

# What Do We Learn About Voter Preferences From Conjoint Experiments?

By SCOTT F ABRAMSON, KORHAN KOCAK, & ASYA MAGAZINNIK\*

*Political scientists frequently interpret the results of conjoint experiments as reflective of majority preferences. In this paper we show that the target estimand of conjoint experiments, the AMCE, is not well-defined in these terms. Even with individually rational experimental subjects, unbiased estimates of the AMCE can indicate the opposite of the true preference of the majority. To show this, we characterize the preference aggregation rule implied by AMCE and demonstrate its several undesirable properties. With this result we provide a method for placing bounds on the proportion of experimental subjects who prefer a given candidate-feature. We provide a testable assumption to show when the AMCE corresponds in sign with the majority preference. Finally, we offer a structural interpretation of the AMCE and highlight that the problem we describe persists even when a model of voting is imposed.*

Conjoint experiments have become a standard part of the political scientist's toolkit. Across the top scholarly journals, political scientists regularly interpret the results of these experiments to make empirical claims about both majority preferences and electoral outcomes. In this paper we show that the target estimand of conjoint experiments, the average marginal component effect (AMCE), does not support such claims. We do so by characterizing the preference aggregation rule implied by the AMCE and, in line with well known results (Arrow, 1950; Gibbard, 1973), demonstrate its undesirable properties for making inferences about voter preferences and electoral outcomes.

\* Abramson: Assistant Professor, Department of Political Science, University of Rochester, email: sabramso@ur.rochester.edu; Kocak: PhD Candidate, Department of Politics, Princeton University, email: kkocak@princeton.edu; Magazinnik: Instructor, Department of Political Science, Massachusetts Institute of Technology, email: asyam@mit.edu. Kocak gratefully acknowledges the support of the Research Program in Political Economy at Princeton University. The authors thank Naoki Egami, Matias Iaryczower, Kosuke Imai, Nolan McCarty, and seminar audiences at the APSA Annual Meeting, the 2019 Conference of the Society for Political Methodology, and the 2019 Toronto Political Behavior Workshop for useful comments and encouragement.

The goal of factorial designs like those in forced-choice conjoint experiments is to mimic the comparisons individual voters make at the ballot box. By randomizing a large number of candidate and platform features, researchers seek to construct realistic approximations of the choices voters face. With a simple difference-in-means or least-squares regression researchers compare the attributes of candidates most frequently chosen to the attributes of the candidates least frequently chosen to make empirical claims about the preferences of voters.

For example, experimental results from conjoint experiments are used to make claims about voters' preferences for particular policies like: "Americans express a pronounced preference for immigrants who are well educated, are in high-skilled professions, and plan to work upon arrival" (Hainmueller and Hopkins, 2015), and "[there is] strong evidence for progressive preferences over taxation among the American public" (Ballard-Rosa, Martin and Scheve, 2017). Even more frequently, conjoint results are used to make statements about candidates for elected office like: "voters prefer experienced or locally born politicians, but do not prefer politicians affiliated with a major political party... and are indifferent with regard to dynastic family ties and gender" (Horiuchi, Smith and Yamamoto, 2018), and "voters and legislators do not seem to hold female candidates in disregard; all else equal, they prefer female to male candidates" (Teele, Kalla and Rosenbluth, 2018).

Put simply, political scientists use conjoint results to make statements about a *binary preference relation* for a representative voter in the context of elections. Moreover, researchers interpret findings from conjoints as evidence that candidates with particular features are most preferred and thereby more likely to win elections (Carnes and Lupu, 2016; Teele, Kalla and Rosenbluth, 2018). This common interpretation has even migrated to the public discourse. CBS News and POLITICO, for example, have both highlighted results from conjoint experiments, asserting that the "[Democratic] party's primary voters prefer female candidates of color in 2020" (Magni and Reynolds, 2019) and that [Democratic] "voters showed a clear preference for females, all else equal" (Khanna, 2019). By

way of example and formal proof, we show that the AMCE does not support empirical claims about winners of electoral contests.

The AMCE is defined as the average effect of varying one attribute of a candidate profile, e.g. the race or gender of the candidate, from  $A$  to  $A'$ , on the probability that the candidate will be chosen by a respondent, where the expectation is defined over the distribution of the other attributes. To be clear, we do not dispute that the estimators proposed by Hainmueller, Hopkins and Yamamoto (2014) for this quantity are unbiased under their assumptions. Rather, we show that even when their assumptions hold, a positive AMCE of candidate-feature  $A$  over  $A'$  does not indicate: 1.) A majority of voters prefer candidates with feature  $A$  to those with  $A'$ ; 2.) all else equal, a random voter prefers candidates with  $A$  to those with  $A'$ ; nor 3.) candidates with feature  $A$  beat candidates with feature  $A'$  in most elections.

This occurs because the AMCE averages over two aspects of individual preferences: their direction (whether or not an individual prefers  $A$  to  $A'$ ) and their intensity (how much they prefer  $A$  to  $A'$ , or how much they care about the issue). Because the AMCE produces a literally average voter, it assigns greater weight to voters who intensely prefer a particular outcome, the consequence of which can be inaccurate out-of-sample predictions about winners of elections. For example, a large majority of people may have a strict preference for male candidates over female candidates, but the AMCE can, nevertheless, be positive for female candidates if there is a small minority of voters who have an intense preference for women. Far from being a theoretical abstraction, we provide empirical evidence that this structure, where the direction and intensity of preferences are correlated, undergirds public opinion across a large swath of important political questions.

Our point is not merely semantic. In the field of market research, where the tools of conjoint experiments were first developed, scholars are typically interested in the demand for a given product, which is determined by both the intensive and extensive margins of consumer choice. By contrast,

political scientists typically care about elections, which are won on the extensive margin. Indeed, outside of fantastical institutional designs (e.g., Lalley and Weyl (2018)), electoral contests are not swayed by *how much* a subset of voters prefer a given candidate but, rather, *how many voters* prefer each candidate. By averaging over both margins of choice, the AMCE can prove largely uninformative with respect to many questions of interest to political scientists.

Since the objective of conjoint experiments is to construct a mapping from individual to aggregate preferences, we build on the literature in positive political theory that formally evaluates mechanisms that do just that. That is, we characterize the AMCE as a preference aggregation rule—a mapping from individual to aggregate preferences (Austen-Smith and Banks, 2000, p. 26). We use results from this exercise to provide a method that, for a given AMCE estimate, allows researchers to place bounds on the proportion of experimental subjects that maintain a strict preference for a candidate-feature. Using this method, we reevaluate the findings of every conjoint experiment published in the *American Political Science Review* and the *Journal of Politics* between 2016 and 2019 and show that, with one exception, their results are consistent with either a majority or minority of respondents holding a strict preference for the candidate-feature that yielded each study’s largest estimated effect.

Next, we describe sufficient conditions under which the AMCE indicates a majority preference. We show that if the direction and intensity of voters’ preferences are uncorrelated, then the sign of the AMCE gives the preference of the majority. We explore the plausibility of this assumption using survey respondents’ stated preferences in the 2016 ANES. For 17 of the 22 items for which measures of both intensity and direction exist, we find strong evidence that they are correlated, suggesting that in many, if not most, areas of politics the AMCE can fail to indicate a majority preference.

Finally, we explore the relationship between the AMCE and a simple model of voting. In providing a structural interpretation of the AMCE we show that it reflects an average of individual ideal points over candidate-features. This highlights how conjoints combine information about both the intensity

and direction of preferences and demonstrates the need to impose additional structure in order to obtain estimates of theoretically relevant quantities of interest. We conclude with some directions for future research on how to make conjoints more informative about voter preferences.

## I An Example

### *Part I: A Positive AMCE Does Not Imply a Majority Preference*

To start, we work through a toy example of how the AMCE aggregates preferences. We aim to make as few assumptions about the underlying preferences of individual voters as possible. While we view our assumptions as benign, we note that if the AMCE exhibits undesirable properties under these assumptions, placing even less structure will not rectify whatever problems we identify and only obscure what drives these results. Furthermore, we emphasize that we are agnostic about the content of voters' preferences. Individuals may be self-interested, other-regarding, or some mixture thereof. *We impose only that individual preferences are complete and transitive; furthermore to vastly simplify the presentation we rule out indifference as is standard in the social choice literature.*<sup>1</sup> Without these assumptions we can learn about neither individual nor aggregate preferences. As such, these are the minimal assumptions about individual preferences we can make and still hope to recover meaningful insight into the AMCE.

Since, fundamentally, the object researchers seek to describe concerns a preference relation over candidate-features, the primitives we begin with are over these features. For simplicity, consider an electorate of five voters (V1, V2, V3, V4, V5), whose preferences over candidates we would like to study with a conjoint experiment. To sidestep concerns about estimation, suppose we can fully observe every potential choice between candidates made by every member of this population. In this world, there are two attributes of candidates that are important to voters: their gender (female or

<sup>1</sup>Formally, completeness is defined as  $x \succeq y$ ,  $y \succeq x$  or both, and transitivity is defined as if  $x \succeq y$  &  $y \succeq z$ , then  $x \succeq z$ . We define a *strict preference relation* as  $x \succ y$  if and only if  $x \succeq y$  and  $y \not\succeq x$  and henceforth refer to it when we write preference.

male) denoted by  $G \in \{F, M\}$ , and their party (Democrat or Republican) denoted  $P \in \{D, R\}$ . Each candidate is an ordered pair of gender and age, so that there are four different candidate profiles:  $FD, FR, MD$ , and  $MR$ . The voters' preferences over attributes are given in the following table:

V1	V2	V3	V4	V5
$M \succ F$	$M \succ F$	$M \succ F$	$F \succ M$	$F \succ M$
$R \succ D$	$R \succ D$	$R \succ D$	$D \succ R$	$D \succ R$

TABLE 1—PREFERENCES OVER ATTRIBUTES

It can easily be seen that a majority of voters prefer male candidates to female candidates, and a majority of voters prefer Republican candidates to Democratic candidates.

We construct preferences over candidates from preferences over attributes in the following way: Voters prefer candidates that have both of the attributes they like to those that have one attribute they like, which in turn they prefer to candidates who have neither of the attributes they like. Notice that there are two types of candidates that have only one attribute that matches a voter's preference. For these candidates, whether a voter prefers one or the other depends on which attribute the voter places a greater weight on. For example, if a voter places more weight on gender, we would expect them to choose a candidate who has their preferred gender but not their preferred party over a candidate who has the voter's preferred party but not the gender.

Formally, such preferences over candidate profiles can be written as the lexicographic preference relation  $\gg$ , where for each voter one attribute is given a greater weight in determining the preference ordering. Accordingly, we assume that voters 1, 2, and 3 place more weight on the candidate's party,  $P \gg G$ , whereas voters 4 and 5 place more on the candidate's gender,  $G \gg P$ . Combining weights with preferences over attributes, we can produce voters' preferences over candidate profiles. These are presented in Table 2.

Given these preferences, in Table 3 we present the votes candidates would obtain in each head-to-

Rank	V1	V2	V3	V4	V5
1.	<i>MR</i>	<i>MR</i>	<i>MR</i>	<i>FD</i>	<i>FD</i>
2.	<i>FR</i>	<i>FR</i>	<i>FR</i>	<i>FR</i>	<i>FR</i>
3.	<i>MD</i>	<i>MD</i>	<i>MD</i>	<i>MD</i>	<i>MD</i>
4.	<i>FD</i>	<i>FD</i>	<i>FD</i>	<i>MR</i>	<i>MR</i>

TABLE 2—PREFERENCES OVER CANDIDATE PROFILES

head election for every possible pairwise comparison. Note that in this example men win three of the four elections when they face off against a woman and four of the six total contests (the winner is bolded in the first column).

Comparison	V1	V2	V3	V4	V5	Tally
<b>MR</b> ,FR	MR	MR	MR	FR	FR	3, 2
<b>MR</b> ,FD	MR	MR	MR	FD	FD	3, 2
<b>MR</b> ,MD	MR	MR	MR	MD	MD	3, 2
MD, <b>FR</b>	FR	FR	FR	FR	FR	0, 5
<b>MD</b> ,FD	MD	MD	MD	FD	FD	3, 2
<b>FR</b> ,FD	FR	FR	FR	FD	FD	3, 2

TABLE 3—AGGREGATE PREFERENCES OVER CANDIDATE PROFILES

In this simple setting, the AMCE is derived as in Hainmueller, Hopkins and Yamamoto (2014), Proposition 3. The intuition behind the comparisons being made when estimating the AMCE is given in Table 4. Here,  $\bar{Y}(C_1, C_2)$  denotes the number of votes candidate  $C_1$  obtains when running against candidate  $C_2$ . For each contest we can obtain  $\bar{Y}$  from the last column of Table 3. To obtain the AMCE for males we compare how male candidates (column 1) fare relative to female candidates (column 2) when they run against the same opponent, then sum this difference over all possible opponents. This sum is finally normalized by the number of possible profiles minus one (3) times the number of possible profiles with a given gender (2) times the number of voters (5). The procedure yields an AMCE for male equal to  $-1/15$ , meaning that the average probability of being chosen is higher for female candidates than it is for male candidates.

Our toy example illustrates the intuition driving our main result. Notice that the AMCE for men

1.	2.			
$\bar{Y}(MR, MD)$	–	$\bar{Y}(FR, MD)$	=	–2
$\bar{Y}(MR, FD)$	–	$\bar{Y}(FR, FD)$	=	0
$\bar{Y}(MR, MR)$	–	$\bar{Y}(FR, MR)$	=	1/2
$\bar{Y}(MR, FR)$	–	$\bar{Y}(FR, FR)$	=	1/2
$\bar{Y}(MD, MD)$	–	$\bar{Y}(FD, MD)$	=	1/2
$\bar{Y}(MD, FD)$	–	$\bar{Y}(FD, FD)$	=	1/2
$\bar{Y}(MD, MR)$	–	$\bar{Y}(FD, MR)$	=	0
$\bar{Y}(MD, FR)$	–	$\bar{Y}(FD, FR)$	=	–2
				–2
(# of profiles – 1) × (# of voters)				= 30
× (# of profiles with a given gender )				
AMCE				= –1/15

TABLE 4—OBTAINING THE AMCE

is *negative*, and yet we know that by construction a majority of the voters prefer male to female candidates and that men will beat women in a majority of head-to-head elections. Holding all else constant (in the case of this example, party), a male candidate would always win.<sup>2</sup> Furthermore, men win more electoral contests. The AMCE produces an estimate that indicates the opposite of the true majority preference because the minority, who place the greatest weight on the gender dimension, also have a preference for female candidates, while the majority, who prefer men, do not place much weight on gender when making their decisions. When aggregating preferences over gender, the AMCE mechanically assigns greater weight to the minority that strongly prefer women.

Crucially, this result is a feature of the target estimand and is not a problem of estimation. Our example is analogous to a survey in which each respondent is asked to evaluate all possible head-to-head comparisons. To highlight this, we conduct a simulation exercise where we run a three question conjoint experiment on a population characterized by the distribution of voter preferences in our toy example. That is, we take a population of five voters with the preferences detailed in Table 3. Then, we randomly construct pairs of candidates, randomizing their gender and age. Knowing voter

<sup>2</sup>Note that in rows 1 and 5 of Table 3 the male candidate gets three votes against the female candidate's two.



preferences for candidate profiles we then obtain a winner in each contest and estimate the AMCE for male candidates. In Figure 1 we present results from conducting this exercise 1,000 times. Of course, because the estimator is unbiased, the effect is centered on  $-1/15$ , despite being generated from a population of voters where  $3/5$  prefer men.

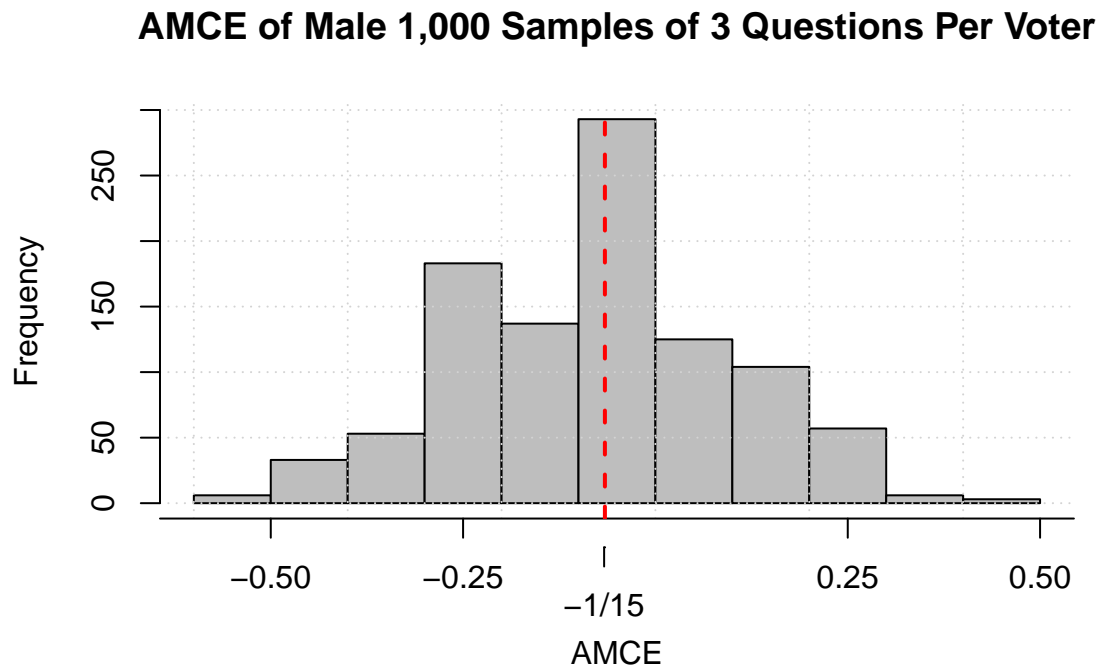


FIGURE 1. THIS FIGURE PRESENTS AN EXERCISE WHERE WE CONDUCT 1,000 CONJOINT EXPERIMENTS ON A POPULATION OF FIVE VOTERS WITH PREFERENCES AS DETAILED IN TABLE 3 AND WHERE CANDIDATES ARE RANDOMLY GENERATED FROM A COMBINATION OF GENDER AND AGE.

*Part II: The AMCE Is Not Robust to the Inclusion of Additional Treatments*

Consider the same population of five voters. However, instead of conducting an experiment where we randomize only party and gender, we now include a third feature, race, which for simplicity takes on only two values, black or white. Denote this  $R \in \{B, W\}$ . Let voters 1, 2, and 3 have the preference  $W \succ B$  and voters 4 and 5 have the preference  $B \succ W$ . Furthermore, let voters 1, 2,

and 3 place the greatest weight on party, the second most on gender, and least weight on race, i.e.  $P \gg G \gg R$ , and let voters 4 and 5 place the greatest weight on race, the next most on gender, and the least on party, i.e.  $R \gg G \gg P$ .

As in the previous section, with this combination of weights and preferences for features we can produce a full ranking of candidate profiles. Voters most prefer candidates with all three of their preferred features and least prefer those with none of their preferred features. Among candidates that have two of the three features voters prefer, they rank candidates with their first and second most preferred feature first, first and third most preferred features second, and second and third most preferred features third. Finally, we assume that voters prefer all candidates with two preferred features to all candidates with just one feature. Preferences over candidates are given in Table 5.

Rank	V1	V2	V3	V4	V5
1.	MRW	MRW	MRW	FDB	FDB
2.	MRB	MRB	FRW	FRB	FRB
3.	FRW	FRW	MDW	MDB	MDB
4.	MDW	MDW	FRB	FDW	FDW
5.	FRB	FRB	MDB	MRW	MRW
6.	MDB	MDB	FDW	FRW	FRW
7.	FDW	FDW	FDB	MDW	MDW
8.	FDB	FRB	FRW	MRW	MRW

TABLE 5—PREFERENCES OVER CANDIDATE PROFILES - EXAMPLE PART II

Again, with these preferences and weights we can fully characterize vote shares in each possible electoral contest. These are presented in Table B2 in the supplemental appendix.

Since these are the same exact voters from the previous example, their preferences with respect to gender have not changed: 3/5 of them prefer men to women. As before, in this example men win a large majority of elections.<sup>3</sup> However, by contrast to our previous example, instead of always ranking female candidates above male candidates, voters 4 and 5 will now sometimes vote for a man.

<sup>3</sup>Thirteen of the sixteen elections where men run head to head against women are won by the male candidate as are nineteen of the twenty-eight overall contests.

Because they place more weight on race than gender, in some contests, voters 4 and 5 will be willing to accept a man even though, all else equal, they prefer women. Since including race changes the relative ranking of male and female candidates, it changes the AMCE researchers would derive from this experiment.

As before, we follow Hainmueller, Hopkins and Yamamoto (2014) Proposition 3 to calculate the AMCE. In Table B3 we show that it is equal to  $1/14$ , yielding the exact opposite substantive result compared to the previous experiment where we considered only gender and party. That is, with the same set of experimental subjects, just by adding an additional feature we can flip the sign of the AMCE. This highlights our second main result: the sign and magnitude of the AMCE depend upon the features included in the experimental design even though individual preferences over these features remain constant across experiments. This occurs because the inclusion of the additional feature changes the relative rankings of candidates with respect to other unrelated, but potentially theoretically important, attributes.

In other words, even with identical subjects, the results researchers obtain from conjoint experiments depend upon the specific set of features included in their experimental design. In the supplemental appendix we describe an example that highlights a variant of this problem that is relevant for applied researchers. There, we show that the exclusion of “unrealistic” feature-combinations alters the AMCE for the same reason as above. For instance, the exclusion of uneducated doctors in a candidate-choice experiment where education, occupation, and gender are randomized changes the AMCE a researcher would obtain for male candidates. This occurs because the exclusion of uneducated doctors can alter the relative ranking of female and male candidates, in turn changing the AMCE we would obtain from this restricted randomization relative to a uniform randomization of features.

## II The AMCE as a Preference Aggregation Rule

In this section, we show that the above example characterizes general features of the AMCE. To accomplish this we start by showing that the AMCE is related to the Borda rule, a voting system that assigns points to candidates according to their order of preference. Borda rule voting is implemented as follows. With  $K$  candidates, the Borda rule assigns zero points to each voter’s least preferred candidate, one point to the candidate preferred to that but no other, and so on until the most preferred candidate receives  $K - 1$  points. Thus for each voter, the Borda score contributed to a candidate corresponds to the number of other candidates to whom he or she is preferred. This in turn is equal to the number of times that candidate would be chosen if the voter was presented with every possible binary comparison. A candidate’s Borda score is the sum of the individual Borda scores assigned to that candidate by each voter, and is equal to the total number of times that candidate would be chosen if each voter was subjected to each binary comparison. This is summarized in Lemma 1:

LEMMA 1: *The Borda score of each profile is equal to the total number of times that profile is chosen in all pairwise comparisons.*

PROOF:

All proofs are in the appendix.

In the context of conjoint experiments, we further define the **Borda score of a feature** as the sum of the Borda scores of each profile that has that feature. For example, the Borda score of “female” is the sum of the Borda scores of all female candidates. This definition allows us to state our first main result that connects the AMCE to the Borda rule:

PROPOSITION 1: *The difference of the Borda scores of two features is proportional to the AMCE.*

The intuition for the proof of Proposition 1 follows from Lemma 1 and the observation that Borda

and AMCE aggregate preferences in analogous ways. They both tally the number of alternatives that are defeated by candidates with a given feature, then use that tally to compare across features. The AMCE is constructed by taking the difference of these tallies and normalizing them to be between  $-1$  and  $1$ . In the appendix we walk through the steps of how to get to AMCE from Borda scores, and produce the same expression as the AMCE in Equation 5 of Hainmueller, Hopkins and Yamamoto (2014).

This connection between the Borda rule and the AMCE is important, because it is well known in the social choice literature that the Borda rule has several undesirable properties. We have shown by way of example that the AMCE has properties similar to the Borda rule. Specifically, the Borda rule violates the independence of irrelevant alternatives (IIA) criterion, which states that the relative ranking of two candidates should not depend on the inclusion of another candidate. In the second part of our example we showed that the AMCE depends upon the particular features included in an experiment. In our example, the estimated AMCE on male versus female depends upon whether or not we include race, precisely because the inclusion of race changes the relative ranking of men and women. Of course, this is problematic since the AMCE is only internally valid with respect to the features included, or feature-combinations excluded, in the particular experimental design that produced it.<sup>4</sup>

A second social choice property of the AMCE tells us that we should be wary of standard interpretations of conjoint results in political science. Specifically, like the Borda rule, the AMCE violates the majority criterion. This states that if a majority of voters prefer one candidate, then that candidate must win. Our example shows that this feature of the Borda rule extends to attributes, where a majority of voters prefer male candidates to female candidates, but the Borda score of  $F$  is greater than that of  $M$ . Here, we establish this result more generally. Specifically, we show that when a

<sup>4</sup>This finding, furthermore, provides a theoretical foundation for the paper of de la Cuesta, Egami and Imai (2019) who demonstrate the sensitivity of the AMCE to changes in the distribution of features randomized.

majority of respondents prefer a feature, the AMCE may still indicate that feature has a negative effect on the average probability of being chosen.

This discrepancy is driven by respondents assigning different weights, or importance, to attributes. For example, if respondents who like a feature also put more weight on it than those who dislike it, the AMCE will be higher. More importantly, a small minority that cares intensely about an attribute can overtake a much larger majority that has the opposite preference but cares less intensely about it. This may result in an AMCE in favor of the feature the minority prefers, even if that feature would in fact lead to a large electoral disadvantage between otherwise similar candidates.<sup>5</sup>

Next, we leverage the relationship between the AMCE and the Borda rule to derive bounds on the fraction of the population that prefers a feature over the benchmark. That is, for a given AMCE, total number of possible candidate profiles in the experiment, and number of values the attribute of interest can take, we characterize the maximum and minimum fractions of voters who prefer that feature over the benchmark attribute. These bounds show that the potential divergence with AMCE grows in the number of unique candidate profiles,  $K$ . Our next result presents these bounds. For simplicity, for our next result and the rest of this section we assume that there are no interactions between preferences over different attributes, that is, voters have unconditional preferences over candidate features. We then relax this assumption in the next section.<sup>6</sup>

**PROPOSITION 2:** *Let  $y$  denote the fraction of voters who prefer  $t_1$  over  $t_0$ . Given an AMCE of*

<sup>5</sup>A third important feature of the Borda Rule is that it is manipulable, or not strategy-proof, when used as a choice mechanism: respondents can be better off by misrepresenting their preferences. However, because conjoint experiments aggregate preferences to describe them rather than to pick an option, we believe strategy-proofness is not a central concern here.

<sup>6</sup>Formally, voter  $i$ 's choices exhibit no interactions when for all  $t_1$  and  $t_0$ , we have

$$Y_i((t_1, T_{[-l]}), (t_0, T_{[-l]})) = Y_i((t_1, T'_{[-l]}), (t_0, T'_{[-l]}))$$

where  $T_{[-l]}$  and  $T'_{[-l]}$  denote two arbitrary vectors of other treatment components.

$\pi(t_1, t_0)$ , it must be that

$$y \in \left[ \max \left\{ \frac{\pi(t_1, t_0)\tau(K-1) + \tau}{K(\tau-1) + \tau}, 0 \right\}, \min \left\{ \frac{\pi(t_1, t_0)\tau(K-1) + K(\tau-1)}{K(\tau-1) + \tau}, 1 \right\} \right]$$

where  $\tau$  is the number of distinct values the attribute of interest can take.

To find these bounds, all we need to calculate are the range of possible Borda scores a respondent can contribute to a feature (as a function of the total number of possible profiles) and the number of distinct values the attribute of interest can take. First, we assume that the attribute of interest has the highest importance for all supporters of the feature of interest, i.e. the respondents who prefer it over the benchmark. For this group, this means that all profiles with the feature of interest are preferred to all profiles without that feature, yielding the highest possible Borda score for the feature of interest and the minimum possible Borda score for the benchmark. Thus we obtain the maximum net Borda score a supporter can contribute to a feature.

Second, we assume that the attribute of interest is least important for all opponents of that feature, i.e. the respondents who prefer the benchmark over the feature. When this is the case the feature of interest will factor into the respondent's choice only if the profiles are otherwise identical. Subject to the constraint that opponents prefer the benchmark, this results in the highest possible Borda score for the feature and the lowest for the benchmark, yielding the minimum net Borda score an opponent can subtract from a feature. Having calculated the maximum Borda score for a feature per supporter and opponent, we can invoke Proposition 1 to calculate the maximum possible AMCE for a given fraction of opponents and supporters. Inverting this function yields the lowest possible fraction of supporters for a given AMCE. The upper bound is calculated analogously. Interested readers can find the details in the proof, where we formally state and carefully trace the arguments summarized here.

In Figure 2, we apply this proposition to compute the bounds for AMCEs of 0.05, 0.10, 0.15, and 0.25 for a binary feature, plotting the upper and lower bounds of the proportion of experimental subjects who prefer a binary feature on the y-axis against the number of potential candidate profiles that respondents can choose from on the x-axis. As the figure shows, even for AMCEs of a fairly large magnitude, it takes fewer than five possible profiles for these bounds to grow to a range that is inconclusive about the preference of the majority. Of course, nearly all conjoint experiments exceed five possible candidate profiles. For instance, with six attributes taking two possible values each—still a conservative design by recent standards—there are already  $2^6 = 64$  possible profiles. Only when the AMCE is extremely large—an effect size of 0.25, which is rarely achieved by anything other than controls such as a candidate’s partisanship or experience—does the bounding exercise assure a majority preference. Even then, if the feature of interest were ternary instead of binary, this would no longer be the case even at an effect size of 0.25.

In Table 6, we conduct this exercise for every forced-choice conjoint experiment in the *APSR* and *JOP* published between 2016 and the first quarter of 2019.<sup>7</sup> We construct our bounds for the largest estimated effect presented in each of these papers. From the eight papers we analyze, only one—that of Mummolo (2016)—produces bounds that guarantee a majority preference. In this paper, the estimated effect is quite large (0.30), the attribute of interest is binary, and the number of possible candidate profiles is the smallest by far of all the included experiments.

The bounding exercise we propose contains the entire range of preferences that are consistent with a given AMCE. In other words, the upper and lower bounds reflect a worst-case scenario for researchers, which is realized when preferences over features and weights over attributes are highly correlated. Thus, Proposition 2 underscores the dangers of making statements about aggregate

<sup>7</sup>We also checked the *AJPS*, but there are no forced-choice conjoint experiments appropriate for our analysis published there in the last three years. Hemker and Rink (2017) have statistically significant findings only when they use non-binary scales as outcomes, and Huff and Kertzer (2017) have a binary outcome (labeling an attack as an act of terrorism) that is not a forced choice between two alternatives.



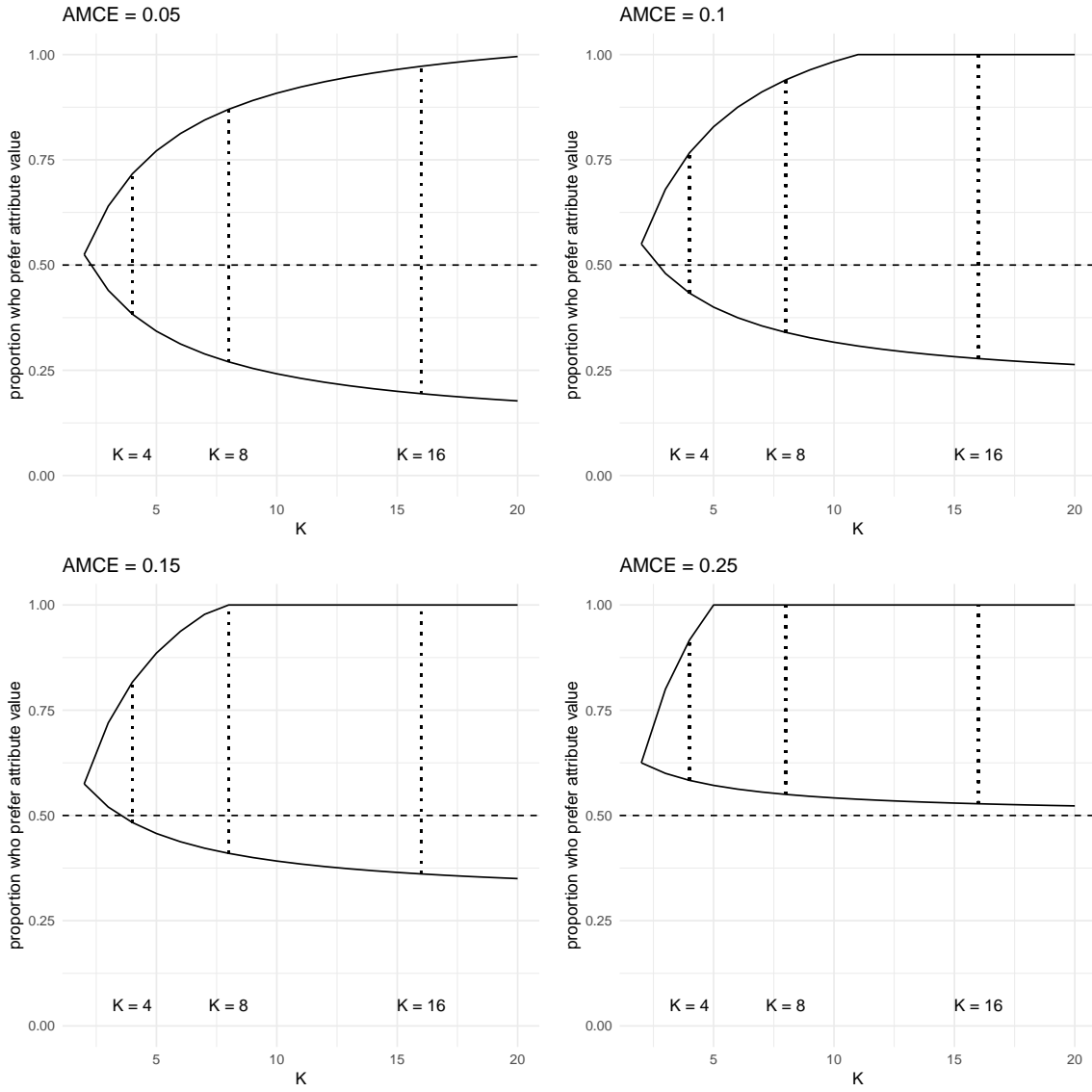


FIGURE 2. UPPER AND LOWER BOUNDS ON FRACTION OF PEOPLE WHO PREFER A BINARY FEATURE, CONSISTENT WITH AN AMCE OF .05, .10, .15, AND .25, RESPECTIVELY, AS A FUNCTION OF NUMBER OF POSSIBLE CANDIDATE PROFILES.

TABLE 6—BOUNDS ON PROPORTION OF SAMPLE HAVING PREFERENCES CONSISTENT WITH AMCE, COMPUTED FOR RECENT PAPERS IN THE TOP POLITICAL SCIENCE JOURNALS

Paper	Estimated effect	AMCE ( $\pi$ )	Number of profiles ( $K$ )	Number of relevant features ( $\tau$ )	Bounds on proportion with consistent preference	95% confidence set
<b>APSR</b>						
Ward (2019)	Proportion of group comprised of university graduates on support for immigration, 30% vs. 0%	0.22	20	4	[0.32, 1.00]	[0.29, 1.00]
Auerbach and Thachil (2018)	Broker education on support, high (BA) vs. none	0.13	1,296	3	[0.20, 1.00]	[0.14, 1.00]
Hankinson (2018)	Height of building on homeowners' support for new construction, 12 vs. 2 stories	-0.16	6,144	4	[0.00, 0.79]	[0.00, 0.81]
Teele, Kalla, and Rosenbluth (2018)	Experience on candidate support among legislators, 8 years vs. 0 years	0.18	864	4	[0.24, 1.00]	[0.21, 1.00]
Carnes and Lupu (2016)	Liberal party label on candidate support (Argentina)	-0.10	32	2	[0.00, 0.76]	[0.00, 0.84]
<b>JOP</b>						
Ballard-Rosa, Martin, and Scheve (2016)	Tax rate on those earning <10k on support for plan, 25% vs. 0%	-0.24	38,400	4	[0.00, 0.68]	[0.00, 0.73]
Mummolo and Nall (2016)	Driving time to work on Democrats' choice of community to live, 75 vs. 10 minutes	-0.23	3,456	4	[0.00, 0.69]	[0.00, 0.72]
Mummolo (2016)	Relevant information on choice to consume, vs. irrelevant (among seniors)	0.30	6	2	[0.63, 1.00]	[0.58, 1.00]

preferences with so little structure on individual choices.

Of course, we may be unlikely to achieve this worst-case scenario, and the correlation between preferences over features and weights may not be so large. Researchers may therefore want to know how the AMCE performs in the best-case scenario. We can use the logic underlying Proposition 2 to show that when voters have uncorrelated weights—when respondents who have a preference for a feature do not systematically prioritize that attribute more than those who have the opposite preference—the AMCE and the majority preference must point in the same direction. That is, if our expectation about the importance of an attribute to a respondent does not change when we learn about the direction of their preference, the sign of the AMCE must correspond to the feature preferred by the majority. Usefully for researchers, under these conditions the AMCE will be smaller in magnitude than the size of the margin, thus providing an attenuated—and therefore conservative—estimate for that quantity.

**PROPOSITION 3:** *When the direction and intensity of preferences across respondents are uncorrelated, the AMCE of a binary attribute has the same sign as the majority preference, but underestimates the size of the margin.*

Proof of Proposition 3 follows closely the logic of Proposition 2: when weights and direction are uncorrelated across supporters and opponents, on the net, each supporter contributes as much to a feature on average as an opponent contributes to the baseline. As such, the points contributed by supporters and opponents cancel out, and the remainder corresponds in sign to the margin of victory for the feature preferred by the majority.

*The Correlation Between Weights and Preferences in American Public Opinion Data*

How realistic is the assumption of no correlation between the direction and intensity of attribute preferences? To answer this we turn to survey data from the 2016 American National Election

Studies (ANES) and assess the degree to which there is a correlation in the expressed direction and intensity on a wide range of survey items. Specifically, the ANES asks about both direction and intensity of preferences for twenty-two issue areas. On seventeen of these—that is, for the vast majority of the issues in the ANES for which we have a measure of both direction and intensity of preferences—we find evidence that the supporters of a given policy or issue area have a meaningfully different assessment of its importance than its opponents.

We conduct this exercise as follows. For every question in the 2016 ANES that accommodates such an analysis, we code a direction variable that has a value of 1 if the respondent takes a clear stance in favor of a position and 0 if they are opposed.<sup>8</sup> We also code a measure of intensity that takes on evenly distributed values over the interval  $[0, 1]$  depending on how many importance categories were included in the question, where 0 is the lowest level of importance and 1 is the highest.<sup>9</sup> We then compute two summary statistics. The first, shown in the first column of Table 7, is the Pearson correlation between the direction and intensity measures, treating both as continuous variables. The second, shown in the second column, is the test statistic from a  $\chi^2$  test of independence of categorical variables. While the  $\chi^2$  test is most appropriate when treating both measures as categorical, the Pearson correlation has the advantage of being informative about the direction of the association: a positive correlation means that supporters assign more importance to the policy than opponents, while a negative correlation indicates the opposite. We report both tests and the two agree, rejecting the null hypothesis that directions and intensities are uncorrelated at  $p < .001$  for 17 out of 22 questions.

Returning to our running example of the preference for women, we see that the divergence between the preference intensities of supporters and opponents turns out to be more pronounced for espoused

<sup>8</sup>We omit respondents who say that they neither favor nor oppose the position, or that they are unsure, because there is no data on the intensity of these respondents' preferences.

<sup>9</sup>For instance, for three importance categories, we code 0 for not important at all, 0.5 for somewhat important, and 1 for very important. A detailed description of the coding of each question can be found in the supplemental appendix.

TABLE 7—CORRELATIONS BETWEEN WEIGHTS AND ATTRIBUTE PREFERENCES IN THE 2016 ANES

Question	Pearson Correlation (p-value)	$\chi^2$ statistic (p-value)	Number of intensity categories
Favor allowing use of bathrooms of identified gender	-0.258 (0.000)	309.2 (0.000)	3
Favor torture for suspected terrorists	-0.246 (0.000)	147.8 (0.000)	3
Favor allowing Syrian refugees into US	-0.246 (0.000)	203.6 (0.000)	3
Favor 2010 health care law	-0.182 (0.000)	125.4 (0.000)	3
Support preferential hiring/promotion of blacks	-0.173 (0.000)	104.9 (0.000)	2
Favor building a wall with Mexico	-0.129 (0.000)	73.6 (0.000)	3
Favor affirmative action in universities	-0.098 (0.000)	15.7 (0.000)	2
Favor sending troops to fight ISIS	-0.080 (0.000)	21.7 (0.000)	3
Think economy has gotten better since 2008	-0.065 (0.000)	13.1 (0.000)	2
Agree that children brought illegally should be sent back	-0.025 (0.103)	2.7 (0.260)	3
Think government should make it harder to own a gun	-0.024 (0.289)	7.1 (0.069)	4
Approve of House incumbent	-0.013 (0.472)	0.5 (0.497)	2
Favor ending birthright citizenship	-0.011 (0.550)	1.6 (0.459)	3
Favor requiring provision of services to same-sex couples	0.020 (0.199)	8.8 (0.012)	3
Favor laws protecting gays against job discrimination	0.110 (0.000)	49.9 (0.000)	2
Think government should take more action on climate change	0.132 (0.000)	66.2 (0.000)	3
Favor requiring employers to give paid leave to new parents	0.149 (0.000)	29.1 (0.000)	2
Favor vaccines in schools	0.174 (0.000)	97.7 (0.000)	3
Support requiring equal pay for men and women	0.201 (0.000)	145.2 (0.000)	3
Favor the death penalty	0.211 (0.000)	184.2 (0.000)	2
Believe benefits of vaccination outweigh risks	0.275 (0.000)	251.4 (0.000)	3
The term 'feminist' describes you extremely/very well	0.328 (0.000)	115.3 (0.000)	5

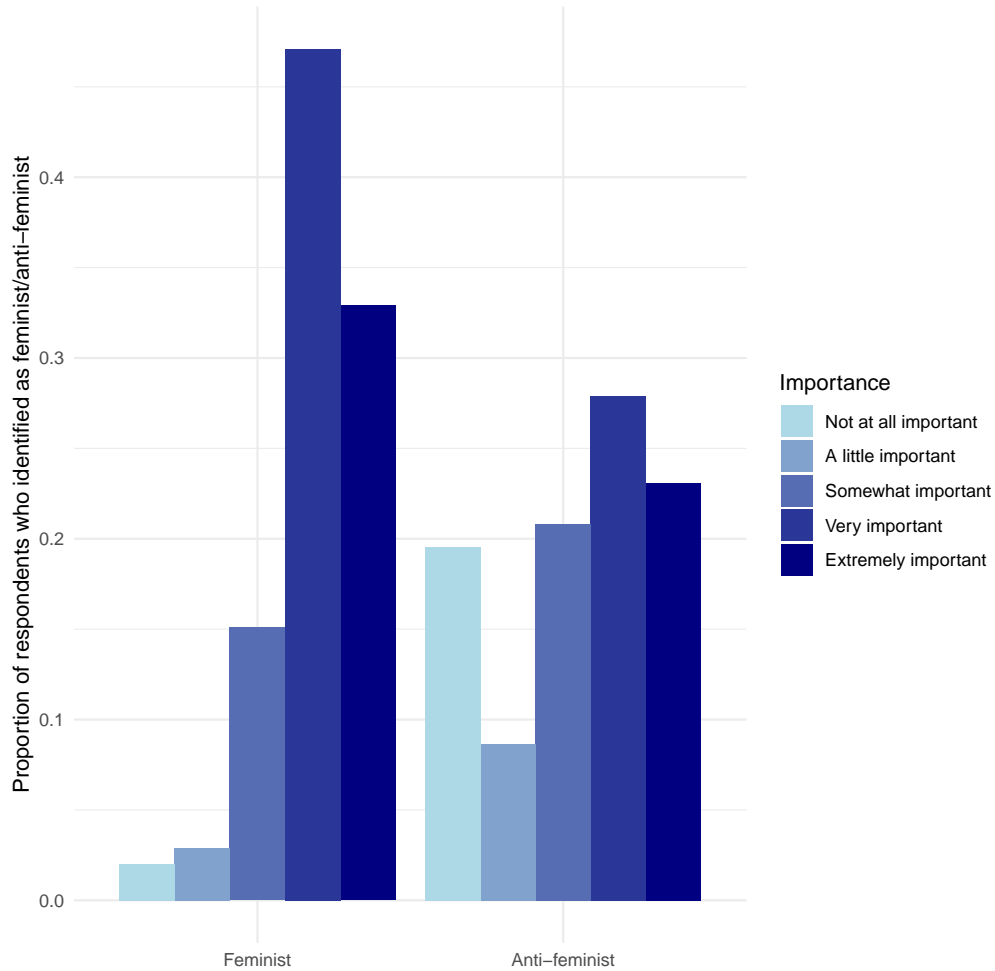
support for feminism than for any other question in the ANES. As Figure 3 shows, self-described feminists tend to attach much more importance to this identity than self-described “anti-feminists.” On the left side of Figure 3, we take the sample of ANES respondents who answered the question “How well does the term ‘feminist’ describe you?” with “Very well” or “Extremely well,”<sup>10</sup> and we plot the proportions of this sample who answered the follow-up question “How important is it to you to be a feminist?” with “Not at all important,” “A little important,” “Somewhat important,” “Very important,” and “Extremely important,” respectively. Nearly half of these feminist identifiers report that this issue is very important to them, with approximately another third calling it extremely important. By contrast, the right side of the figure shows the same distribution for the sample of respondents who answered the question “How well does the term ‘anti-feminist’ describe you?” with “Very well” or “Extremely well.” The distribution of this intensity measure for “anti-feminists” is much flatter than the one for feminists: roughly half of the sample lands between “Not at all important” and “Somewhat important,” with the other half reporting “Very important” or “Extremely important.” Crucially, the sample on the right is those who identify strongly as *anti-feminists*, not merely those who fail to identify strongly as feminists, who would naturally be expected not to care deeply about the issue. Figure 3 thus presents strong empirical evidence in favor of the very dynamic that drove our stylized running example: there are a majority of voters who prefer men but care little about the issue, with a minority that prefers women but cares a great deal.

*Using the No-Interactions Assumption to Compute Tighter Bounds*

Next, we take advantage of the structure of conjoint data to compute tighter bounds on the proportion of survey respondents who prefer a feature over the baseline than the general bounds derived above. To do so, we use the insight that when one or more attributes are held fixed at the

<sup>10</sup>The other choices were “Somewhat well,” “Not very well,” and “Not at all.”

FIGURE 3. RESPONDENTS' IDENTIFICATION WITH FEMINIST/ANTI-FEMINIST LABELS, BY ISSUE IMPORTANCE



same value in a given head-to-head comparison, the respondent makes her decision based only on the values of the remaining attributes (those that differ from one another), assuming that there are no interactions—that is, that the choice between any two features is not contingent on the value of another attribute. Under this key assumption, we can compute tighter bounds as a weighted average of our standard bounds computed within all subsets of the data, where the subsets are defined according to which attributes differ and which are the same in the randomly generated candidate pairings. We recompute  $\pi$ ,  $K$ , and  $\tau$  within each subgroup, where  $K$ —the number of possible candidate profiles—is computed ignoring the attributes that are the same; thus, it is guaranteed

to be smaller than the aggregate  $K$  when there is at least one common attribute. Formally, these tighter bounds are given by:

$$\left[ \sum_{s=1}^S \frac{n_s}{N} l_s(\pi_s, K_s, \tau), \sum_{s=1}^S \frac{n_s}{N} u_s(\pi_s, K_s, \tau) \right]$$

where  $l_s$  and  $u_s$  are the lower and upper bounds for a subset  $s$ , respectively. To illustrate how we create these subsets of the data, we walk through an example of a conjoint experiment with four attributes: gender (male, female), party (Democrat, Republican), race (white, Black, Hispanic, other), and age (young, middle, and old). Supposing we are interested in the effect of gender (female vs. male), we divide the data into groups based on the three remaining attributes: a group where the candidate pairs have different values of party, race, and age; three groups in which they have the same party, race, and age, respectively; three groups with two matched attributes and a third unmatched (party and race, party and age, and race and age); and a final group with all matched attributes. Generically, this will yield  $S = 2^{A-1}$  groups, the power set of all attributes other than the attribute of interest for the AMCE. Within each of these subsets, we compute an AMCE and a  $K$  that ignores the matched attributes: for instance, holding fixed party and race, there are six possible candidate profiles (2 values of gender  $\times$  3 values of age). Finally, we compute a weighted average of these subset-specific bounds, where the weight is determined by the number of observations in that subset.<sup>11</sup>

Table 8 reports these tighter bounds for the papers in Table 6 where the replication data allows for such an analysis.<sup>12</sup> In one case (Mummolo and Nall 2016), this approach narrow the bounds enough to conclude that a minority of respondents prefers the feature over the baseline, though these gains in precision are lost once we incorporate the uncertainty of the estimate.

<sup>11</sup>Together, the subsets form a partition of the full dataset. In some cases, a subset may be too small to compute an AMCE, but this will not affect the bounds dramatically precisely because it only has a small number of observations.

<sup>12</sup>This analysis was not possible for all of the papers in Table 6 because not all replication data contained unique respondent and question-level identifiers, which are needed to determine which features are held fixed (but not to calculate an AMCE).



TABLE 8—TIGHTER BOUNDS ON PROPORTION OF SAMPLE HAVING PREFERENCES CONSISTENT WITH AMCE

Paper	Estimated effect	AMCE ( $\pi$ )	Bounds on proportion with consistent preference	Tighter bounds on proportion with consistent preference
<b>APSR</b>				
Ward (2019)	Proportion of group comprised of university graduates on support for immigration, 30% vs. 0%	0.22	[0.32, 1.00] (0.30, 1.00)	[0.33, 1.00] (0.31, 1.00)
Auerbach and Thachil (2018)	Broker education on support, high (BA) vs. none	0.13	[0.20, 1.00] (0.15, 1.00)	[0.26, 0.94] (0.20, 0.99)
Teele, Kalla, and Rosenbluth (2018)	Experience on candidate support among legislators, 8 years vs. 0 years	0.18	[0.24, 1.00] (0.21, 1.00)	[0.25, 1.00] (0.21, 1.00)
Carnes and Lupu (2016)	Liberal party label on candidate support (Argentina)	-0.10	[0.00, 0.76] (0.00, 0.84)	[0.12, 0.63] (0.06, 0.69)
<b>JOP</b>				
Ballard-Rosa, Martin, and Scheve (2016)	Tax rate on those earning <10k on support for plan, 25% vs. 0%	-0.23	[0.00, 0.70] (0.00, 0.73)	[0.00, 0.70] (0.00, 0.73)
Mummolo and Nall (2016)	Driving time to work on Democrats' choice of community to live, 75 vs. 10 minutes	-0.23	[0.00, 0.69] (0.00, 0.71)	[0.00, 0.45] (0.00, 0.71)
Mummolo (2016)	Relevant information on choice to consume, vs. irrelevant (among seniors)	0.30	[0.63, 1.00] (0.59, 1.00)	[0.65, 0.90] (0.59, 0.90)

*Notes:* AMCE and standard bounds may differ slightly from those reported in paper (and thus Table 6 above) because we reestimate the AMCE without survey weights and only on sample having two candidate profiles per respondent (unmatched profiles appear in some replication datasets). 95% confidence sets computed using a block bootstrap are reported in parentheses below the bounds.

### III Allowing for Interactions Between Features

We have thus far focused on the scenario where voters had unconditional preferences over candidate features. In this section we explore the implications of altering this definition of preferences for features and now allow for arbitrary interactions. For instance, we allow for the possibility that men are preferred to women only when the candidate is a Republican and the reverse when the candidate is a Democrat.<sup>13</sup> We derive a summary statistic for aggregate feature preferences that captures this more complex, and potentially more realistic, preference structure and show that the bounds derived in Proposition 2 with no interactions are always smaller than the bounds we can construct when we allow for interactions across features. Thus, the AMCE is *less* informative about the fraction of voters who prefer an attribute when preferences over features can interact. Furthermore, we discuss some interpretive limitations that applied researchers face when they allow respondents to have interactive preferences over features.

To start, we define an **individual feature preference** for feature  $t_1$  over feature  $t_0$  as the proportion of the time respondent  $i$  selects a profile with feature  $t_1$  over an otherwise identical profile with feature  $t_0$ , over all all-else-equal head-to-head contests that can be constructed from all values of the other features. Formally:

$$\Pi_i(t_1, t_0) = \frac{1}{K/\tau} \sum_{j=1}^{K/\tau} Y_i(x_{j1}, x_{j0})$$

where  $K$  and  $\tau$  are defined as before, and thus  $K/\tau$  represents the number of possible all-else-equal comparisons for the feature of interest. As in our example, we denote by  $Y_i(x_{j1}, x_{j0}) = 1$  if voter  $i$  chooses profile  $x_{j1}$  with feature  $t_1$  over an otherwise identical profile  $x_{j0}$  with feature  $t_0$  in a pairwise comparison, and  $Y_i(x_{j1}, x_{j0}) = 0$  otherwise.

<sup>13</sup>That is, the *feature* they prefer is a function of the other features—not their preferred candidate profile, which is, of course, also a function of the other features in our main example.

Note that with no interactions  $\Pi_i(t_1, t_0)$  can take only two values, 0 or 1, since voters make the same choice regardless of the other candidate features. Moreover, without interactions, averaging the individual feature preference over respondents yields the proportion of individuals who prefer  $t_1$  to  $t_0$  as defined in the previous section. When we allow for interactions across features,  $\Pi_i(t_1, t_0)$  is continuous in  $[0, 1]$ . We now define a preference for  $t_1$  over  $t_0$  as having  $\Pi_i(t_1, t_0) > 1/2$  in this interactive setting, and we derive the bounds on the proportion of respondents who prefer  $t_1$  to  $t_0$  according to this definition.

PROPOSITION 4: *When there are interactive preferences, the bounds on the fraction of voters who prefer  $t_1$  over  $t_0$  are wider for any given AMCE than in the case of no interactions.*

In our supplemental appendix we show that when there are no interactions between attributes, this bounded quantity—the proportion of experimental subjects who prefer  $t_1$  to  $t_0$ —is also indicative of an electoral advantage, but when we allow for interactions this is no longer the case. In other words, we show that with interactive preferences, even with tight bounds indicating a majority of respondents having  $\Pi_i > 1/2$ , researchers should not conclude that candidates with feature  $t_1$  will beat candidates with feature  $t_0$  in most all-else-equal contests.

Furthermore, when interactive preferences are admitted, the individual feature preference is potentially undesirable because it does not satisfy transitivity. To see this, suppose there are two ternary variables of interest,  $P \in \{L, C, R\}$  and  $E \in \{H, U, G\}$ , and consider a voter whose ranking over candidate profiles is as follows:

$$RG \succ LG \succ CG \succ LU \succ CU \succ RU \succ CH \succ RH \succ LH$$

Looking at all-else-equal comparisons, this voter chooses  $R$  over  $L$ ,  $L$  over  $C$ , and  $C$  over  $R$  in two of three comparisons, or  $\Pi_i(R, L) = \Pi_i(L, C) = \Pi_i(C, R) = 2/3$ . Thus, voter  $i$  prefers  $R$  to  $L$ ,  $L$

to  $C$ , and  $C$  to  $R$ . Put simply, if we admit interactions between features, then the very notion of a preference for features becomes difficult to pin down from a theoretical perspective.

#### IV Structural Interpretation of the AMCE

Although the proposed estimator of the AMCE of Hainmueller, Hopkins and Yamamoto (2014) is “model free,” in this section we demonstrate how it relates to an underlying model of choice. Our purpose in providing this simple structural interpretation of the AMCE is to illustrate from another angle the same aggregation problem that we have already identified in the preceding sections, wherein we cannot disentangle the intensity and direction of individual preferences. To start, consider two candidates  $c \in \{1, 2\}$  running in contest  $j$  who offer platforms  $\mathbf{x}_{ijc}$  to voter  $i$ . A platform  $\mathbf{x}_{ijc}$  is a vector of policies of length  $M$  that fully characterizes a candidate in contest  $j$ . Let  $b_i$  represent an  $M$  length vector of voter  $i$ ’s preferred policy locations (e.g., their issue-specific ideal-points), and assume that voters have quadratic utility functions. Thus, voter  $i$ ’s utility is maximized when candidate  $c$  offers a platform that exactly matches her preferred policy positions, and the loss she obtains is a function of the distance between the candidate’s policies and her ideal platform. Her utilities from Candidate 1 and 2’s respective platforms is given by:

$$(1) \quad \begin{aligned} U_i(\mathbf{x}_{ij1}) &= -(b_i - \mathbf{x}_{ij1})^2 + \eta_{ij1} \\ U_i(\mathbf{x}_{ij2}) &= -(b_i - \mathbf{x}_{ij2})^2 + \eta_{ij2} \end{aligned}$$

While the imposition of quadratic loss may seem restrictive, in the appendix we show that our results are numerically identical if we assume an absolute linear loss utility function. Regardless, it

follows that:

$$\begin{aligned}
\Pr(y_{ij1} = 1) &= \Pr(U_i(\mathbf{x}_{ij1}) > U_i(\mathbf{x}_{ij2})) \\
(2) \qquad &= \Pr(-(b_i - \mathbf{x}_{ij1})^2 + \eta_{ij1} > -(b_i - \mathbf{x}_{ij2})^2 + \eta_{ij2}) \\
&= \Pr(\eta_{ij2} - \eta_{ij1} < 2(b'_i(\mathbf{x}_{ij1} - \mathbf{x}_{ij2}) + \mathbf{x}'_{ij2}\mathbf{x}_{ij2} - \mathbf{x}'_{ij1}\mathbf{x}_{ij1}))
\end{aligned}$$

where  $y_{ij1}$  is a binary indicator that equals 1 when respondent  $i$  chooses Candidate 1 in contest  $j$  and 0 otherwise. Now consider data generated from a conjoint experiment, where  $\mathbf{x}_{ij1}$  and  $\mathbf{x}_{ij2}$  are vectors of randomized candidate attributes that have been discretized into binary indicators with an omitted category.

Typically, we would estimate Equation 2 with a probit or logit-like regression. Instead consider a linear model of the form:

$$\begin{aligned}
y_{ij1} &= 2(b'_i(\mathbf{x}_{ij1} - \mathbf{x}_{ij2}) + \mathbf{x}'_{ij2}\mathbf{x}_{ij2} - \mathbf{x}'_{ij1}\mathbf{x}_{ij1}) + \eta_{ij1} - \eta_{ij2} \\
&= \sum_k (2b_{im}(x_{ijm1} - x_{ijm2}) + x_{ijm2}^2 - x_{ijm1}^2) + \eta_{ij1} - \eta_{ij2} \\
(3) \qquad &= \sum_k (2b_{im} - 1)(x_{ijm1} - x_{ijm2}) + \eta_{ij1} - \eta_{ij2} \\
&= \sum_k \beta_{im} \Delta x_{ijm} + \epsilon_{ij}
\end{aligned}$$

where  $\mathbb{E}(\epsilon_{ij}) = \mathbb{E}(\eta_{ij1} - \eta_{ij2}) = 0$  follows from the randomization of  $\mathbf{x}_{ij1}$  and  $\mathbf{x}_{ij2}$ , and the third line follows from the fact that  $x_{ijmc}^2 = x_{ijmc}$ , as this is a dummy. The slope,  $\beta_{im} = 2b_{im} - 1$ , gives the change in probability for individual  $i$  of choosing Candidate 1 when Candidate 1 has feature  $m$  and Candidate 2 does not, holding all their other features constant. Implicitly, it also constrains each element of  $b_i$  to the  $[0, 1]$  line. When  $b_{im} = 0$  (and  $\beta_{im} = -1$ ) the manipulation  $\Delta x_{ijm} = 1$  holding all other features constant gives a predicted reduction in the probability of choosing Candidate 1

of one-hundred percent. When  $b_{im} = 1$  (and  $\beta_{im} = 1$ ), the same manipulation gives a predicted increase in the probability of choosing Candidate 1 of one-hundred percent. When  $b_{im} = \frac{1}{2}$  (and  $\beta_{im} = 0$ ), this indicates that voter  $i$  is perfectly indifferent.

Finally, averaging over all individuals, we obtain  $\mathbb{E}(\beta_{im})$  as the coefficient from the regression:

$$(4) \quad y_{ij1} = \sum_m \Delta x_{ijm} \beta_m + \epsilon_{ij}$$

where the estimated coefficient  $\hat{\beta}_m$  recovers the AMCE for feature  $m$ .<sup>14</sup> Thus we see that, under this simple model of choice, the AMCE can be interpreted as an average of respondents' ideal points. This insight illuminates why the AMCE is such an inappropriate summary statistic for making claims about winners of elections or representative voters' preferences. Under majority rule, elections are won by the median voter, and the magnitudes of the ideal points of the most extreme voters should do nothing to change the probability of a given candidate winning the election. Measuring the probability of winning as a function of preference intensity essentially gives citizens voting power commensurate with the strength of their opinions—a feature almost never observed in real-life institutional designs.<sup>15</sup>

## V Conclusion

We have shown that the AMCE, the target estimand of most conjoint experiments, does not support most interpretations ascribed to it by political scientists. A positive AMCE for a particular candidate-feature does not imply that the majority of respondents prefer that feature over the benchmark. It does not indicate that they prefer a candidate with that feature to a candidate

<sup>14</sup>For a simple proof, see the supplemental appendix.

<sup>15</sup>It is true that voters with intense preferences may be more likely to turn out to vote or to be politically active, and may therefore exercise outsized influence on electoral outcomes. However, these are not the mechanisms that researchers are currently thinking about when interpreting the results of conjoint experiments, and if we think they are at work, we should study and model them more explicitly.

without it, all else equal. It does not mean that voters are more likely to elect a candidate with that feature than candidates without it. None of this is the consequence of uncertainty introduced by sampling or measurement; rather, it is inherent to the AMCE's properties as an aggregation mechanism. Even when the universe of respondents is fully observed and every conceivable contest between candidates is assessed carefully and honestly, claims about voter preferences and electoral outcomes are not generally supported by the results from conjoint experiments.

Instead, what we have demonstrated is that the AMCE can be thought of as an average of the direction and intensity of voters' preferences, or an average of ideal points. As a consequence, it can point in the opposite direction as the majority preference when there is a minority that intensely prefers a feature and a majority that feels the opposite, but less strongly. The larger the correlation between direction and intensity, the more misleading the AMCE with respect to quantities of interest in a one-person, one-vote setting. Far from a statistical accident, this preference structure pervades the sorts of issues that interest political scientists (and for which conjoints are often deployed), such as gender parity in elected office (Teele, Kalla and Rosenbluth, 2018) or who should be favored by the nation's immigration policy (Hainmueller and Hopkins, 2015).

Building on well-known results from the literature on social choice, we have characterized the AMCE as a preference aggregation mechanism and shown its relationship to the Borda rule. Using this correspondence, we then derived bounds on the proportion of a sample that prefers a feature based on a given AMCE. Unfortunately, the vast majority of findings published in the top political science journals in the past few years fail to support claims about majority preferences. That said, we have also shown that if one is willing to assume no correlation between preference direction and intensity, then at the very least the sign of the AMCE from these experiments corresponds to the will of the majority. Since this assumption is unlikely to hold, the discipline must reevaluate what we have learned from conjoint experiments with this clearer understanding of the AMCE in mind.

A simple corrective, which we strongly encourage, is for applied researchers to use precise language when interpreting the results of conjoint experiments, placing the “representative voter” implied by the AMCE in the appropriate context. While common interpretations such as “voters prefer  $A$  to  $A'$ ” are not well-defined,<sup>16</sup> typically researchers interpret conjoint results to evoke some notion of a majority—one that is not supported by the target estimand of these experiments. By the same token, political scientists should, on the whole, stop making inferences about electoral contests from the AMCE unless these claims are supported by further evidence about there being no correlation between voters’ priorities and preferences.

Nevertheless, if researchers must rely upon forced-choice conjoint experiments, our paper suggests they may find themselves in a bind. On the one hand, our results indicate that if they want to interpret their findings with respect to a majority preference, then they should restrict themselves to conservative randomization schemes that limit the number of attributes and potential candidate-profiles. Only with a conservative design and a small number of binary attributes is there hope of producing sufficiently small bounds on an estimated AMCE to conclusively reflect a majority preference. On the other hand, because the AMCE is dependent upon the particular features included in an experiment, for a result to be externally valid researchers must include the full set of theoretically relevant attributes in their randomization scheme. That is, for a conjoint experiment to provide substantively relevant results and, moreover, for it to recover point estimates that are stable with respect to the inclusion of additional features, researchers must get the distribution of randomized attributes exactly right. Unfortunately, it may prove difficult to construct a “Goldilocks” experimental design that both randomizes a conservative number of features—to enable researchers’ claims about a majority preference—and includes a sufficiently large number of attributes—to be assured that results are insensitive to the randomization distribution.

<sup>16</sup>Does “voters prefer” mean all voters? A subset of voters? If so, what subset?



If researchers want to make claims about majority preferences from conjoint experiments, one potential way forward may be to combine them with experiments designed to recover voters' priorities. As we have shown in Proposition 3, if the direction of respondents' preferences is uncorrelated with the weights they assign to the dimensions of choice, claims about a majority preference can be sustained with existing research designs. However, this avenue is only fruitful to the extent that this uncorrelated preference structure is present in realistic political contexts. Instead, we suggest that researchers be willing to make stronger assumptions to gain the ability to make claims about electoral outcomes. A fully structural approach to conjoint analysis may prove most versatile in combining the realistic approximations of candidates that randomizing a large number of candidate-features provides with an ability to make well-supported claims about electoral contests. By imposing and estimating a model of choice, researchers may be able to have their cake and eat it too.

## REFERENCES

- Arrow, Kenneth J.** 1950. "A difficulty in the concept of social welfare." *Journal of political economy*, 58(4): 328–346.
- Austen-Smith, David, and Jeffrey S Banks.** 2000. *Positive Political Theory I: Collective Preference*. Vol. I, University of Michigan Press.
- Ballard-Rosa, Cameron, Lucy Martin, and Kenneth Scheve.** 2017. "The structure of American income tax policy preferences." *The Journal of Politics*, 79(1): 1–16.
- 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.
- de la Cuesta, Brandon, Naoki Egami, and Kosuke Imai.** 2019. "Experimental Design and Statistical Inference for Conjoint Analysis: The Essential Role of Population Distribution." Working Paper.
- Gibbard, Allan.** 1973. "Manipulation of voting schemes: a general result." *Econometrica*, 41(4): 587–601.
- Hainmueller, Jens, and Daniel J Hopkins.** 2015. "The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants." *American Journal of Political Science*, 59(3): 529–548.
- 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.
- Horiuchi, Yusaku, Daniel M Smith, and Teppei Yamamoto.** 2018. "Identifying voter preferences for politicians' personal attributes: A conjoint experiment in Japan." *Political Science Research and Methods*, 1–17.
- Khanna, Kabir.** 2019. "What traits are Democrats prioritizing in 2020 candidates?" *CBS News*.  
<https://www.cbsnews.com/news/democratic-voters-hungry-for-women-and-people-of-color-in-2020-nomination/>.
- Lalley, Steven P, and E Glen Weyl.** 2018. "Quadratic voting: How mechanism design can radicalize democracy." *AEA Papers and Proceedings*, 108: 33–37.
- Magni, Gabriele, and Andrew Reynolds.** 2019. "Democrats Don't Want to Nominate a Candidate Who Looks Like Bernie or Joe." *POLITICO*.  
<https://www.politico.com/magazine/story/2019/05/24/democrats-dont-want-to-nominate-another-white-man-for-president-226977>.
- Mummolo, Jonathan.** 2016. "News from the other side: How topic relevance limits the prevalence of partisan selective exposure." *The Journal of Politics*, 78(3): 763–773.
- Teale, 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.

## PROOFS

## PROOF OF LEMMA 1:

Suppose there are  $N$  voters and  $K$  profiles. Consider voter  $i$ 's preference ranking over profiles. For any pair of profiles  $x_j, x_k$ , denote by  $Y_i(x_j, x_k) = 1$  if  $i$  chooses profile  $x_j$  over  $x_k$  in a pairwise comparison, and  $Y_i(x_j, x_k) = 0$  otherwise. Without loss of generality, reorder the profiles such that the profile most preferred by  $i$  is  $x_1$ , the second most preferred is  $x_2$ , and so on such that the least preferred is  $x_K$ . Assign  $i$ 's most preferred profile a Borda score of  $b_i(x_1) = K - 1$ , their second most preferred profile a score of  $b_i(x_2) = K - 2$ , and so on such that their least preferred profile has a score of zero. Suppose now  $i$  is presented with each pairwise comparison. Then,  $i$  chooses their most preferred profile  $x_1$  every time it is on the ballot, against every other profile, so

$$\sum_{j \neq 1} Y_i(x_1, x_j) = \underbrace{1 + 1 + 1 + \dots + 1}_{K-1 \text{ times}} = K - 1$$

times. The second most preferred will be chosen every time except when compared with the most preferred profile, so

$$\sum_{j \neq 2} Y_i(x_2, x_j) = 0 + \underbrace{1 + 1 + 1 + \dots + 1}_{K-2 \text{ times}} = K - 2$$

times. Going this way, we see that individual Borda scores over profiles match exactly with the number of times each profile is chosen when every pairwise comparison is made. Finally, the least preferred profile will never be chosen in a pairwise comparison made by voter  $i$ ,  $\sum_{j \neq K} Y_i(x_K, x_j) = 0 + 0 + 0 + \dots + 0 = 0$ . Thus, for each individual voter, the Borda score of a profile is equal to the number of times it is chosen when that voter makes all pairwise comparisons,  $b_i(x_m) = \sum_{j \neq m} Y_i(x_m, x_j)$ .

The aggregate Borda score of a profile is the sum of individual voters' Borda scores of that profile. When we sum across voters the times each profile  $x_m$  is chosen in all pairwise comparisons, their

sums must be equal to the sum of individual Borda scores. Formally,

$$b(x_m) \equiv \sum_{i=1}^N b_i(x_m) = \sum_{i=1}^N \sum_{j \neq m} Y_i(x_m, x_j).$$

LEMMA 2: *When there are no interactions and only binary attributes, a profile has the highest Borda score if and only if all its features have the highest Borda scores for their respective attributes.*

PROOF OF LEMMA 2:

Let us first restate the formal definition of no interactions. Voter  $i$ 's choices exhibit no interactions when for all  $t_1$  and  $t_0$ , we have

$$Y_i((t_1, T_{[-l]}), (t_0, T_{[-l]})) = Y_i((t_1, T'_{[-l]}), (t_0, T'_{[-l]}))$$

where  $T_{[-l]}$  and  $T'_{[-l]}$  denote two arbitrary vectors of other treatment components.

Formally, Borda score of a feature  $t_1$ ,  $B(t_1)$  is

$$B(t_1) \equiv \sum_{i=1}^N \sum_{x_1 \in \kappa(t_1)} \sum_{x_j \neq x_1} Y_i(x_1, x_j)$$

where  $\kappa(t_1)$  denotes the set of all profiles that have the feature  $t_1$ .

No interactions implies

$$b_i(t_1, T_{[-l]}) - b_i(t_1, T'_{[-l]}) = b_i(t_0, T_{[-l]}) - b_i(t_0, T'_{[-l]})$$

for all  $t_1, t_0, T_{[-l]}$ , and  $T'_{[-l]}$  by a straightforward application of Lemma 1. Then, summing these up,

we observe

$$\sum_{i=1}^N b_i(t_1, T_{[-l]}) - \sum_{i=1}^N b_i(t_0, T_{[-l]}) = \sum_{i=1}^N b_i(t_1, T'_{[-l]}) - \sum_{i=1}^N b_i(t_0, T'_{[-l]}).$$

Suppose now  $(t_1, T_{[-l]}^*)$  is the profile with the highest Borda score. This means:

$$\sum_{i=1}^N b_i(t_1, T_{[-l]}^*) - \sum_{i=1}^N b_i(t_0, T_{[-l]}^*) \geq 0.$$

By the no interactions assumption, it follows that for any arbitrary vector of treatments  $T_{[-l]}$ :

$$\sum_{i=1}^N b_i(t_1, T_{[-l]}) - \sum_{i=1}^N b_i(t_0, T_{[-l]}) \geq 0$$

Because this is true for each vector of treatments  $T_{[-l]}$ , it is also true when we sum over them and get the Borda score of  $t_1$ . Therefore, the Borda score of  $t_1$  must be greater than that of  $t_0$  because

$$B(t_1) = \sum_{T_{[-l]}} \sum_{i=1}^N b_i(t_1, T_{[-l]}) \geq \sum_{T_{[-l]}} \sum_{i=1}^N b_i(t_0, T_{[-l]}) = B(t_0).$$

#### PROOF OF PROPOSITION 1:

The number of profiles that have  $t_1$  is equal to the number of profiles that have  $t_0$ , which is in turn equal to the total number of profiles divided by the number of unique values the attribute of interest can take:  $|\kappa(t_1)| = |\kappa(t_0)| = \frac{K}{\tau}$ . Then, by dividing the Borda score of a feature,  $B(t_1)$  by the total number of pairwise comparisons  $t_1$  appears in,  $\frac{K}{\tau}N(K-1)$ , and taking the difference with the Borda score  $B(t_0)$  of the benchmark attribute  $t_0$ , divided by  $\frac{K}{\tau}N(K-1)$  yields exactly the AMCE of  $t_1$  as defined in Hainmueller, Hopkins and Yamamoto (2014):

$$\pi(t_1, t_0) = \frac{\sum_{i=1}^N \sum_{x \in \kappa(t_1)} \sum_{x_j \neq x} Y_i(x, x_j)}{|\kappa(t_1)|N(K-1)} - \frac{\sum_{i=1}^N \sum_{x \in \kappa(t_0)} \sum_{x_j \neq x} Y_i(x, x_j)}{|\kappa(t_0)|N(K-1)} = \frac{\tau}{NK(K-1)} (B(t_1) - B(t_0)).$$

#### PROOF OF PROPOSITION 2:

We prove this proposition by finding the range of Borda scores of  $t_1$  and  $t_0$  that can be rationalized

for a given proportion of respondents who prefer  $t_1$  over  $t_0$ ; and then inverting this range to find the minimum and maximum proportions of respondents who prefer  $t_1$  over  $t_0$  for a given AMCE.

Let us find the minimum fraction of respondents who prefer  $t_1$  over  $t_0$  that is consistent with an AMCE. Notice that for a fixed fraction of respondents, the AMCE is maximized when respondents in favor of  $t_1$  assign the highest priority to the attribute, they rank  $t_1$  the best, and  $t_0$  the worst; whereas those who prefer  $t_0$  like  $t_1$  next, and assign the lowest priority to it. In other words, when those who prefer  $t_1$  rank all profiles with  $t_1$  at the top, and all profiles with  $t_0$  at the bottom, this drives the AMCE up. To help with the intuition, the preferences of such a voter might look like:

$$\underbrace{t_1\alpha\beta\gamma}_{K-1} \succ \underbrace{t_1\alpha'\beta\gamma}_{K-2} \succ \dots \succ \underbrace{t_1\alpha'\beta'\gamma'}_{K-\frac{K}{\tau}} \succ t_2\alpha\beta\gamma \succ \dots \succ t_2\alpha'\beta'\gamma' \succ \dots \succ \underbrace{t_0\alpha\beta\gamma}_{\frac{K}{\tau}-1} \succ \underbrace{t_0\alpha'\beta\gamma}_{\frac{K}{\tau}-2} \succ \dots \succ \underbrace{t_0\alpha'\beta'\gamma'}_0$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  represent a collection of other features of candidates included in the experiment. Holding constant the other features, the difference in Borda scores of a profile with  $t_1$  and with  $t_0$  is thus  $K - \frac{K}{\tau}$ . Formally, for any vector of other attributes  $T_{[-l]}$ , the profile  $(t_1, T_{[-l]})$  is maximally chosen  $K - \frac{K}{\tau}$  more times than  $(t_0, T_{[-l]})$  when every pairwise comparison is made. From Proposition 1 we know that this implies the maximum difference in Borda scores,  $b_i(t_1, T_{[-l]}) - b_i(t_0, T_{[-l]}) = K - \frac{K}{\tau}$ , for any arbitrary combination of other attributes,  $T_{[-l]}$ . Because there are  $\frac{K}{\tau}$  possible unique combinations of other attributes, each respondent makes  $\frac{K}{\tau}$  such comparisons between  $t_1$  and  $t_0$ . Thus, each respondent who prefers  $t_1$  maximally generates a  $\frac{K^2(\tau-1)}{\tau^2}$  higher Borda score for  $t_1$  than  $t_0$ .

Similarly, the maximum AMCE is only obtained when those who prefer  $t_0$  assign the lowest priority to this attribute, and rank profiles with  $t_1$  just below otherwise identical profiles with  $t_0$ .

Such preferences might look like:

$$\underbrace{t_0\alpha\beta\gamma}_{K-1} \succ \underbrace{t_1\alpha\beta\gamma}_{K-2} \succ t_2\alpha\beta\gamma \succ \dots \succ \underbrace{t_0\alpha'\beta\gamma}_{K-\tau-1} \succ \underbrace{t_1\alpha'\beta\gamma}_{K-\tau-2} \succ \dots \succ \underbrace{t_0\alpha'\beta'\gamma'}_{\tau-1} \succ \underbrace{t_1\alpha'\beta'\gamma'}_{\tau-2} \succ t_2\alpha'\beta'\gamma' \succ \dots$$

When other features are held constant, the difference in Borda scores of a profile with  $t_1$  and  $t_0$  is  $-1$ . In other words, for respondents who prefer  $t_0$  to  $t_1$ , the maximum difference is  $b_j(t_1, T_{[-l]}) - b_j(t_0, T_{[-l]}) = -1$ , for any arbitrary combination of other attributes,  $T_{[-l]}$ . Again, because there are  $\frac{K}{\tau}$  possible combinations of other features and thus as many comparisons between profiles with  $t_1$  and  $t_0$ , each respondent who prefers  $t_0$  minimally generates  $\frac{K}{\tau}$  more points for  $t_0$  than  $t_1$ .

Thus, for a given AMCE  $\pi(t_1, t_0)$ , we can derive the minimum fraction  $y$  of voters who prefer  $t_1$ ,  $y^{\min}$ , by summing these scores and normalizing:

$$\pi(t_1, t_0) = \frac{(y^{\min}) \frac{K^2(\tau-1)}{\tau^2} - (1 - y^{\min}) \frac{K}{\tau}}{\binom{K}{2} \frac{2}{\tau}}.$$

Simple algebra reveals

$$y^{\min} = \max \left\{ \frac{\pi(t_1, t_0)\tau(K-1) + \tau}{K(\tau-1) + \tau}, 0 \right\}.$$

A very similar argument establishes the upper bound of  $y$ .

### PROOF OF PROPOSITION 3:

Denote by  $n_1$  the number of respondents who prefer  $t_1$  to  $t_0$ . Similarly, let  $n_0 = N - n_1$  refer to the number of respondents who prefer  $t_0$  to  $t_1$ . Without loss of generality, reorder respondents so those who prefer  $t_1$  to  $t_0$  have the lowest rank, that is  $i \in \{1, \dots, n_1\}$ . Suppose direction and intensity of preferences are uncorrelated across respondents. Then, the average net contribution to  $t_1$  from a supporter of  $t_1$  is the same as the average net contribution to  $t_0$  from an opponent of  $t_1$ .

Formally, we can write this as

$$(A.11) \quad \frac{1}{n_1} \sum_{i=1}^{n_1} B_i(t_1) - B_i(t_0) = \frac{1}{n_0} \sum_{i=n_1+1}^N B_i(t_0) - B_i(t_1).$$

for any  $t_1, t_0$ , and  $i$ .

We know from the proof of Proposition 1 that we can write the AMCE as:

$$(A.12) \quad \pi(t_1, t_0) = \frac{\tau}{NK(K-1)} \sum_{i=1}^N B_i(t_1) - B_i(t_0).$$

Then, we can rewrite expression A.12 as

$$\pi(t_1, t_0) = \frac{\tau}{NK(K-1)} \left( \sum_{i=1}^{n_1} B_i(t_1) - B_i(t_0) - \sum_{i=n_1+1}^N B_i(t_0) - B_i(t_1) \right)$$

From Equation A.11, when preference direction and intensity are uncorrelated:

$$\pi(t_1, t_0) = \frac{\tau \mathbb{E}_{i \leq n_1} [B(t_1) - B(t_0)]}{NK(K-1)} (n_1 - n_0).$$

Thus,  $\pi(t_1, t_0)$  is positive if and only if a majority of respondents prefer  $t_1$  to  $t_0$ , or  $n_1 > 1/2$ .

#### PROOF OF PROPOSITION 4:

When preferences over features interact with one another, the bounds on the fraction of voters who prefer  $t_1$  to  $t_0$  for an AMCE of  $\pi(t_1, t_0)$ , in an experiment with  $K$  possible profiles, and when



the attribute of interest can take  $\tau$  distinct values, are given by

$$y \in \left[ \max \left\{ \frac{K(\pi(t_1, t_0)\tau(K-1) - K(\tau-1))}{\tau^2 \left( \lceil \frac{K+\tau}{2\tau} \rceil \left( 1 - 2 \lfloor \frac{1}{2} \left( \frac{K}{\tau} - 1 \right) \rfloor \right) - \lfloor \frac{1}{2} \left( \frac{K}{t} - 1 \right) \rfloor^2 \right) + K^2(t-1)} + 1, 0 \right\}, \right. \\ \left. \min \left\{ \frac{K(\pi(t_1, t_0)\tau(K-1) + K(\tau-1))}{\tau^2 \left( \lceil \frac{K+\tau}{2\tau} \rceil \left( 1 - 2 \lfloor \frac{1}{2} \left( \frac{K}{\tau} - 1 \right) \rfloor \right) - \lfloor \frac{1}{2} \left( \frac{K}{t} - 1 \right) \rfloor^2 \right) + K^2(t-1)}, 1 \right\} \right],$$

where  $\lfloor \cdot \rfloor$  and  $\lceil \cdot \rceil$  are the floor and ceiling functions respectively.<sup>17</sup>

Similarly to the proof of Proposition 2, these bounds obtain when both the voters who prefer  $t_1$  and those who prefer  $t_0$  give the maximum and minimum net Borda scores to  $t_1$  versus  $t_0$ . The bounds in this case are wider because interactions allow for more leeway when constructing preferences. Below we lay out the arguments for the lower bound. The upper bound is constructed analogously.

For respondents who prefer  $t_1$ , the maximum possible net Borda score given to  $t_1$  versus  $t_0$  with interactions is same as the case without:  $\frac{K^2(\tau-1)}{\tau^2}$ . Now consider a respondent who prefers  $t_0$ . When we allow for interactions, such a respondent prefers profiles with  $t_0$  to otherwise identical profiles with  $t_1$  in majority of the cases, but in others they may have a preference for profiles with  $t_1$ . Specifically, a respondent who prefers  $t_0$  gives the maximum possible net Borda score to  $t_1$  versus  $t_0$  when her preferences look like the following:

$$\underbrace{t_1\alpha\beta\gamma \succ t_1\alpha'\beta\gamma \succ \dots}_{\lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor \text{ profiles}} \succ \underbrace{t_0\alpha'\beta'\gamma \succ t_1\alpha'\beta'\gamma \succ \dots \succ t_0\alpha'\beta'\gamma' \succ t_1\alpha'\beta'\gamma'}_{2\lceil \frac{K}{2\tau} + \frac{1}{2} \rceil \text{ profiles}} \succ \underbrace{t_0\alpha\beta\gamma \succ t_0\alpha'\beta\gamma \succ \dots}_{\lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor \text{ profiles}}$$

where again  $\alpha$ ,  $\beta$ , and  $\gamma$  represent a collection of other features of candidates included in the experiment. In words, this respondent has the minimal distance of one between the profiles with  $t_0$  she prefers to otherwise identical profiles with  $t_1$ , and the maximal distance of  $K - \lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor$  between

<sup>17</sup>The floor and ceiling functions are necessary because of how we define a preference; strictly more than half of all all-else-equal comparisons. If there is an odd (even) number of all-else-equal comparisons, then minimally the profiles with the preferred feature are chosen once (twice) more than those without. The floor and ceiling functions account for this difference.

the profiles with  $t_1$  she prefers to otherwise identical profiles with  $t_0$ . To check that for this respondent we have  $\Pi_i(t_1, t_0) < \frac{1}{2}$ , notice there are  $\lceil \frac{K}{2\tau} + \frac{1}{2} \rceil$  comparisons where she prefers  $t_0$  over  $t_1$  and  $\lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor$  comparisons where  $t_1$  is preferred to  $t_0$ . Thus, the maximum net contribution to  $t_1$  of a respondent who prefers  $t_0$  to  $t_1$  is  $(\lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor) (K - \lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor) - \lceil \frac{K}{2\tau} + \frac{1}{2} \rceil$ . Notice that for  $\frac{K}{\tau} > 2$ , we have  $(\lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor) (K - \lfloor \frac{K}{2\tau} - \frac{1}{2} \rfloor) > \lceil \frac{K}{2\tau} + \frac{1}{2} \rceil$ . This means that when we allow for interactions, a respondent who prefers  $t_0$  to  $t_1$  may still contribute more Borda points to  $t_1$  than  $t_0$ .

When we calculate the bounds as in the proof of Proposition 2, we find that

$$\pi(t_1, t_0) = \frac{(y^{\min})^{\frac{K^2(\tau-1)}{\tau^2}} + (1 - y^{\min}) \left( \lceil \frac{K+t}{2t} \rceil (2 \lfloor \frac{1}{2} (\frac{K}{t} - 1) \rfloor - 1) + \lfloor \frac{1}{2} (\frac{K}{t} - 1) \rfloor^2 \right)}{\binom{K}{2} \frac{2}{\tau}}.$$

Algebra reveals

$$y^{\min} = \max \left\{ 1 - \frac{(K-1)K(K - \pi(t_1, t_0)\tau)}{\tau^2 \left( \lceil \frac{K}{2\tau} \rceil^2 - (K+2) \lceil \frac{K}{2\tau} \rceil + \lfloor \frac{K}{2\tau} \rfloor \right) + (K-1)K^2 + (K+2)\tau^2}, 0 \right\}$$

It can be confirmed that this is equal to the lower bound in Proposition 2 when  $\frac{K}{\tau} = 2$ , and strictly lower when  $\frac{K}{\tau} > 2$ .

#### ELECTORAL ADVANTAGE IN INTERACTIVE AND NON-INTERACTIVE SETTINGS

In this section we demonstrate that the proportion of respondents who prefer  $t_1$  to  $t_0$ , or have an individual feature preference  $\Pi_i > 1/2$ , is indicative of electoral advantage only in noninteractive settings. That is, when interactions between attributes are admitted, even tight bounds indicating a majority of respondents having  $\Pi_i > 1/2$  are not sufficient evidence to conclude that candidates with  $t_1$  will beat candidates with  $t_0$  in most all-else-equal contests.

We define **electoral advantage** of  $t_1$  over  $t_0$  as the difference between the proportion of the time  $t_1$  beats  $t_0$  in an all-else-equal contest, out of all possible all-else-equal contests, and one-half:

$$(A.21) \quad A(t_1, t_0) = \frac{1}{K/\tau} \sum_{j=1}^{K/\tau} \mathbb{1} \left\{ \left( \frac{1}{N} \sum_{i=1}^N Y_i(x_{j1}, x_{j0}) \right) > \frac{1}{2} \right\} - \frac{1}{2}$$

In other words,  $A(t_1, t_0)$  is the difference between the electorate-level analogue of  $\Pi_i(t_1, t_0)$ —the proportion of the time an electorate selects  $t_1$  over  $t_0$  in a simple-majority vote between all-else-equal alternatives, out of all possible all-else-equal contests—and one-half, and thus it captures the electoral (dis)advantage enjoyed by a candidate with feature  $t_1$  compared to  $t_0$ .

First, consider the baseline case with no interactions. Here, note that whenever a majority of voters prefers  $t_1$  to  $t_0$ ,  $x_{j1}$  will beat  $x_{j0}$  in every all-else-equal contest  $j$ , and  $A(t_1, t_0)$  will achieve its maximum value of  $\frac{1}{2}$ , so we can be confident that  $t_1$  carries an electoral advantage over  $t_0$ . But this is no longer the case when we allow for interactions. We can illustrate this by way of a simple example. Consider a population of three voters with the following interactive preferences over gender  $\in \{M, F\}$  and party  $\in \{D, R, I\}$ :

Rank	V1	V2	V3
1.	MD	MR	MI
2.	FR	FI	MD
3.	MR	MI	MR
4.	FI	FD	FI
5.	MI	MD	FD
6.	FD	FR	FR
$\Pi_i(F, M)$	2/3	2/3	0

TABLE B1—PREFERENCES OVER CANDIDATE PROFILES - EXAMPLE WHERE BOUNDS DO NOT INDICATE ELECTORAL ADVANTAGE IN INTERACTIVE SETTINGS

Here,  $\Pi_i(F, M) > 1/2$  for two out of three respondents, but  $A(F, M) = -1/6$ , indicating an electoral *disadvantage* for females despite the fact that the majority prefers this feature.

PROOF THAT EQUATION 4 IS EQUIVALENT TO THE AMCE:

To show that the estimation of Equation 4 would yield the AMCE note first that Hainmueller,

Hopkins and Yamamoto (2014) show that the following regression recovers an unbiased estimate of the AMCE:

$$y_{ijc} = \delta + x_{jmc}\rho_k + v_{ijmc}$$

where  $\hat{\rho}_m$  gives the AMCE for feature  $m$ . From the randomization of  $x$ , it follows from standard results that the vector of coefficients  $\beta$  from Equation 4 can be obtained from the separate regression of the outcome  $y_{ij1}$  on each column  $k$  of the matrix  $\Delta X_{ij}$ , e.g.  $y_{ij1} = \Delta x_{ijm}\beta_m + \epsilon_{ijm}$ . It is sufficient to show that  $\hat{\rho}_m = \hat{\beta}_m$ . The above equation implies  $\hat{\rho}_m = \frac{\text{Cov}(x_{ijmc}, y_{ijc})}{\text{Var}(x_{ijmc})}$ . Similarly, estimating Equation 4 via least squares without an intercept implies  $\hat{\beta}_m = \frac{\mathbb{E}(\Delta x_{ijm} y_{ij1})}{\mathbb{E}(\Delta x_{ijm}^2)}$ . Since  $\mathbb{E}(\Delta x_{ijm}) = 0$ , it follows that  $\hat{\beta}_m = \frac{\text{Cov}(x_{ijm1} - x_{ijm2}, y_{ij1})}{\text{Var}(x_{ijm1} - x_{ijm2})}$ .

Consider the numerator.

$$\begin{aligned} \text{Cov}(x_{ijm1} - x_{ijm2}, y_{ij1}) &= \text{Cov}(x_{ijm1}, y_{ij1}) - \text{Cov}(x_{ijm2}, y_{ij1}) \\ &= \text{Cov}(x_{ijm1}, y_{ij1}) - \text{Cov}(x_{ijm2}, 1 - y_{ij2}) \\ &= 2\text{Cov}(x_{ijmc}, y_{ijmc}) \end{aligned}$$

The last line follows from the fact that  $\text{Cov}(x_{ijm1}, y_{ij1}) = \text{Cov}(x_{ijm2}, y_{ij2})$

Next consider the denominator.

$$\begin{aligned} \text{Var}(x_{ijm1} - x_{ijm2}) &= \text{Var}(x_{ijm1}) + \text{Var}(-x_{ijm2}) - 2\text{Cov}(x_{ijm1}, x_{ijm2}) \\ &= 2\text{Var}(x_{ijmc}) \end{aligned}$$

Which again follows from the randomization of features. It directly follows that  $\hat{\beta}_m = \hat{\rho}_m = \text{AMCE}$ .

PROOF OF THE EQUIVALENCE OF THE QUADRATIC LOSS AND ABSOLUTE LOSS:

$$(A.22) \quad \begin{aligned} U_i(x_{j1}) &= -|x_{j1} - b_i| + \eta_{ij} \\ U_i(x_{j2}) &= -|x_{j2} - b_i| + \nu_{ij} \end{aligned}$$

Assume  $0 \leq b_i \leq 1$

$$(A.23) \quad \begin{aligned} \Pr(y_{ij1} = 1) &= \Pr(U_i(\mathbf{x}_{ij1}) > U_i(\mathbf{x}_{ij2})) \\ &= \Pr(\eta_{ij} - \nu_{ij} < |x_{j2} - b_i| - |x_{j1} - b_i|) \end{aligned}$$

Since  $x_{j1}$  &  $x_{j2}$  can take on only two values  $\{0, 1\}$ , it follows  $x_{j1} \leq b_i \leq x_{j2}$  or  $x_{j2} \leq b_i \leq x_{j1}$ . This yields:

$$(A.24) \quad \Pr(y_{ij1} = 1) = \Pr(\eta_{ij} - \nu_{ij} < \Delta x_j(2b_i - 1))$$

If we were to estimate this via a linear probability model we obtain

$$(A.25) \quad \begin{aligned} y_{ij1} &= \Delta x_j(2b_i - 1) + \eta_{ij} - \nu_{ij} \\ &= \Delta x_j \beta_i + \epsilon_{ij} \end{aligned}$$

Comparison	V1	V2	V3	V4	V5	Tally
FDB, <b>FDW</b>	FDW	FDW	FDW	FDB	FDB	2,3
FDB, <b>FRB</b>	FRB	FRB	FRB	FDB	FDB	2,3
FDB, <b>FRW</b>	FRW	FRW	FRW	FDB	FDB	2,3
FDB, <b>MDB</b>	MDB	MDB	MDB	FDB	FDB	2,3
FDB, <b>MDW</b>	MDW	MDW	MDW	FDB	FDB	2,3
FDB, <b>MRB</b>	MRB	MRB	MRB	FDB	FDB	2,3
FDB, <b>MRW</b>	MRW	MRW	MRW	FDB	FDB	2,3
FDW, <b>FRB</b>	FRB	FRB	FRB	FRB	FRB	0,5
FDW, <b>FRW</b>	FRW	FRW	FRW	FDW	FDW	2,3
FDW, <b>MDB</b>	MDB	MDB	MDB	MDB	MDB	0,5
FDW, <b>MDW</b>	MDW	MDW	MDW	FDW	FDW	2,3
FDW, <b>MRB</b>	MRB	MRB	MRB	FDW	FDW	2,3
FDW, <b>MRW</b>	MRW	MRW	MRW	FDW	FDW	2,3
FRB, <b>FRW</b>	FRW	FRW	FRW	FRB	FRB	2,3
<b>FRB</b> ,MDB	FRB	FRB	FRB	FRB	FRB	5,0
FRB, <b>MDW</b>	MDW	MDW	MDW	FRB	FRB	2,3
FRB, <b>MRB</b>	MRB	MRB	MRB	FRB	FRB	2,3
FRB, <b>MRW</b>	MRW	MRW	MRW	FRB	FRB	2,3
<b>FRW</b> ,MDB	FRW	FRW	FRW	MDB	MDB	3,2
<b>FRW</b> ,MDW	FRW	FRW	FRW	FRW	FRW	5,0
FRW, <b>MRB</b>	MRB	MRB	MRB	MRB	MRB	0,5
FRW, <b>MRW</b>	MRW	MRW	MRW	FRW	FRW	2,3
MDB, <b>MDW</b>	MDW	MDW	MDW	MDB	MDB	2,3
MDB, <b>MRB</b>	MRB	MRB	MRB	MDB	MDB	2,3
MDB, <b>MRW</b>	MRW	MRW	MRW	MDB	MDB	2,3
MDW, <b>MRB</b>	MRB	MRB	MRB	MRB	MRB	0,5
MDW, <b>MRW</b>	MRW	MRW	MRW	MDW	MDW	2,3
MRB, <b>MRW</b>	MRW	MRW	MRW	MRB	MRB	2,3

TABLE B2—AGGREGATE PREFERENCES OVER CANDIDATE PROFILES - EXAMPLE PART II

AN EXAMPLE OF THE AMCE VIOLATING IIA:

Consider three types of voters with preferences over three candidate-features, Gender ( $M$  or  $F$ ), Age ( $O$  or  $Y$ ), and Race ( $B$  or  $W$ ). Preferences over features are given in Table B4.

Assume priorities over features as follows. V1:  $R \succ A \succ G$ ; V2:  $A \succ R \succ G$ ; V3:  $A \succ G \succ R$ . With this information we can construct preferences over candidates for each type as presented in Table B5.

Consider a population of five V1s, two V2s, and two V3s. Table B6 gives the AMCE estimate with the full set of candidate-features and then restricting each combination of Age and Race. We see that

1.	2.	
$\bar{Y}(MDB, MDB)$	$-\bar{Y}(FDB, MDB)$	$= 1/10$
$\bar{Y}(MDB, FDB)$	$-\bar{Y}(FDB, FDB)$	$= 1/10$
$\bar{Y}(MDB, MRB)$	$-\bar{Y}(FDB, MRB)$	$= 0$
$\bar{Y}(MDB, FRB)$	$-\bar{Y}(FDB, FRB)$	$= -2/5$
$\bar{Y}(MDB, MDW)$	$-\bar{Y}(FDB, MDW)$	$= 0$
$\bar{Y}(MDB, FDW)$	$-\bar{Y}(FDB, FDW)$	$= 3/5$
$\bar{Y}(MDB, MRW)$	$-\bar{Y}(FDB, MRW)$	$= 0$
$\bar{Y}(MDB, FRW)$	$-\bar{Y}(FDB, FRW)$	$= 0$
$\bar{Y}(MRB, MDB)$	$-\bar{Y}(FRB, MDB)$	$= -2/5$
$\bar{Y}(MRB, FDB)$	$-\bar{Y}(FRB, FDB)$	$= 0$
$\bar{Y}(MRB, MRB)$	$-\bar{Y}(FRB, MRB)$	$= 1/10$
$\bar{Y}(MRB, FRB)$	$-\bar{Y}(FRB, FRB)$	$= 1/10$
$\bar{Y}(MRB, MDW)$	$-\bar{Y}(FRB, MDW)$	$= 3/10$
$\bar{Y}(MRB, FDW)$	$-\bar{Y}(FRB, FDW)$	$= -2/5$
$\bar{Y}(MRB, MRW)$	$-\bar{Y}(FRB, MRW)$	$= 0$
$\bar{Y}(MRB, FRW)$	$-\bar{Y}(FRB, FRW)$	$= 3/5$
$\bar{Y}(MDW, MDB)$	$-\bar{Y}(FDW, MDB)$	$= 3/5$
$\bar{Y}(MDW, FDB)$	$-\bar{Y}(FDW, FDB)$	$= 0$
$\bar{Y}(MDW, MRB)$	$-\bar{Y}(FDW, MRB)$	$= -2/10$
$\bar{Y}(MDW, FRB)$	$-\bar{Y}(FDW, FRB)$	$= 3/5$
$\bar{Y}(MDW, MDW)$	$-\bar{Y}(FDW, MDW)$	$= 1/10$
$\bar{Y}(MDW, FDW)$	$-\bar{Y}(FDW, FDW)$	$= 1/10$
$\bar{Y}(MDW, MRW)$	$-\bar{Y}(FDW, MRW)$	$= 0$
$\bar{Y}(MDW, FRW)$	$-\bar{Y}(FDW, FRW)$	$= -2/5$
$\bar{Y}(MRW, MDB)$	$-\bar{Y}(FRW, MDB)$	$= 0$
$\bar{Y}(MRW, FDB)$	$-\bar{Y}(FRW, FDB)$	$= 0$
$\bar{Y}(MRW, MRB)$	$-\bar{Y}(FRW, MRB)$	$= 3/5$
$\bar{Y}(MRW, FRB)$	$-\bar{Y}(FRW, FRB)$	$= 0$
$\bar{Y}(MRW, MDW)$	$-\bar{Y}(FRW, MDW)$	$= -2/5$
$\bar{Y}(MRW, FDW)$	$-\bar{Y}(FRW, FDW)$	$= 0$
$\bar{Y}(MRW, MRW)$	$-\bar{Y}(FRW, MRW)$	$= 1/10$
$\bar{Y}(MRW, FRW)$	$-\bar{Y}(FRW, FRW)$	$= 1/10$
		<b>2</b>

$$\begin{aligned}
 (\# \text{ of profiles} - 1) \times (\# \text{ of features} - 1) &= 28 \\
 \times \# \text{ of values for gender} &
 \end{aligned}$$

$$\text{AMCE} = 1/14$$

TABLE B3—OBTAINING THE AMCE - EXAMPLE II

V1	V2	V3
$M \succ F$	$F \succ M$	$F \succ M$
$O \succ Y$	$Y \succ O$	$Y \succ O$
$B \succ W$	$B \succ W$	$W \succ B$

TABLE B4—PREFERENCES OVER ATTRIBUTES

Rank	V1	V2	V3
1.	MOB	FYB	FYW
2.	FOB	MYB	FYB
3.	MYB	FYW	MYW
4.	MOW	FOB	FOW
5.	FYB	MYW	MYB
6.	FOW	MOB	FOB
7.	MYW	FOW	MOW
8.	FYW	MOW	MOB

TABLE B5—PREFERENCES OVER ATTRIBUTES

the sign flips when we omit either OB and YW, indicating that the AMCE for Male is dependent upon the other feature-combinations, violating IIA.

Omitted Features	O	Y
B	1/168	-4/189
W	-4/189	1/168

No Omitted Features: -1/126

TABLE B6—AMCE ESTIMATES OF MALE, RESTRICTING AGE-RACE FEATURE COMBINATIONS



## ADDITIONAL INFORMATION ABOUT THE ANES

Here we provide the exact wording of the questions we drew from the ANES, with our coding of direction and intensity described below each question.

1) *Favor allowing use of bathrooms of identified gender*

- *Direction (Question V161228)*

Should transgender people—that is, people who identify themselves as the sex or gender different from the one they were born as—have to use the bathrooms of the gender they were born as, or should they be allowed to use the bathrooms of their identified gender?

0 = “1. Have to use the bathrooms of the gender they were born with”

1 = “2. Be allowed to use the bathrooms of their identified gender”

- *Intensity (Question V161228a)*

How strongly do you feel about that?

0 = “3. Slightly”

0.5 = “2. Moderately”

1 = “1. Very strongly”

2) *Favor torture for suspected terrorists*

- *Direction (Question V162295)*

Do you favor, oppose, or neither favor nor oppose the U.S. government torturing people who are suspected of being terrorists, to try to get information?

0 = “2. Oppose”

1 = “1. Favor”

- *Intensity (Questions V162295a, V162295b)*

Do you [favor/oppose] that [a great deal, moderately, or a little / a little, moderately, or a great deal]?

0 = “3. A little”

0.5 = “2. Moderately”

1 = “1. A great deal”

3) *Favor allowing Syrian refugees into US*

- *Direction (Question V161214)*

Do you favor, oppose, or neither favor nor oppose allowing Syrian refugees to come to the United States?

0 = “2. Oppose”

1 = “1. Favor”

- *Intensity (Question V161214a)*

Do you [favor/oppose] that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?

0 = “3. A little”

0.5 = “2. A moderate amount”

1 = “1. A great deal”

4) *Favor 2010 health care law*

- *Direction (Question V161113)*  
Do you favor, oppose, or neither favor nor oppose the health care reform law passed in 2010? This law requires all Americans to buy health insurance and requires health insurance companies to accept everyone.  
0 = “2. Oppose”  
1 = “1. Favor”
- *Intensity (Question V161114a)*  
Do you favor that [a great deal, moderately, or a little / a little, moderately, or a great deal]?  
0 = “3. A little”  
0.5 = “2. Moderately”  
1 = “1. A great deal”

5) *Support preferential hiring/promotion of blacks*

- *Direction (Question V162238)*  
What about your opinion—are you for or against preferential hiring and promotion of blacks?  
0 = “2. Against preferential hiring and promotion of blacks”  
1 = “1. For preferential hiring and promotion of blacks”
- *Intensity (Question V162238a, V162238b)*  
Do you [favor/oppose] preference in hiring and promotion strongly or not strongly?  
0 = “2. Not strongly”  
1 = “1. Strongly”

6) *Favor building a wall with Mexico*

- *Direction (Question V161196)*  
Do you favor, oppose, or neither favor nor oppose building a wall on the U.S. border with Mexico?  
0 = “2. Oppose”  
1 = “1. Favor”
- *Intensity (Question V161196a)*  
Do you [favor/oppose] that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?  
0 = “3. A little”  
0.5 = “2. A moderate amount”  
1 = “1. A great deal”

7) *Favor affirmative action in universities*

- *Direction (Question V161204)*  
Do you favor, oppose, or neither favor nor oppose allowing universities to increase the number of black students studying at their schools by considering race along with other factors when choosing students?  
0 = “2. Oppose”  
1 = “1. Favor”
- *Intensity (Question V161204a, V161204b)*  
Do you [favor/oppose] that [a great deal, a moderate amount, or a little / a little, a

moderate amount, or a great deal]?

0 = "3. A little"

0.5 = "2. A moderate amount"

1 = "1. A great deal"

8) *Favor sending troops to fight ISIS*

- *Direction (Question V161213)*

Do you favor, oppose, or neither favor nor oppose the U.S. sending ground troops to fight Islamic militants, such as ISIS, in Iraq and Syria?

0 = "2. Oppose"

1 = "1. Favor"

- *Intensity (Question V161213a)*

Do you [favor/oppose] that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?

0 = "3. A little"

0.5 = "2. A moderate amount"

1 = "1. A great deal"

9) *Think economy has gotten better since 2008*

- *Direction (Question V161235)*

Would you say that compared to 2008, the nation's economy is now better, worse, or about the same?

0 = "2. Worse"

1 = "1. Better"

- *Intensity (Question V161235a)*

Much [better/worse] or somewhat [better/worse]?

0 = "2. Somewhat"

1 = "1. Much"

10) *Agree that children brought illegally should be sent back*

- *Direction (Question V161195)*

What should happen to immigrants who were brought to the U.S. illegally as children and have lived here for at least 10 years and graduated high school here? Should they be sent back where they came from, or should they be allowed to live and work in the United States?

0 = "2. Should be allowed to live and work in the U.S."

1 = "1. Should be sent back where they came from"

- *Intensity (Question V161195a)*

Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, a great deal]?

0 = "3. A little"

0.5 = "2. A moderate amount"

1 = "1. A great deal"

11) *Think government should make it harder to own a gun*

- *Direction (Question V161187)*  
Do you think the federal government should make it more difficult for people to buy a gun than it is now, make it easier for people to buy a gun, or keep these rules about the same as they are now?  
0 = “2. Easier”  
1 = “1. More difficult”
- *Intensity (Question V161188)*  
How important is this issue to you personally?  
0 = “5. Not important at all”  
0.25 = “4. Not too important”  
0.5 = “3. Somewhat important”  
0.75 = “2. Very important”  
1 = “1. Extremely important”

12) *Approve of House incumbent*

- *Direction (Question V162114)*  
In general, do you approve or disapprove of the way [PRELOAD: name of U.S. House Representative preceding the election] has been handling [his/her] job?  
0 = “2. Disapprove”  
1 = “1. Approve”
- *Intensity (Question V162114a, V162114b)*  
Do you [approve/disapprove] strongly or not strongly?  
0 = “2. Not strongly”  
1 = “1. Strongly”

13) *Favor ending birthright citizenship*

- *Direction (Question V161193)*  
Some people have proposed that the U.S. Constitution should be changed so that the children of unauthorized immigrants do not automatically get citizenship if they are born in this country. Do you favor, oppose, or neither favor nor oppose this proposal?  
0 = “2. Oppose”  
1 = “1. Favor”
- *Intensity (Question V161193a)*  
Do you [favor/oppose] that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?  
0 = “3. A little”  
0.5 = “2. A moderate amount”  
1 = “1. A great deal”

14) *Favor requiring provision of services to same-sex couples*

- *Direction (Question V161227)*  
Do you think business owners who provide wedding-related services should be allowed to refuse services to same-sex couples if same-sex marriage violates their religious beliefs, or do you think business owners should be required to provide services regardless of a couple’s sexual orientation?  
0 = “1. Should be allowed to refuse”  
1 = “2. Should be required to provide services”

- *Intensity (Question V161227a)*  
How strongly do you feel that way?  
0 = “3. A little”  
0.5 = “2. Moderately”  
1 = “1. Very strongly”
- 15) *Favor laws protecting gays against job discrimination*
- *Direction (Question V161229)*  
Do you favor or oppose laws to protect gays and lesbians against job discrimination?  
0 = “2. Oppose”  
1 = “1. Favor”
  - *Intensity (Question V161229a)*  
Do you [favor/oppose] such laws strongly or not strongly?  
0 = “2. Not strongly”  
1 = “1. Strongly”
- 16) *Think government should take more action on climate change*
- *Direction (Question V161224)*  
Do you think the federal government should be doing more about rising temperatures, should be doing less, or is it currently doing the right amount?  
0 = “2. Should be doing less”  
1 = “1. Should be doing more”
  - *Intensity (Question V161224a)*  
Should it be doing a great deal [more/less], a moderate amount [more/less], or a little [more/less]?  
0 = “3. A little”  
0.5 = “2. A moderate amount”  
1 = “1. A great deal”
- 17) *Favor requiring employers to give paid leave to new parents*
- *Direction (Question V161226)*  
Do you favor/oppose, or neither favor nor oppose requiring employers to offer paid leave to parents of new children?  
0 = “2. Oppose”  
1 = “1. Favor”
  - *Intensity (Question V161226a)*  
Do you [favor/oppose] that a great deal, a moderate amount, or a little ?  
0 = “3. A little”  
0.5 = “2. A moderate amount”  
1 = “1. A great deal”
- 18) *Favor vaccines in schools*
- *Direction (Question V162146)*  
Do you favor, oppose, or neither favor nor oppose requiring children to be vaccinated in order to attend public schools?  
0 = “2. Oppose”  
1 = “1. Favor”

- *Intensity (Question V162147)*  
Do you [favor/oppose] that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?  
0 = “3. A little”  
0.5 = “2. A moderate amount”  
1 = “1. A great deal”

19) *Support requiring equal pay for men and women*

- *Direction (Question V162149)*  
Do you favor, oppose, or neither favor nor oppose requiring employers to pay women and men the same amount for the same work?  
0 = “2. Oppose”  
1 = “1. Favor”
- *Intensity (Question V162150)*  
Do you favor that [a great deal, a moderate amount, or a little / a little, a moderate amount, or a great deal]?  
0 = “3. A little”  
0.5 = “2. A moderate amount”  
1 = “1. A great deal”

20) *Favor the death penalty*

- *Direction (Question V161233)*  
Do you favor or oppose the death penalty for persons convicted of murder?  
0 = “2. Oppose”  
1 = “1. Favor”
- *Intensity (Question V161233a)*  
Do you [favor / oppose] the death penalty for persons convicted of murder strongly or not strongly?  
0 = “2. Not strongly”  
1 = “1. Strongly”

21) *Believe benefits of vaccination outweigh risks*

- *Direction (Question V162161)*  
Do the health benefits of vaccinations generally outweigh the risks, do the risks outweigh the benefits, or is there no difference?  
0 = “2. Risks outweigh benefits”  
1 = “1. Benefits outweigh risks”
- *Intensity (Question V162162)*  
Are the [health benefits/risks] [much greater, moderately greater, or slightly greater / slightly greater, moderately greater, or much greater]?  
0 = “3. Slightly greater”  
0.5 = “2. Moderately greater”  
1 = “1. Much greater”

22) *The term ‘feminist’ describes you extremely/very well*

- *Direction (Questions V161346, V161348)*  
Here, we coded one variable from two questions.

- a) 0 = Responded with “2. Very well” or “1. Extremely well” to the question “How well does the term ‘anti-feminist’ describe you?” (V161348)
- b) 1 = Responded with “2. Very well” or “1. Extremely well” to the question “How well does the term ‘feminist’ describe you?” (V161346)
- *Intensity (Question V161347, V161349)*  
How important is it to you to be a [feminist/anti-feminist]?
  - 0 = “5. Not important at all”
  - 0.25 = “4. Not too important”
  - 0.5 = “3. Somewhat important”
  - 0.75 = “2. Very important”
  - 1 = “1. Extremely important”