *median* treatment effect, which represents the effect of the treatment on an individual unit who is at the median of the treatment effect distribution.[15]

This is analogous to the relationship between the AMCE and other alternative aggregations of treatment effects. Whereas the AMCE is nonparametrically identified by the observed difference in means by virtue of the random assignment of the treatments, quantities involving non-linear mappings such as the probability of winning require additional assumptions and/or more complicated modeling techniques. There is little wonder, then, that recent empirical applications of conjoint experiments have gravitated towards the estimation of AMCEs: they offer critical advantages over potential alternatives.

Have scholars shied away from randomized experiments because the ATE is not directly informative about whether the treatment effect is positive for a majority of units? Certainly not. Rather, 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 have a slight preference for its inverse, a'. This is true not only of the AMCE but also of the ATE; if a small number of people's 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. Like the ATE, the AMCE is an average, and so it necessarily combines directionality and intensity. This is indeed appropriate in many political applications: in a great many cases, a minority of people with intense preferences over a certain attribute can drive its electoral significance. And this is not merely a rhetorical point because; as we showed above, the AMCE identifies the difference in the expected vote shares.

4 Practical Recommendations

Our analysis has demonstrated that the AMCE can be used to recover meaningful quantities of interest for elections scholars. At the same time, our discussion also uncovers nuances in the interpretation of the AMCE and raises a caution against possible misinterpretations. In

[15] Note that this quantity is different from what is known in the literature as the (.5) *quantile treatment effect*. The latter refers to the difference between the medians of the two marginal potential outcome distributions, for which various methods have been developed (see Imbens and Wooldridge, 2009, Section 3.2 for a review).

this section, we provide practical guidance on what type of language applied researchers can use to summarize their empirical findings based on the AMCE estimates obtained from conjoint experiments.

There are at least two straightforward, accurate ways to describe AMCE estimates. First, consider the most generic case in which respondents choose between profiles (e.g. candidates, products, etc.) in a forced-choice design. Here, the AMCE can be described as the effect on the expected probability of preferring or choosing the profile when an attribute changes from one value to another (averaging over the randomization distribution of the profiles included in the conjoint). So, for example, one could say: "Changing the age of the candidate from young to old increases the probability of choosing the candidate profile by δ percentage points."

Second, consider the case in which researchers conduct a conjoint in an electoral context that involves choosing between candidates or parties. Here, the AMCE can also be interpreted as the effect on the expected vote share of the candidate or party when an attribute changes from one value to another (averaging over the randomization distribution of the profiles). So for example one could state: "Changing the age of the candidate from young to old increases the expected vote share of the candidate by δ percentage points." Thus, AMCEs in electoral conjoints allow applied researchers to make succinct, direct empirical statements about one of their core quantities of interest.

This simple language, of course, hides important nuances about the quantity of interest, which researchers should familiarize themselves with before applying the methodology. In particular, the difference in the expected vote share here specifically refers to the difference in the vote share that any candidate with a young age would on average obtain against an opponent versus the vote share that any candidate with an old age would on average obtain against an opponent, where the opponent is randomly drawn from the randomization distribution of the attributes (see Section 2.2). This language works similarly if there are multiple profiles, for example, in a conjoint with three candidate profiles.

Needless to say, the usual caveats about interpreting survey experiments apply: one needs to exercise caution when the goal is to extrapolate empirical findings from survey experiments to actual behavioral outcomes in real elections (but see Hainmueller, Hangartner and Yamamoto, 2015; Auerbach and Thachil, 2018). In addition, researchers should keep in mind that the AMCE averages the effect of an attribute over two different distributions: the randomization distribution of the other attributes and the distribution of respondents. The sampling strategy and the experimental design should therefore be informed by the target distributions (i.e., $\mathcal{A}$ and $\mathcal{V}$ as defined in Section 2.4) to which researchers want to make an inference about (Hainmueller, Hopkins and Yamamoto, 2014; de la Cuesta, Egami and Imai, 2019). In addition, subgroup analysis can be helpful to examine how the AMCEs may depend on a particular group of respondents or choice of the attribute distributions (see Bansak et al., forthcoming, for advice on conjoint design).

5 Concluding Remarks

In this paper, we employed a general framework for analyzing voter preferences in electoral conjoint experiments with multiple candidate/party attributes to study the microfoundations of the AMCE, a frequent estimand in recent applications of conjoint experiments in political science. A key result that emerges is that, as long as voters have a preference ranking over the set of multi-attribute candidate/party profiles and vote for their preferred profiles, the AMCE directly recovers a core quantity of interest to election scholars: the effects of candidate or party attributes on their expected vote shares for elections that mirror the conjoint design. Importantly, this crucial property of the AMCE holds regardless of the structure of voter preferences or the electoral formulae used to map votes into seats. In addition, we explored several other possible quantities of interest to electoral scholars in the context of conjoint experiments and discussed possible estimation strategies. This exercise further demonstrates the theoretical and practical advantages of the AMCE. Finally, we also provided practical guidance on interpreting AMCEs for researchers applying conjoint experiments.

Our study has several implications. First, our results highlight the essential role of the AMCE for analyzing elections using conjoint experiments. In contrast to a recent critique that suggested that AMCEs are largely uninformative with respect to questions of interest to political scientists (Abramson, Koçak and Magazinnik, 2019), our results demonstrate that AMCEs are in fact of fundamental importance for scholarship on elections. AMCEs—under general conditions—identify the effects of changes in attributes on the expected vote shares of candidates or parties. And as our literature review has shown, vote shares are the central outcome of interest for much of the literature on elections and voting. The bottom line for applied scholars is simple: if one is interested in effects of candidate or party attributes on vote shares, then 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 do not rely on arbitrary functional form assumptions.

Second, by going beyond AMCEs, our study highlights that conjoint experiments can also be informative about other, less widely used causal quantities of interest for studying elections. In particular, we have defined several estimands that capture the effects of changes in attributes on the probability of winning and sketched procedures for their estimation. This revealed that it is important to precisely define what is being compared when considering relative probabilities of winning and also that such estimation requires additional modeling assumptions that go beyond those guaranteed by the randomization. We have also examined a quantity that relates to the fraction of voters who prefer a specific attribute, the focus of Abramson, Koçak and Magazinnik (2019). We highlighted how this quantity is rarely informative in multi-attribute elections while also being harder to estimate and involving a higher-dimensional modeling problem than the case of estimating the probability of winning. In other words, Abramson, Koçak and Magazinnik not only ask the ill-posed question of whether one estimator (that for the AMCE) recovers a quantity it was not designed for, but also do so for a quantity rarely informative about actual vote choices in multi-attribute elections. Overall, our analysis has revealed that the AMCE, thanks to its ease-of-use and clear interpretability, has many advantages over these alternative estimands.

Third, our study points to some fruitful avenues for future research. We have proposed procedures for estimating alternative quantities of interest related to candidates'/parties' probability of winning elections/seats that we hope may serve as a starting point for future research into possible modeling approaches. If researchers are willing to make additional assumptions, such approaches could be used to get even more mileage out of conjoint data than what is currently employed in applied work.

References

Abrajano, Marisa A., Christopher S. Elmendorf and Kevin M. Quinn. 2015. “Using Experiments to Estimate Racially Polarized Voting.”. UC Davis Legal Studies Research Paper Series, No. 419.

Abramson, Scott F., Korhan Koçak and Asya Magazinnik. 2019. “What Do We Learn About Voter Preferences From Conjoint Experiments?”. Working paper presented at PolMeth XXXVI.

Adida, Claire L, Adeline Lo and Melina Platas. 2017. “Engendering empathy, begetting backlash: American attitudes toward Syrian refugees.”.

Arnesen, Sveinung, Dominik Duell and Mikael Poul Johannesson. 2019. “Do citizens make inferences from political candidate characteristics when aiming for substantive representation?” Electoral Studies 57:46–60.

Auer, Daniel, Giuliano Bonoli, Flavia Fossati and Fabienne Liechti. 2019. “The matching hierarchies model: evidence from a survey experiment on employersâ hiring intent regarding immigrant applicants.” International migration review 53(1):90–121.

Auerbach, Adam Michael and Tariq Thachil. 2018. “How Clients Select Brokers: Competition and Choice in India’s Slums.” American Political Science Review 112(4):775–791.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2019. “Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments.” Political Science Research and Methods Forthcoming.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. forthcoming. Conjoint Survey Experiments. In The Cambridge Handbook of Advances in Experimental Political Science, ed. James N. Druckman and Donald P. Green. Cambridge, MA: Cambridge University Press.

Bansak, Kirk, Jens Hainmueller and Dominik Hangartner. 2016. “How economic, humanitarian, and religious concerns shape European attitudes toward asylum seekers.” Science 354(6309):217–222.

Bechtel, Michael M., Federica Genovese and Kenneth F. Scheve. 2016. "Interests, Norms, and Support for the Provision of Global Public Goods: The Case of Climate Cooperation." British Journal of Political Science Forthcoming.

Ben-Akiva, Moshe, Daniel McFadden and Kenneth Train. 2019. "Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis." Foundations and Trends in Econometrics 10(1-2):1–144.

Carnes, Nicholas and Noam Lupu. 2016. "Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class." American Political Science Review 110(4):832–844.

Clayton, Katherine, Jeremy Ferwerda and Yusaku Horiuchi. 2019. "Exposure to Immigration and Admission Preferences: Evidence From France." Political Behavior .

Coughlin, Peter J. 1992. Probabilistic voting theory. Cambridge University Press.

Crowder-Meyer, Melody, Shana Kushner Gadarian, Jessica Trounstine and Kau Vue. 2018. "A Different Kind of Disadvantage: Candidate Race, Cognitive Complexity, and Voter Choice." Political Behavior pp. 1–22.

de la Cuesta, Brandon, Naoki Egami and Kosuke Imai. 2019. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution.". Working paper presented at PolMeth XXXVI.

Egami, Naoki and Kosuke Imai. 2019. "Causal interaction in factorial experiments: Application to conjoint analysis." Journal of the American Statistical Association 114(526):529–540.

Enelow, James M and Melvin J Hinich. 1989. "A general probabilistic spatial theory of elections." Public Choice 61(2):101–113.

Flores, René D and Ariela Schachter. 2018. "Who Are the âIllegalsâ? The Social Construction of Illegality in the United States." American Sociological Review 83(5):839–868.

Franchino, Fabio and Francesco Zucchini. 2014. "Voting in a Multi-dimensional Space: A Conjoint Analysis Employing Valence and Ideology Attributes of Candidates." Political Science Research and Methods pp. 1–21.

Gampfer, Robert, Thomas Bernauer and Aya Kachi. 2014. "Obtaining public support for North-South climate funding: Evidence from conjoint experiments in donor countries." Global Environmental Change 29:118–126.

Goggin, Stephen N, John A Henderson and Alexander G Theodoridis. 2019. "What goes with red and blue? Mapping partisan and ideological associations in the minds of voters." Political Behavior pp. 1–29.

Hainmueller, Jens, Daniel J Hopkins and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." Political Analysis 22(1):1–30.

Hainmueller, Jens, Dominik Hangartner and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-world Behavior." Proceedings of the National Academy of Sciences 112(8):2395–2400.

Hankinson, Michael. 2018. "When do renters behave like homeowners? High rent, price anxiety, and NIMBYism." American Political Science Review 112(3):473–493.

Hanretty, Chris, Benjamin E. Lauderdale and Nick Vivyan. 2020. "A Choice-Based Measure of Issue Importance in the Electorate." American Journal of Political Science forthcoming.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12470*

Holland, Paul W. 1986. "Statistics and causal inference." Journal of the American statistical Association 81(396):945–960.

Horiuchi, Yusaku, Daniel M. Smith and Teppei Yamamoto. 2018. "Measuring Voters' Multidimensional Policy Preferences with Conjoint Analysis: Application to Japan's 2014 Election." Political Analysis 26(2):190–209.

Imbens, Guido W. and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." Journal of Economic Literature 47(1):5–86.
URL: *https://www.aeaweb.org/articles?id=10.1257/jel.47.1.5*

Jenke, Libby, Kirk Bansak, Jens Hainmueller and Dominik Hangartner. Forthcoming. "Using Eye-Tracking to Understand Decision-Making in Conjoint Experiments." Political Analysis .

Kirkland, Patricia A and Alexander Coppock. 2018. "Candidate choice without party labels." Political Behavior 40(3):571–591.

Leeper, Thomas J, Sara B Hobolt and James Tilley. forthcoming. "Measuring Subgroup Preferences in Conjoint Experiments." Political Analysis .

Lindbeck, Assar and Jörgen W Weibull. 1987. "Balanced-budget redistribution as the outcome of political competition." Public choice 52(3):273–297.

Loewen, Peter John, Daniel Rubenson and Arthur Spirling. 2012. "Testing the Power of Arguments in Referendums: A Bradley–Terry Approach." Electoral Studies 31(1):212–221.

Matsuo, Akitaka and Seonghui Lee. 2018. "Multi-dimensional policy preferences in the 2015 British general election: A conjoint analysis." Electoral studies 55:89–98.

McFadden, Daniel. 1986. "The choice theory approach to market research." Marketing science 5(4):275–297.

McFadden, Daniel L. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In Frontiers in Econometrics, ed. P. Zarembka. New York: Academic Press pp. 105–142.

Mummolo, Jonathan and Clayton Nall. 2016. "âWhy Partisans Donât Sort: The Constraints on Political Segregationâ." The Journal of Politics Forthcoming.

Ono, Yoshikuni and Barry C Burden. 2019. "The contingent effects of candidate sex on voter choice." Political Behavior 41(3):583–607.

Prinzie, Anita and Dirk Van den Poel. 2008. "Random forests for multiclass classification: Random multinomial logit." Expert systems with Applications 34(3):1721–1732.

Reid, Stephen and Rob Tibshirani. 2014. "Regularization paths for conditional logistic regression: the clogitL1 package." Journal of statistical software 58(12).

Ryan, Timothy J. and J. Andrew Ehlinger. 2019. "Issue Publics: Fresh Relevance for an Old Concept.". Working paper presented at the Annual Meeting of the American Political Science Association, August 2019, Washington, DC.

Schachter, Ariela. 2016. "From âdifferentâ to âsimilarâ an experimental approach to understanding assimilation." American Sociological Review 81(5):981–1013.

Schofield, Norman. 2007. The spatial model of politics. Routledge.

Stokes, Leah C and Christopher Warshaw. 2017. "Renewable energy policy design and framing influence public support in the United States." Nature Energy 2(8):17107.

Taagepera, Rein and Matthew Soberg Shugart. 1989. Seats and Votes. New Haven, CT: Yale University Press.

Teele, Dawn Langan, Joshua Kalla and Frances Rosenbluth. 2018. "The Ties That Double Bind: Social Roles and Women's Underrepresentation in Politics." American Political Science Review 112(3):525–541.

Train, Kenneth E. 2009. Discrete choice methods with simulation. Cambridge university press.

Tsai, Tsung-han. 2017. "A balance between candidate-and party-centric representation under mixed-member systems: The evidence from voter behavior in Taiwan." Electoral Studies 49:17–25.

Wright, Matthew, Morris Levy and Jack Citrin. 2016. "Public Attitudes Toward Immigration Policy Across the Legal/Illegal Divide: The Role of Categorical and Attribute-Based Decision-Making." Political Behavior 38(1):229–253.

Appendix

A.1 Details on the Estimation of the Probability of Winning

Due to the model dependence of the procedure for estimating the probability of winning described in Section 3.1, validation of the final model is paramount. There is of course no reason to believe, nor do we even need to assume, that the final fitted model perfectly represents the true underlying data-generating process. After all, the purpose of these procedures is not to estimate model-specific parameters that themselves are meant to represent particular estimands of interest. Instead, the goal is to learn a model $\hat{f}$ that produces good predictions such that $\hat{f}(a, b, c, a', b', c') \approx f(a, b, c, a', b', c')$. Model validation and evaluation can thus proceed according to standard best practices in machine learning and statistical learning theory, making use of performance metrics that are a function of out-of-sample or cross-validation predictions and the corresponding true outcome values.

Given the focus on estimating the probability of winning, one's first instinct might be to simply compute the out-of-sample or cross-validation classification accuracy of $\hat{M}$. However, while classification accuracy would be informative, it would be insufficient and potentially misleading in terms of the usefulness of the model for predicting the majority-vote outcomes of matchups. For instance, consider a profile matchup $([abc], [a'b'c'])$ where the true average vote share is 0.55 (i.e. 55% of the population of interest would choose $[abc]$ over $[a'b'c']$). In this case, even if one had perfectly modeled f and had data on this matchup for the entire population, the classification accuracy of $\hat{M}$ at the individual level would be 0.55. This is an underwhelming classification accuracy, but it does not suggest a poorly trained model for our purposes; quite to the contrary, a perfect model would exhibit a classification accuracy of 0.55 if applied to randomly sampled voters' evaluations of this matchup.

In other words, the focus on predicting the outcome of a matchup at the aggregate level (i.e. which of two candidate profiles would win the majority of votes among a population of interest) means that the classification accuracy of $\hat{M}$ at the individual level (i.e. whether or not $\hat{M}$ accurately predicts a randomly sampled individual's vote $\{0, 1\}$ for a particular matchup) is neither of primary interest nor necessarily even indicative of the quality of the model $\hat{f}$. Since estimates of $M([ABC], [A'B'C'])$ must necessarily happen at some level of aggregation, validation/evaluation of the model must also occur at some level of aggregation. Calibration analysis methods from statistical and machine learning are well-suited for this purpose.

Calibration analysis is a method of assessing the reliability of predicted probabilities. In an ideal world, one would have a perfectly specified and fitted model and hence its predicted probabilities would equal the true probabilities. This is of course not possible in reality, but we may still hope that the predicted probabilities closely approximate the true probabilities. However, empirically assessing this at the individual level is impossible because underlying probabilities are never truly observed. In addition, the true underlying vote share for any matchup $([abc], [a'b'c'])$ is also unobserved given the dimensionality of the feature space and randomization of the attributes. However, the reliability of a model's predicted probabilities can still be (partly) assessed by aggregating the data into bins.

Specifically, for each data point (i.e. each observed matchup evaluation), $\hat{f}$ would be applied to formulate a cross-validated predicted probability, and those predicted probabilities would then be binned into intervals (e.g. 20 intervals of length 0.05 from 0 to 1). Within each bin, the average predicted probability would be computed and compared against the true fraction of 1's in the data points belonging to that bin. Predicted probability averages that are approximately equal to the true fraction of 1's in each bin would be evidence of a well-calibrated model. This would then provide support, albeit not definitive, to the claim that $\hat{f}(A, B, C, A', B', C')$ is meaningfully approximating $E_{\mathcal{V}}[Y_i([ABC], [A'B'C'])]$, in which case it would then be possible to provide reliable estimates of $M([ABC], [A'B'C'])$.

Note also that if one's focus is solely on predicting the ultimate election outcome in a particular matchup, with no additional interest in accurately estimating the vote share, it need only be the case that for any matchup whose vote share is above 0.5, the estimate of the vote share (i.e. predicted probability) is also above 0.5. What that means is one does not necessarily need a model that is calibrated along the entire $[0, 1]$ interval. Instead, it would be sufficient to have, for instance, a model's whose calibration curve hits the identity line at 0.5 and is otherwise monotonically increasing, which is a strictly easier condition for classification models to satisfy.