Testing for Bias in Order Assignment

with an Application to Texas Election Ballots*

Sheridan Grant†, Michael D. Perlman‡, Darren Grant§
University of Washington and Sam Houston State University

June 18, 2018

Abstract

Statistical methods are developed for assessing the likelihood of prejudicial bias in agent-assigned permutations, such as the ordering of candidates on an election ballot. The null hypothesis of an unbiased order assignment is represented by several forms of probabilistic exchangeability of the random orderings, while bias is represented either by compatibility with an assumed ranking of the items with respect to a hypothesized preference criterion (PC) or by linear concordance with assumed scores of the items on a PC scale. These methods are applied to the ordering of candidates on 2014 Texas Republican primary election ballots. Significant evidence of bias is found in three of the five races studied; in two of these races significant evidence of bias is found in at least six of Texas's 254 counties.

*Key words: Elections; ballot-order effect; primacy effect; bias; permutation; exchangeability; ranking; linear concordance score; confidence bound; hypothesis test. Declarations of interest: none.

†Department of Statistics, UW; slgstats@uw.edu. Research supported in part by National Institutes of Health Grant HHSN268201-600310A.

‡Department of Statistics, UW; mdperlma@uw.edu. Research supported in part by U.S. Department of Defense Grant H98230-10-C-0263.

§Department of Economics and International Business, SHSU; dgrant@shsu.edu.

1. Introduction.

Each year an unusual ritual takes place in school district offices, city halls, and county courthouses across the state of Texas: the drawing of the order in which candidates for public office will be placed on the ballot, as required by state law. Candidates often attend these drawings to ensure that the agent conducting them does not manipulate the ordering and place a competitor higher on the ballot, conferring upon them an electoral advantage known as the *ballot-order effect*. In Texas primary and runoff elections for statewide office Grant (2017) finds this effect to be sizeable and monotonic in ballot order, especially in low-profile or low-information races. This finding is corroborated for other states by several studies cited therein, while Meredith and Salant (2013) obtain broadly similar results for local elections in California.

The possibility that such orderings might be prejudicially biased, either consciously or unconsiously, by the agents executing them is not far-fetched. Darcy and McAllister (1990) noted the evidence for such bias in their review of the early literature on the ballot order effect.[1] One can test statistically for such bias when orderings (= permutations) of the same set of k items are conducted repeatedly and (presumably) independently N times, as in Texas, where ballot order for primary and runoff elections is determined by agents at the county level, even for offices contested statewide. A test for uniform random ordering when $k = 2$ is elementary, but this is not the case when multiple items are being ordered.

The general problem of testing for the uniform randomness of permutations arises in a variety of applications, notably testing random number generators (Knuth (1981)).[2] Despite the generality of the problem, however, a consensus on testing methodology has not emerged. Even within the ballot order literature, a variety of options are used. Grant (2017) applies Fisher's Exact Test to the cross-tabulation of candidates and ballot positions in Texas, to determine if this cross-tabulation is likely to have occurred

[1] In one extreme example, Grant (2017) notes that in the 2014 Texas Democratic gubernatorial (two-candidate) runoff, the candidate disfavored by the party establishment–she had likened President Obama to Hitler during the campaign–was listed first on the ballot in only about one-quarter of the state's counties.

[2] Tests developed in this context–typically chi-square tests for each of the $k!$ permutations being equally likely–are designed to be extremely sensitive to any departure from uniform randomness but require extremely large sample sizes, a luxury social scientists may not have.

by random chance. Meredith and Salant (2013) apply a variant of the chi-squared goodness of fit test to determine if incumbent candidates are equally likely to end up at any position on the ballot. And Ho and Imai (2008) apply a series of rank tests to randomized alphabets used for ballot ordering in California, testing statistics such as the average absolute difference in rank between pairs of letters, to see if they differ significantly from the null mean.

All of these approaches have limitations. The first procedure, Fisher's Exact Test, aggregates the N observed orderings into a $k \times k$ contingency table, where k is the number of items and order positions. Among other problems,[3] such aggregation loses relevant information contained in the orderings themselves. It is possible–and, in political applications, probable–that agents with opposing prejudicial biases manipulate orderings in opposite directions, but these offsetting manipulations may not be apparent in the aggregated table. The chi-squared goodness of fit-type test also ignores important information: by aggregating all of the candidates into incumbents and non-incumbents, it avoids Fisher's Exact Test's problems with dependency in the counts, but it loses all information about individuals.

In addition, unless k is very small or N very large, these tests, along with those of Ho and Imai (2008), may lack the sensitivity needed to detect the specific deviations from uniform randomness that may be encountered. This limited sensitivity would derive, in part, from the omnibus nature of such tests, which do not utilize *a priori* information that could increase their power. Such information is often available in political and other applications in which human agents perform the orderings; any deviations from uniform randomness are likely to reflect these agents' preferences.

In this paper, procedures are developed that utilize the individual orderings and not their aggregation, and that will detect the departures from uniform randomness to be expected from *a priori* information about the characteristics of the items and the preference criteria of the agents executing the ordering. In Section 2, the null hypothesis of an unbiased order assignment is represented by several forms of exchangeability of a random permutation. In Section 3, the alternative hypothesis of bias in order assignment is repre-

[3]Each agent's ordering contributes k dependent counts, rather than a single count, to the cross-tabulation, violating the multinomial distribution model for the aggregated counts that is assumed by Fisher's Exact Test. Since the test ignores the dependence among the counts, it is anti-conservative. See Appendix D for a simulation illustration of this phenomenon.

sented by compatibility with an assumed preferential ranking (ties permitted) of the items, while in Section 4 bias is represented by linear concordance with assumed preference scores of the items. In both cases methods for detecting the corresponding alternatives are obtained. Appendix C illustrates these methods' ability to detect prejudicial bias even for relatively small sample sizes.

In Section 5 these procedures are applied to five races in the 2014 Texas Republican primary. Significant evidence of bias in at least one of the approximately 245 reporting counties is found in three of the five races; in two of these, significant evidence is found for bias in at least six and ten counties.

Although developed for the ballot ordering process, the subject of our empirical application, these methods apply far more broadly. The ballot order effect is an example of a more general psychological phenomenon, the primacy effect (cf. Murdock, 1962) in which the first-listed of a set of options tends to be chosen more frequently. Thus, in many scenarios in which a set of competing decisions must be made sequentially without prejudice, the agents ordering those decisions may be tempted to manipulate the orderings in accordance with their preferences. In such cases our procedures would be directly applicable.

2. Unbiased and biased order assignments.

Suppose that k items, labeled $1, \dots, k$ in alphabetical or numerical order, are being ordered by each of N agents. An ordering is simply a permutation $\pi \equiv (\pi_1, \dots, \pi_k)$ of $(1, \dots, k)$ such that item i is assigned position π_i, $i = 1, \dots, k$. The primacy effect implies that positions near 1 ($\pi_i \approx 1$) in the ordering are advantageous, while positions near k ($\pi_i \approx k$) are disadvantageous.

We shall assume that the ordering π selected by an agent is a realization of a random permutation $\Pi \equiv (\Pi_1, \dots, \Pi_k)$ (quite possibly degenerate). The gold standard for unbiased order assignment is the null hypothesis H_0 that Π is *uniformly* random, i.e., $\Pi \sim \mathbf{U}_0$, the uniform distribution over all $k!$ permutations $\pi \equiv (\pi_1, \dots, \pi_k)$ of $(1, \dots, k)$. Equivalently, Π is *exchangeable*, i.e.,

$$(1) \qquad H_0 : (\Pi_1, \dots, \Pi_k) \sim (\Pi_{\pi_1}, \dots, \Pi_{\pi_k})$$

for every $\pi \equiv (\pi_1, \dots, \pi_k)$ of $(1, \dots, k)$.[4]

[4]Several states, including Texas, California, and Virginia, require that the process by

A less restrictive hypothesis H_1 that still indicates a degree of unbiasedness is that of *second-order exchangeability*, so that Π satisfies

$$(2) \qquad H_1 : \begin{cases} \mathrm{E}(\Pi_1, \ldots, \Pi_k) = \mathrm{E}(\Pi_{\pi_1}, \ldots, \Pi_{\pi_k}), \\ \mathrm{Cov}(\Pi_1, \ldots, \Pi_k) = \mathrm{Cov}(\Pi_{\pi_1}, \ldots, \Pi_{\pi_k}), \end{cases}$$

for all π, where Cov denotes the covariance matrix.[5] That is, the first and second moments (mean and variance) are constant across positions, and the relationship between pairs of positions is also constant. Clearly, exchangeability implies second-order exchangeability, i.e., $H_0 \Rightarrow H_1$, but not conversely: for $k = 3$, for example, the uniform distribution over $(1, 2, 3)$, $(3, 1, 2)$, and $(2, 3, 1)$ is second-order exchangeable but not exchangeable.[6]

We will also apply a restriction of the previous hypothesis H_1: the intermediate hypothesis H_2 of *fourth-order exchangeability*. That is, Π satisfies

$$(3) \qquad H_2 : \mathrm{E}\big(\Pi_1^{r_1} \ldots \Pi_k^{r_k}\big) = \mathrm{E}\big(\Pi_{\pi_1}^{r_1} \ldots \Pi_{\pi_k}^{r_k}\big)$$

for all choices of nonnegative integers $r_1, \ldots, r_k$ such that $r_1 + \cdots + r_k \leq 4$.

Clearly $H_0 \Rightarrow H_2 \Rightarrow H_1$. The hypotheses H_1 and H_2 of higher-order exchangeability describe forms of symmetry in the order-generating distribution that are weaker than full uniformity and allow us to demonstrate that our methods apply in more general settings than when the null hypothesis is simple exchangeability, though in many applications H_0 will be sufficient.

We shall develop methods for testing the null hypothesis that H_0, H_1, or H_2 holds for all N agents, against alternative hypotheses that represent bias in terms of compatibility of the ordering with a partitioning of the items that is derived from a conjectured preference criterion (PC). A PC may be expressed either qualitatively via a ranking (with ties permitted) of the items, or quantitatively via numerical scoring; of course, a ranking can be assessed immediately from a scoring.

---

which ballot orders are generated satisfies H_0 for at least some of their elections, cf. Krosnick, Miller, and Tichy (2004).

[5]Equivalently, $\mathrm{E}(\Pi_i) = \mathrm{E}(\Pi_j)$ for all $i, j \in \{1, \ldots, k\}$, $\mathrm{Var}(\Pi_i) = \mathrm{Var}(\Pi_j)$ for all $i, j \in \{1, \ldots, k\}$, and $\mathrm{Cov}(\Pi_i, \Pi_j) = \mathrm{Cov}(\Pi_{i'}, \Pi_{j'})$ for all $i \neq j$ and $i' \neq j' \in \{1, \ldots, k\}$.

[6]This is the uniform distribution on the set of all cyclic permutations of order $k = 3$. The uniform distribution over the set of cyclic permutations of order $k \geq 4$ is not second-order exchangeable.

In our framework, a PC can be either uni-directional or bi-directional. It is uni-directional when the vast majority of observers have the same preference; this occurs in the election ballot ordering context when the PC is based on name recognition or experience level, for example. The PC is bi-directional when observers may prefer one end of a spectrum to the other, e.g., politically conservative-to-progressive in the ballot-ordering example.[7]

The case where the PC is expressed by a ranking is treated in Section 3. There the evidence for bias is simply the number of agents whose assigned orderings are fully compatible with the conjectured ranking of the items. The overall extent of this rank-compatibility (RC) bias is expressed by lower confidence bounds for the number of agents with actual RC bias.

Section 4 treats the case where the PC is expressed via scores, either uni- or bi-directional. There the evidence for bias is measured by the linear concordance (LC) of the observed ordering with the vector of scores. The strength of this evidence is measured by p-values from LC tests designed to detect bias in one or more agents. Unlike the RC statistics, the score-based LC statistics are robust to small departures from the conjectured ranking, that is, they maintain sensitivity at alternatives that are near, but not identical to, the assumed ranking of the items.

The RC and LC methods are applied to 2014 Texas election data in Section 5. Proofs of some results are deferred to the Appendices.

[7]In the bi-directional setting, it is assumed that if an agent is in fact biased, then her/his preference is monotonically increasing or decreasing along the PC spectrum. This does not admit a "middle-ground" case where intermediate items may be favored, thereby introducing a "bias" that prefers moderation. Our assumption simplifies the analysis and is likely to be approximately correct in highly polarized climates, or, for example, when an election has two major candidates and a collection of minor candidates. Future work will address such a bias for moderation.

3. Preference criteria (PC) expressed by ranks.

Suppose that an agent ranks the k items (with ties permitted) in accordance with a particular PC, so that a low (high) rank indicates agreement (disagreement) with the PC. This yields a partitioning $\mathcal{B} \equiv (B_1, \ldots, B_r)$ of the items into blocks $B_1, \ldots, B_r$ of sizes $k_1, \ldots, k_r$ ($r \geq 2$, $k_1 + \cdots + k_r = k$). Items in block B_j have a lower (the same) PC ranking than (as) those in $B_{j'}$ if $j < j'$ ($j = j'$). If each $k_j = 1$ then the items are totally ordered w.r. to the PC, while if one or more $k_j > 1$ then they are partially ordered.[8]

An ordering given by a permutation $\pi \equiv (\pi_1, \ldots, \pi_k)$ is *compatible* with the partitioning $\mathcal{B}$ if

$$i \in B_j \iff \pi_i \in B_j. \tag{4}$$

The set of all $k_1! \cdots k_r!$ permutations compatible with $\mathcal{B}$ is denoted by $\mathcal{P}(\mathcal{B})$.

3.1. A uni-directional alternative based on ranks. For a uni-directional PC and associated partitioning $\mathcal{B}$, if no factors other than PC influence order assignment then it is reasonable to represent an agent's bias for PC by the alternative hypothesis

$$A_{\mathcal{B}} : \Pi \sim \mathbf{U}_{\mathcal{B}} \equiv \mathbf{U}(k_1) \otimes \cdots \otimes \mathbf{U}(k_r)\,, \tag{5}$$

the uniform distribution over $\mathcal{P}(\mathcal{B})$. Here $\mathbf{U}(k_j)$ denotes the uniform distribution over all $k_j!$ permutations of B_j and $\otimes$ denotes independence among blocks. Under $A_{\mathcal{B}}$, Π is rank-compatible (RC) ($\Pr[\Pi \in \mathcal{P}(\mathcal{B})] = 1$) and assigns equal probability $1/k_1! \cdots k_r!$ to each permutation in $\mathcal{P}(\mathcal{B})$. Under H_0,

$$\Pr[\Pi \in \mathcal{P}(\mathcal{B})] = \frac{k_1! \cdots k_r!}{k!} \equiv p_{\mathcal{B}}. \tag{6}$$

Let $\Pi^{(1)}, \ldots, \Pi^{(N)}$ denote the random orderings from agents $1, \ldots, N$. It is assumed that $\Pi^{(1)}, \ldots, \Pi^{(N)}$ are independent but not necessarily identically distributed: $\Pi^{(n)}$ is drawn either according to H_0 or to $A_{\mathcal{B}}$ if the nth agent is biased or not. Let $\mathcal{N}_{\mathcal{B}}$ ($\emptyset \subseteq \mathcal{N}_{\mathcal{B}} \subseteq \{1, \ldots, N\}$) denote the subset of the N agents who are biased as specified by $A_{\mathcal{B}}$, so

$$\Pi^{(n)} \sim \begin{cases} \mathbf{U}_0 & \text{if } n \notin \mathcal{N}_{\mathcal{B}}, \\ \mathbf{U}_{\mathcal{B}} & \text{if } n \in \mathcal{N}_{\mathcal{B}}. \end{cases} \tag{7}$$

[8] For candidates on an election ballot, each of the one or two most prominent candidates might determine her/his own singleton block, whereas candidates who are not well-known by the public and/or hard to distinguish ideologically might be grouped into a single block.

Here $\mathcal{N}_\mathcal{B}$ itself is treated as the unknown parameter. The extent to which bias is present is represented by $\tau_\mathcal{B} \equiv |\mathcal{N}_\mathcal{B}|$ (the cardinality of $\mathcal{N}_\mathcal{B}$), where $0 \le \tau_\mathcal{B} \le N$.

For each $n = 1, \dots, N$, the likelihood ratio for $\Pi^{(n)}$ is given by

$$\frac{\Pr_{A_\mathcal{B}}[\Pi^{(n)} = \pi]}{\Pr_{H_0}[\Pi^{(n)} = \pi]} = \frac{k!}{k_1! \cdots k_r!} I_{\mathcal{P}(\mathcal{B})}(\pi), \tag{8}$$

hence the MLE of $\mathcal{N}_\mathcal{B}$ is given by

$$\hat{\mathcal{N}_\mathcal{B}} \equiv \hat{\mathcal{N}_\mathcal{B}}(\Pi^{(1)}, \dots, \Pi^{(N)}) \equiv \{n \mid \Pi^{(n)} \in \mathcal{P}(\mathcal{B})\}. \tag{9}$$

Therefore the MLE of $\tau_\mathcal{B}$ is given by the *rank-compatibility (RC) statistic*

$$\hat{\tau}_\mathcal{B} \equiv |\hat{\mathcal{N}_\mathcal{B}}| = \sum\nolimits_{n=1}^{N} I_{\mathcal{P}(\mathcal{B})}(\Pi^{(n)}) \tag{10}$$

and the distribution of $\hat{\tau}_\mathcal{B}$ depends on $\mathcal{N}_\mathcal{B}$ only through $\tau_\mathcal{B}$:

$$\hat{\tau}_\mathcal{B} \sim \tau_\mathcal{B} + \text{Binomial}(N - \tau_\mathcal{B},\, p_\mathcal{B}), \quad \tau_\mathcal{B} = 0, 1, \dots, N. \tag{11}$$

Here $p_\mathcal{B}$ is known, $\tau_\mathcal{B}$ is the unknown parameter to be inferred, and $\hat{\tau}_\mathcal{B}$ is the number of biased agents plus those unbiased agents where an apparently RC ordering occurs merely through random chance (with probability $p_\mathcal{B}$).

It is straightforward to obtain a lower $(1-\alpha)$ confidence bound for $\tau_\mathcal{B}$. For any integer $b = 0, 1, \dots, N$, it follows from (11) that

$$\begin{aligned}
\Pr[\hat{\tau}_\mathcal{B} - b \le \tau_\mathcal{B}] &= \Pr[\text{Binomial}(N - \tau_\mathcal{B},\, p_\mathcal{B}) \le b] && (12)\\
&\ge \Pr[\text{Binomial}(N, p_\mathcal{B}) \le b] \\
&= \sum_{n=0}^{b} \binom{N}{n} p_\mathcal{B}^l (1 - p_\mathcal{B})^{N-n} \\
&\equiv \gamma(b; N, p_\mathcal{B}). && (13)
\end{aligned}$$

Therefore $\hat{\tau}_\mathcal{B} - b$ is a (conservative) lower $\gamma(b; N, p_\mathcal{B})$ confidence bound for $\tau_\mathcal{B}$. Clearly the confidence probability bound $\gamma(b; N, p_\mathcal{B})$ increases with b and decreases with N and $p_\mathcal{B}$.

Example 3.1. If $k = 8$, $r = 3$, $k_1 = k_3 = 3$, and $k_2 = 2$, then $p_\mathcal{B} = 3!2!3!/8! = 1/560 \approx .0018$. For $N \in \{50, 150, 254^9\}$ we compute the following lower bounds for the confidence probability of $\hat{\tau}_\mathcal{B} - b$:

[9] There are 254 Texas counties, each having an agent ordering a ballot, so $N = 254$ in this application.

b	0	1	2	3
$\gamma(b; 254, 1/560)$	.635	.924	.989	.999
$\gamma(b; 150, 1/560)$	.765	.970	.997	1.00
$\gamma(b; 50, 1/560)$	.915	.996	1.00	1.00

Example 3.2. If $k = 8$, $r = 4$, $k_1 = k_2 = k_3 = k_4 = 2$, then $p_{\mathcal{B}} = (2!)^4/8! = 1/2520 \approx .00040$:

b	0	1	2	3
$\gamma(b; 254, 1/2520)$	.904	.995	.999	1.00
$\gamma(b; 150, 1/2520)$	.942	.998	1.00	1.00
$\gamma(b; 50, 1/2520)$	.980	1.00	1.00	1.00

Example 3.3. If $k = 8$, $r = 2$, and $k_1 = k_2 = 4$, then $p_{\mathcal{B}} = (4!)^2/8! = 1/70 \approx .0143$:

b	0	1	2	3	4	5	6	7	8	9	10
$\gamma(b; 254, 1/70)$	.023	.121	.296	.508	.701	.841	.926	.969	.988	.996	.999
$\gamma(b; 150, 1/70)$	.115	.367	.638	.832	.935	.979	.994	.998	1.00	1.00	1.00
$\gamma(b; 50, 1/70)$	.487	.840	.965	.994	.999	1.00	1.00	1.00	1.00	1.00	1.00

The tightest lower confidence bound of the form $\hat{\tau}_{\mathcal{B}} - b$ is $\hat{\tau}_{\mathcal{B}}$ itself, attained when $b = 0$, with a confidence probability bound $\gamma(0; N, p_{\mathcal{B}}) = (1 - p_{\mathcal{B}})^N$. This exceeds a target value $1 - \alpha$ if

$$p_{\mathcal{B}} \leq 1 - (1 - \alpha)^{1/N}. \tag{14}$$

Values of $p_{\mathcal{B}}$ below which $\hat{\tau}_{\mathcal{B}}$ is a (conservative) $(1 - \alpha)$ lower confidence bound for τ are shown here for $\alpha = .10, .05$ and $.01$:

N	50	150	254
$1 - (.90)^{1/N}$	.0021	.00070	.00041
$1 - (.95)^{1/N}$	.0010	.00034	.00020
$1 - (.99)^{1/N}$	.0002	.00007	.00004

This shows that very small values of $p_{\mathcal{B}}$ may be required for $\hat{\tau}_{\mathcal{B}}$ itself to be a satisfactory lower confidence bound for $\tau_{\mathcal{B}}$. Thus a bound $\hat{\tau}_{\mathcal{B}} - b$ with $b \geq 1$ is usually preferable.

3.2. A bi-directional alternative based on ranks. If the PC is bi-directional and an associated partitioning $\mathcal{B} \equiv (B_1, \ldots, B_r)$ of the k items is specified as in §3.1, then bias in an ordering Π can appear either as compatibility with $\mathcal{B}$, i.e. $\Pi \in \mathcal{P}(\mathcal{B})$, or else as compatibility with its opposite $\tilde{\mathcal{B}}$, i.e. $\Pi \in \mathcal{P}(\tilde{\mathcal{B}})$, where $\tilde{\mathcal{B}} \equiv (\tilde{B}_1, \ldots, \tilde{B}_r) \equiv (B_r, \ldots, B_1)$ and block $\tilde{B}_j$ has size $\tilde{k}_j = k_{k-j+1}$. Define $\mathbf{U}_{\tilde{\mathcal{B}}}$, $A_{\tilde{\mathcal{B}}}$, $p_{\tilde{\mathcal{B}}}$, $\mathcal{N}_{\tilde{\mathcal{B}}}$, $\tau_{\tilde{\mathcal{B}}}$, $\hat{\mathcal{N}}_{\mathcal{B}}$, and $\hat{\tau}_{\tilde{\mathcal{B}}}$ as above with $(\mathcal{B}, B_j, k_j)$ replaced by $(\tilde{\mathcal{B}}, \tilde{B}_j, \tilde{k}_j)$. Note that $\mathcal{P}(\mathcal{B}) \cap \mathcal{P}(\tilde{\mathcal{B}}) = \emptyset$, $\mathcal{N}(\mathcal{B}) \cap \mathcal{N}(\tilde{\mathcal{B}}) = \emptyset$, and $\hat{\mathcal{N}}(\mathcal{B}) \cap \hat{\mathcal{N}}(\tilde{\mathcal{B}}) = \emptyset$. Clearly $p_{\mathcal{B}} = p_{\tilde{\mathcal{B}}}$.

As before, assume that the orderings $\Pi^{(1)}, \ldots, \Pi^{(N)}$ from agents $1, \ldots, N$ are independent, where now each $\Pi^{(n)}$ is drawn either according to H_0, $A_{\mathcal{B}}$, or $A_{\tilde{\mathcal{B}}}$ depending on whether, and in which direction, the nth is biased. Thus

$$\Pi^{(n)} \sim \begin{cases} \mathbf{U}_0 & if\ n \notin \mathcal{N}_{\mathcal{B}} \cup \mathcal{N}_{\tilde{\mathcal{B}}}, \\ \mathbf{U}_{\mathcal{B}} & if\ n \in \mathcal{N}_{\mathcal{B}}, \\ \mathbf{U}_{\tilde{\mathcal{B}}} & if\ n \in \mathcal{N}_{\tilde{\mathcal{B}}}, \end{cases} \tag{15}$$

so the pair $(\mathcal{N}_{\mathcal{B}}, \mathcal{N}_{\tilde{\mathcal{B}}})$ is the unknown parameter, with MLE $(\hat{\mathcal{N}}_{\mathcal{B}}, \hat{\mathcal{N}}_{\tilde{\mathcal{B}}})$. The extent to which bias for (respectively, against) the PC is measured by $\tau_{\mathcal{B}} \equiv |\mathcal{N}_{\mathcal{B}}|$ ($\tau_{\tilde{\mathcal{B}}} \equiv |\mathcal{N}_{\tilde{\mathcal{B}}}|$), where $0 \leq \tau_{\mathcal{B}} + \tau_{\tilde{\mathcal{B}}} \leq N$. The MLE of $(\tau_{\mathcal{B}}, \tau_{\tilde{\mathcal{B}}})$ is the pair of RC statistics $(\hat{\tau}_{\mathcal{B}}, \hat{\tau}_{\tilde{\mathcal{B}}}) \equiv (|\hat{\mathcal{N}}_{\mathcal{B}}|, |\hat{\mathcal{N}}_{\tilde{\mathcal{B}}}|)$ with distribution given by

$$(\hat{\tau}_{\mathcal{B}}, \hat{\tau}_{\tilde{\mathcal{B}}}) \sim (\tau_{\mathcal{B}}, \tau_{\tilde{\mathcal{B}}}) + \text{Trinomial}\left(N - \tau_{\mathcal{B}} - \tau_{\tilde{\mathcal{B}}};\, p_{\mathcal{B}}, p_{\mathcal{B}}\right), \tag{16}$$

the (incomplete) trinomial distribution based on $N - \tau_{\mathcal{B}} - \tau_{\tilde{\mathcal{B}}}$ trials, with cell probabilities $(p_{\mathcal{B}}, p_{\mathcal{B}}, 1 - 2p_{\mathcal{B}})$.

Lower confidence bounds for $\tau_{\mathcal{B}}$ and $\tau_{\tilde{\mathcal{B}}}$ can be obtained as in (12)-(13):

$$\Pr[\hat{\tau}_{\mathcal{B}} - b \leq \tau_{\mathcal{B}}] \geq \gamma(b; N, p_{\mathcal{B}}), \tag{17}$$

$$\Pr[\hat{\tau}_{\tilde{\mathcal{B}}} - \tilde{b} \leq \tau_{\tilde{\mathcal{B}}}] \geq \gamma(\tilde{b}; N, p_{\mathcal{B}}). \tag{18}$$

Similarly, a lower confidence bound can be obtained for $\tau_{\mathcal{B}} + \tau_{\tilde{\mathcal{B}}}$, the total number of agents biased either for or against the PC: for any integer $c \in \{0, 1, \ldots, N\}$,

$$\Pr[\hat{\tau}_{\mathcal{B}} + \hat{\tau}_{\tilde{\mathcal{B}}} - c \leq \tau_{\mathcal{B}} + \tau_{\tilde{\mathcal{B}}}] \geq \gamma(c; N, 2p_{\mathcal{B}}). \tag{19}$$

Conservative simultaneous lower confidence bounds for $\tau_{\mathcal{B}}$ and $\tau_{\tilde{\mathcal{B}}}$ can be obtained via Bonferroni's inequality:

$$\Pr[\hat{\tau}_{\mathcal{B}} - b \leq \tau_{\mathcal{B}},\, \hat{\tau}_{\tilde{\mathcal{B}}} - \tilde{b} \leq \tau_{\tilde{\mathcal{B}}}] \geq \gamma(b; N, p_{\mathcal{B}}) + \gamma(\tilde{b}; N, p_{\tilde{\mathcal{B}}}) - 1. \tag{20}$$

4. Preference criteria (PC) expressed by scores.

In Section 3 it was assumed that if an agent is unbiased, he or she will select a random ordering according to the exchangeable model H_0, while if biased, he or she will select an ordering that conforms exactly to a partitioning $\mathcal{B}$ and/or its opposite $\tilde{\mathcal{B}}$, specified by a uni-directional or bi-directional preference criterion (PC). In some cases, however, this may be a simplification of the true behavioral processes that generate bias. Rather than only two options–randomizing the order or conforming fully to the PC–there may be intermediate cases in which orderings are generated with a partial degree of conformity with the PC. These "nearby orderings" could occur because of minor differences in biased agents' preferences, randomness in the process that generates biased orderings, or other small frictions in the generation of a biased ordering from a PC.

While the detection methods based on the RC statistics $\hat{\tau}_{\mathcal{B}}$ and/or $\hat{\tau}_{\tilde{\mathcal{B}}}$ are sensitive to orderings that conform exactly to the specified partitioning $\mathcal{B}$ and/or its opposite $\tilde{\mathcal{B}}$, they do not detect nearby orderings. Therefore, we now present procedures that are sensitive both to orderings in $\mathcal{B}$ and/or $\tilde{\mathcal{B}}$ and also to such nearby orderings.

4.1. A uni-directional alternative based on scores. Suppose that, in quantifying the PC, the items $1, \ldots, k$ are assigned quantitative scores $s_1, \ldots, s_k$ (rather than qualitative ranks) that indicate agents' preferences across items. As with ranks, it is assumed that small (large) values of s_i indicate conformity (non-conformity) with the PC, and ties are represented by equalities among the s_i. As in §3, the score vector $\mathbf{s} \equiv (s_1, \ldots, s_k)$ determines a partitioning $\mathcal{B}(\mathbf{s}) \equiv (B_1(\mathbf{s}), \ldots, B_r(\mathbf{s}))$ of $\{1, \ldots, k\}$ into blocks of sizes $k_1, \ldots, k_r$. Items in block $B_j(\mathbf{s})$ have a lower (the same) PC score than (as) those in $B_{j'}(\mathbf{s})$ if and only if $j < j'$ ($j = j'$).

To illustrate, consider Example 3.1 above, where $k = 8$ and $r = 3$. Assume that $k_1 \equiv 3$ items have the same low ($\equiv$ strong) PC score s_1, $k_2 \equiv 2$ have the same intermediate PC score s_2, and $k_3 \equiv 3$ have the same high ($\equiv$ weak) PC score s_3, with no finer distinctions among them. The values of the s_i are either specified *a priori* or chosen (see (57)) to maximize the value of the linear concordance score statistic $\bar{T}$ (see (37)) when bias for PC is present. For now, take the s_i to be the block ranks averaged over ties, so in this example, $s_1 = s_2 = s_3 = 2$, $s_4 = s_5 = 4.5$, $s_6 = s_7 = s_8 = 7$.

As before it is assumed that the order of the items is determined by a random permutation $\Pi \equiv (\Pi_1, \ldots, \Pi_k)$ of $(1, \ldots, k)$. Consider the *linear condordance (LC) score*

$$\sum\nolimits_{i=1}^{k} \Pi_i s_i \equiv \langle \Pi, \mathbf{s} \rangle, \tag{21}$$

the inner product of the position vector Π and the score vector $\mathbf{s}$. This score indicates the degree of linear concordance of Π with $\mathbf{s}$: by the Rearrangement Inequality,

$$\sum i s_{(i)} \geq \sum \Pi_i s_i \geq \sum (k - i + 1) s_{(i)}\,, \tag{22}$$

where $s_{(1)} \leq \cdots \leq s_{(k)}$ are the ordered scores. (Note that $i = \Pi_{(i)}$.) Furthermore, the first (second) inequality in (22) is strict unless $\Pi \in \mathcal{P}(\mathcal{B}(\mathbf{s}))$ ($\Pi \in \mathcal{P}(\tilde{\mathcal{B}}(\mathbf{s}))$). Thus large (small) values of $\langle \Pi, \mathbf{s} \rangle$ indicate concordance (disconcordance) of Π with $\mathbf{s}$.

Recall the well-known identities

$$\sum \Pi_i = \sum i = k(k+1)/2, \tag{23}$$

$$\sum \Pi_i^2 = \sum i^2 = k(k+1)(2k+1)/6. \tag{24}$$

From (23),

$$\sum \Pi_i s_i = \sum \Pi_i (s_i - \bar{s}) + [k(k+1)/2]\bar{s}, \tag{25}$$

where $\bar{s} = (\sum s_i)/k$. Therefore we can impose the restriction that $\bar{s} = 0$ (equivalently, $\sum s_i = 0$) on $\mathbf{s}$ without altering the linear concordance ordering. (Also see Remark 4.1.) Thus in Example 3.1 we now will take $s_1 = s_2 = s_3 = -2.5$, $s_4 = s_5 = 0$, and $s_6 = s_7 = s_8 = 2.5$.

Suppose first that Π is second-order exchangeable, i.e., satisfies H_1 in (2), a less restrictive form of unbiasedness than H_0. Then $\mathrm{E}(\Pi_i)$, $\mathrm{Var}(\Pi_i)$, and therefore $\mathrm{E}(\Pi_i^2)$ do not depend on i, while $\mathrm{Cov}(\Pi_i, \Pi_j)$ does not depend on

i, j ($i \neq j$). From (23) and (24) it follows that for $i = 1, \dots, k$,

$$\text{(26)} \qquad \mathrm{E}(\Pi_i) = (k+1)/2,$$

$$\text{(27)} \qquad \mathrm{E}(\Pi_i^2) = (k+1)(2k+1)/6,$$

$$\text{(28)} \qquad \mathrm{Var}(\Pi_i) = \mathrm{E}(\Pi_i^2) - (\mathrm{E}(\Pi_i))^2 = (k^2-1)/12.$$

Because $\sum \Pi_i$ is constant,

$$0 = \mathrm{Var}\left(\sum \Pi_i\right) = k\mathrm{Var}(\Pi_i) + k(k-1)\mathrm{Cov}(\Pi_i, \Pi_j)$$

for $i \neq j$, so

$$\text{(29)} \qquad \mathrm{Cov}(\Pi_i, \Pi_j) = -(k+1)/12.$$

These moment relations can be expressed in vector form as

$$\text{(30)} \qquad \mathrm{E}(\Pi_1, \dots, \Pi_k) = \tfrac{k+1}{2}\,\mathbf{e},$$

$$\text{(31)} \qquad \mathrm{Cov}(\Pi_1, \dots, \Pi_k) = \tfrac{k+1}{12}(kI - \mathbf{e}'\mathbf{e}),$$

where $\mathbf{e} = (1, \dots, 1) : 1 \times k$ and I is the $k \times k$ identity matrix.

It follows from (30) that if Π satisfies H_1 then[10]

$$\text{(32)} \qquad \mathrm{E}(\langle \Pi, \mathbf{s} \rangle) = \langle \mathrm{E}(\Pi), \mathbf{s} \rangle = \langle \tfrac{k+1}{2}\,\mathbf{e},\, \mathbf{s} \rangle = \tfrac{k+1}{2} \sum s_i = 0.$$

If instead $\Pr[\Pi \in \mathcal{P}(\mathcal{B}(\mathbf{s}))] = 1$ then by the Rearrangement Inequality and (32),

$$\text{(33)} \qquad \langle \Pi, \mathbf{s} \rangle = \max_{\pi} \langle \pi,\, \mathbf{s} \rangle > \mathrm{E}(\langle \Pi_0,\, \mathbf{s} \rangle) = 0\,,$$

where $\Pi_0 \sim \mathbf{U}_0$; strict inequality holds because $\mathrm{Var}(\langle \Pi_0,\, \mathbf{s} \rangle) > 0$ by (38) below. Thus, testing H_1 against the alternative $A_{1,\mathbf{s}} : \Pr[\Pi \in \mathcal{P}(\mathcal{B}(\mathbf{s}))] = 1$ can be accomplished by testing instead

$$\text{(34)} \qquad H_1 \quad \text{vs.} \quad A'_{1,\mathbf{s}} : \mathrm{E}(\langle \Pi,\, \mathbf{s} \rangle) > 0,$$

a uni-directional ($\equiv$ one-sided) alternative, based on the linear concordance score $\langle \Pi,\, \mathbf{s} \rangle$. Here $E(\langle \Pi,\, \mathbf{s} \rangle)$ provides a measure of the mean linear concordance of Π with the score vector $\mathbf{s}$.

[10] The converse is false: if Π is the random permutation that puts mass $1/2$ on $(1, \dots, k)$ and $(k, \dots, 1)$ each, then $\mathrm{E}(\Pi) = \frac{k+1}{2}\,\mathbf{e}$ so $\langle \mathrm{E}(\Pi), \mathbf{s} \rangle = 0$ but Π does not satisfy H_1.

Whereas $I_{\mathcal{P}(\mathcal{B}(\mathbf{s}))}(\Pi) > 0$ if and only if $\Pi \in \mathcal{P}(\mathcal{B}(\mathbf{s}))$, $\langle \Pi, \mathbf{s} \rangle > 0$ not only if $\Pi \in \mathcal{P}(\mathcal{B}(\mathbf{s}))$ (by (33)) but also if Π is near $\mathcal{P}(\mathcal{B}(\mathbf{s}))$.[11] Therefore tests that reject H_1 for large values of $\langle \Pi, \mathbf{s} \rangle$ will be sensitive to the enlarged alternative $A'_{1,\mathbf{s}}$, unlike the test based on $\hat{\tau}_{\mathcal{B}}$ in (10) that detects bias only if $I_{\mathcal{P}(\mathcal{B}(\mathbf{s}))}(\Pi) = 1$. Two such tests are now presented.

Assume again that the orderings $\Pi^{(1)}, \ldots, \Pi^{(N)}$ for agents $1, \ldots, N$ are independent but not necessarily identically distributed. Based on the linear concordance scores $\langle \Pi^{(n)}, \mathbf{s} \rangle$, $n = 1, \ldots, N$, suppose that we now wish to test the combined null hypothesis $\breve{H}_1$ that H_1 for all N agents, i.e.,

$$\breve{H}_1: \ \Pi^{(n)} \text{ is second} - \text{order exchangeable}, \quad n = 1, \ldots, N, \tag{35}$$

against the alternative $\breve{A}'_{1,\mathbf{s}}$ that $A'_{1,\mathbf{s}}$ holds for at least one agent, which can be expressed in the form

$$\breve{A}'_{1,\mathbf{s}} : \mathrm{E}(\langle \Pi^{(n)}, \mathbf{s} \rangle) \geq 0, \qquad n = 1, \ldots, N, \tag{36}$$

with at least one inequality strict. Here $\breve{A}'_{1,\mathbf{s}}$ is a standard multivariate one-sided alternative (cf. Silvapulle and Sen (2005), p.100), for which several combined test statistics are available.

The first and simplest is the *average linear concordance (LC) score*

$$\bar{T}_{\mathbf{s}} \equiv \frac{1}{N} \sum\nolimits_{n=1}^{N} \langle \, \Pi^{(n)}, \, \mathbf{s} \, \rangle. \tag{37}$$

Under $\breve{H}_1$, $\Pi^{(1)}, \ldots, \Pi^{(N)}$ are independent and second-order exchangeable, so $\langle \, \Pi^{(n)}, \, \mathbf{s} \, \rangle$ has mean 0 (by (32)) and variance obtained from (31) as follows:

$$\mathrm{Var}(\langle \Pi^{(n)}, \mathbf{s} \rangle) = \mathbf{s} \, \mathrm{Cov}(\Pi^{(n)}) \, \mathbf{s}' = \tfrac{k+1}{12} \mathbf{s}(kI - \mathbf{e}'\mathbf{e})\mathbf{s}' = \tfrac{k(k+1)}{12} \|\mathbf{s}\|^2 \tag{38}$$

since $\mathbf{s}\,\mathbf{e}' \equiv \langle \mathbf{s}, \mathbf{e} \rangle = \sum s_i = 0$. Furthermore, for fixed k and $\mathbf{s}$,

$$|\langle \Pi^{(n)}, \mathbf{s} \rangle| \leq \sum\nolimits_{i=1}^{k} \Pi_i^{(n)} |s_i| \leq k \sum |s_i|, \qquad n = 1, \ldots, N.$$

Thus, by the Central Limit Theorem for independent, uniformly bounded random variables (cf. Feller (1966), p.264(e)), if N is moderately large then from (37) and (38),

$$\sqrt{N} \, \bar{T}_{\mathbf{s}} \overset{d}{\approx} \mathcal{N} \left(0, \, \frac{k(k+1)\|\mathbf{s}\|^2}{12} \right).$$

[11] For example, if $k = 3$ and $\mathbf{s} = (-1, 0, 1)$ then $\mathcal{P}(\mathcal{B}(\mathbf{s})) = \{(1, 2, 3)\}$ but $\langle \Pi, \mathbf{s} \rangle > 0$ for $\Pi = (1, 2, 3)$, $(2, 1, 3)$, and $(1, 3, 2)$.

Therefore $\breve{H}_1$ is rejected in favor of the one-sided alternative $\breve{A}'_{1,\mathbf{s}}$ at approximate significance level α if

$$\bar{T}_{\mathbf{s}} \geq \sqrt{\frac{k(k+1)\|\mathbf{s}\|^2}{12N}}\, z_\alpha, \tag{39}$$

where z_α is the upper α quantile of the standard normal distribution $\mathcal{N}(0,1)$. The approximate p-value is

$$1 - \Phi\left(\bar{T}_{\mathbf{s};\mathrm{obs}} \sqrt{\frac{12N}{k(k+1)\|\mathbf{s}\|^2}}\right), \tag{40}$$

where $\bar{T}_{\mathbf{s};\mathrm{obs}}$ is the observed value of $\bar{T}_{\mathbf{s}}$ and Φ is the cdf of $\mathcal{N}(0,1)$.

Another combined test statistic is the *maximum standardized LC score*

$$T_{\mathbf{s},\max} \equiv \max_{n=1,\dots,N} \langle\, \Pi^{(n)},\, \mathbf{s}\,\rangle. \tag{41}$$

Under $\breve{H}_1$, it follows from (38) and the one-sided Chebyshev Inequality (cf. Feller (1966), eqn.(7.5), p.152) that

$$\Pr\left[\langle \Pi^{(n)},\, \mathbf{s}\rangle \geq \|\mathbf{s}\|\, t\right] \leq \frac{k(k+1)\|\mathbf{s}\|^2/12}{\|\mathbf{s}\|^2 t^2 + k(k+1)\|\mathbf{s}\|^2/12} = \frac{k(k+1)}{12t^2 + k(k+1)}$$

for $t > 0$. Thus by independence,

$$\begin{aligned}
\Pr\left[T_{\mathbf{s},\max} \geq \|\mathbf{s}\| t\right] &= 1 - \Pr\left[T_{\mathbf{s},\max} < \|\mathbf{s}\| t\right] \\
&= 1 - \prod_{n=1}^{N} \Pr\left[\langle \Pi^{(n)},\, \mathbf{s}\rangle < \|\mathbf{s}\|\, t\right] \\
&= 1 - \prod_{n=1}^{N} \left\{1 - \Pr\left[\langle \Pi^{(n)},\, \mathbf{s}\rangle \geq \|\mathbf{s}\|\, t\right]\right\} \\
&\leq 1 - \left[\frac{12t^2}{12t^2 + k(k+1)}\right]^N.
\end{aligned}$$

Therefore an upper bound for the p-value of the $T_{\mathbf{s},\max}$-test is given by

$$1 - \left[\frac{12T^2_{\mathbf{s},\max;\mathrm{obs}}}{12T^2_{\mathbf{s},\max;\mathrm{obs}} + k(k+1)\|\mathbf{s}\|^2}\right]^N. \tag{42}$$

However, this bound is of practical use only if N is small.

The $T_{\mathbf{s},\max}$-test will have higher (lower) sensitivity than the $\bar{T}_{\mathbf{s}}$-test when only a few (most) of the mean concordance scores $\langle \mathrm{E}(\Pi^{(n)}),\, \mathbf{s}\rangle$ are positive, that is, when only a few (most) of the N agents are biased.

Remark 4.1. Thus far it has been assumed that $s_1, \ldots, s_k$ are predetermined quantitative PC scores. Suppose instead that, as in §3, only a qualitative ranking of the items $1, \ldots, k$ under PC is assumed, thereby determining a partitioning $\mathcal{B} \equiv (B_1, \ldots, B_r)$ of $\{1, \ldots, k\}$. As before we let $k_j = |B_j|$, $j = 1, \ldots, r$. A score vector $\mathbf{s} \equiv (s_1, \ldots, s_k)$ conforms to this partitioning if and only if the ordered scores $\{s_{(i)} \,|\, i = 1, \ldots, k\}$ satisfy

$$s_{(i)} = \tilde{s}_j \quad \text{for} \quad K_j + 1 \le i \le K_{j+1},\ 1 \le j \le r, \tag{43}$$

where $K_j = k_1 + \cdots + k_{j-1}$ for $2 \le j \le r$, $K_1 \equiv 0$, and $\tilde{s}_1 < \cdots < \tilde{s}_r$ are the scores within the blocks $(B_1, \ldots, B_r)$. The assumed condition $\bar{s} = 0$ is equivalent to $\sum_{j=1}^{r} \tilde{s}_j k_j = 0$. In Appendix A we determine $\tilde{s}_1^* < \cdots < \tilde{s}_r^*$ such that the associated conforming score vector $\mathbf{s}^*$ maximizes the value of the ratio $\langle \Pi^{(n)},\, \mathbf{s}\rangle / \|\mathbf{s}\|$ over all conforming $\mathbf{s}$ when $\Pi^{(n)} \in \mathcal{P}(\mathcal{B})$, thereby maximizing the contribution of such $\Pi^{(n)}$ to the values of $\bar{T}_{\bar{s}}$ and $T_{\mathbf{s},\max}$. The maximizing scores $\tilde{s}_j^*$ are just the ranks averaged over ties, then centered to sum to 0. (These $\tilde{s}_j^*$ also maximize the bi-directional statistic $\tilde{T}_{\mathbf{s}}$ in (48).) □

4.2. A bi-directional alternative based on scores. As in §4.1, suppose that $\mathbf{s} \equiv (s_1, \ldots, s_k)$ is a score vector for a preference criterion PC, now assumed to be bi-directional as in §3.2. Again we can take $\bar{s} = 0$ without altering the linear concordance ordering. Now large positive values of $\langle \Pi,\, \mathbf{s}\rangle$ indicate concordance of the ordering Π with the PC score vector $\mathbf{s}$ while large negative values indicate concordance with $-\mathbf{s}$. Thus large positive values of the test statistic $\bar{T}_{\mathbf{s}}$ (cf. (37)) lead to rejection of $\breve{H}_1$ in favor of the one-sided alternative $\breve{A}'_{1,\mathbf{s}}$, while large negative values of $\bar{T}_{\mathbf{s}} \equiv -\bar{T}_{-\mathbf{s}}$ lead to rejection of $\breve{H}_1$ in favor of the opposite one-sided alternative $\breve{A}'_{1,-\mathbf{s}}$.

Suppose, however, that it is wished to test unbiasedness in order assignment against the bi-directional ($\equiv$ two-sided) alternative

$$A'_{2,\mathbf{s}} : \mathrm{E}(\langle \Pi, \mathbf{s}\rangle) \neq 0 \tag{44}$$

(compare to $A'_{1,\mathbf{s}}$ in (34)). The orderings $\Pi^{(1)}, \ldots, \Pi^{(N)}$ again are assumed to be independent but not necessarily identically distributed. We shall test the combined null hypothesis $\breve{H}_2$ that H_2 holds for all N agents (recall (3)), i.e.,

$$\breve{H}_2 : \Pi^{(n)} \text{ is fourth} - \text{order exchangeable}, \quad n = 1, \ldots, N, \tag{45}$$

against the multivariate two-sided alternative

$$\breve{A}'_{2,\mathbf{s}} : \big(\mathrm{E}(\langle \Pi^{(1)}, \mathbf{s}\rangle), \ldots, \mathrm{E}(\langle \Pi^{(N)}, \mathbf{s}\rangle)\big) \neq (0, \ldots, 0) \tag{46}$$

(compare to $\breve{A}'_{1,\mathbf{s}}$ in (36)), using the LC scores $\langle \Pi^{(n)}, \mathbf{s}\rangle$.

Because $\breve{A}'_{2,\mathbf{s}}$ can be expressed equivalently as

$$\breve{A}'_{2,\mathbf{s}} : \sum_{n=1}^{N} \big[\mathrm{E}(\langle \Pi^{(n)}, \mathbf{s}\rangle)\big]^2 > 0, \tag{47}$$

an appropriate test rejects $\breve{H}_2$ in favor of $\breve{A}'_{2,\mathbf{s}}$ for large values of

$$\tilde{T}_{\mathbf{s}} \equiv \frac{1}{N} \sum\nolimits_{n=1}^{N} [\langle\, \Pi^{(n)}, \, \mathbf{s}\,\rangle]^2. \tag{48}$$

Under $\breve{H}_2$, $\Pi^{(1)}, \ldots, \Pi^{(N)}$ are independent and fourth-order exchangeable, $[\langle\, \Pi^{(n)}, \, \mathbf{s}\,\rangle]^2$ has mean

$$\mathrm{E}\big([\langle\, \Pi^{(n)}, \, \mathbf{s}\,\rangle]^2\big) = \mathrm{Var}(\langle \Pi^{(n)}, \, \mathbf{s}\rangle) = \tfrac{k(k+1)}{12} \|\mathbf{s}\|^2 \equiv \Gamma(\mathbf{s}) \tag{49}$$

by (38), and variance obtained from (65) in Appendix B:

$$\mathrm{Var}(\langle \Pi^{(n)}, \, \mathbf{s}\rangle^2) = \tfrac{k(k+1)}{360} \Big[(5k^2 - k - 9)\sigma_2^2 - 3k(k+1)\sigma_4\Big] \equiv \Psi(\mathbf{s}), \tag{50}$$

where $\sigma_2 = \sum s_i^2 \equiv \|\mathbf{s}\|^2$ and $\sigma_4 = \sum s_i^4$. For fixed k and $\mathbf{s}$,

$$[\langle \Pi^{(n)}, \, \mathbf{s}\rangle]^2 \leq \|\Pi^{(m)}\|^2 \, \|s\|^2 \leq \tfrac{k(k+1)(2k+1)}{6} \|\mathbf{s}\|^2, \quad n = 1, \ldots, N,$$

so the Central Limit Theorem for independent, uniformly bounded random variables implies that if N is moderately large, then

$$\sqrt{N} \left(\tilde{T}_{\mathbf{s}} - \Gamma(\mathbf{s})\right) \overset{d}{\approx} \mathcal{N}\left(0, \, \Psi(\mathbf{s})\right). \tag{51}$$

Thus $\breve{H}_2$ is rejected in favor of $\breve{A}'_{2,\mathbf{s}}$ at approximate significance level α if

$$\tilde{T}_{\mathbf{s}} \geq \Gamma(\mathbf{s}) + \sqrt{\frac{\Psi(\mathbf{s})}{N}} \, z_\alpha, \tag{52}$$

where z_α is the upper α quantile of $\mathcal{N}(0,1)$. The approximate p-value is

$$1 - \Phi\left[\sqrt{\frac{N}{\Psi(\mathbf{s})}}\left(\tilde{T}_{\mathbf{s};\mathrm{obs}} - \Gamma(\mathbf{s})\right)\right], \tag{53}$$

where $\tilde{T}_{\mathbf{s};\mathrm{obs}}$ is the observed value of $\tilde{T}_{\mathbf{s}}$ and Φ is the cdf of $\mathcal{N}(0,1)$.

5. Results from the 2014 Texas Republican primary election.

Ballot-order data (used in Grant, 2017) are available for $N = 245$ of the 254 Texas counties.[12] Five Republican primary election races are studied: U.S. Senate, State Comptroller, Lieutenant Governor, Railroad Commissioner, and Attorney General. These are discussed individually in §5.1-5.5 below.

Numerical results based on the rank-compatibility (RC) statistics $\hat{\tau}_{\mathcal{B}}$, $\hat{\tau}_{\tilde{\mathcal{B}}}$ (cf. (10)) and linear concordance (LC) statistics $\bar{T}_{\mathbf{s}^*}$, $\tilde{T}_{\mathbf{s}^*}$ (cf. (37), (48)) are summarized in Table 1. Here $\mathbf{s}^*$ is the maximizing score vector determined from (57)). For the RC statistics, an entry n^* (n^{**}) indicates confidence (strong confidence) for the presence of bias in n or more of the N counties recorded, while p^* (p^{**}) indicates a significant (strongly significant) p-value for the presence of bias in n or more of the N counties recorded. For the LC statistics, p^* (p^{**}) indicates a significant (strongly significant) p-value for the presence of bias in one or more of the N counties recorded.

In three of the five races, we find significant statistical evidence for the presence of bias in ballot-order assignments in one or more of the N reporting counties. The evidence is strongly significant in two (Lieutenant Governor and Railroad Commissioner) of these four races.

5.1. U.S. Senate. We take the preference criterion (PC) to be bi-directional political ideology on the Republican establishment-Tea Party spectrum. There were $k = 8$ candidates: in alphabetical order Cleaver $\equiv 1$, Cope $\equiv 2$, Cornyn $\equiv 3$, Mapp $\equiv 4$, Reasor $\equiv 5$, Stockman $\equiv 6$, Stovall $\equiv 7$, Vega $\equiv 8$. John Cornyn was the establishment incumbent so there were no establishment challengers; Stockman, Stovall, and Reasor were Tea Party candidates;

[12]Except for the U.S. Senate and Railroad Commissioner races, where data is available for $N = 243$ and $N = 244$ counties, respectively. Data were missing due to some counties not holding Republican primaries, some counties failing to provide data to the author of Grant (2017), and there being errors on some counties' ballots for certain elections.

Data from the 2014 Texas Republican Primary Election						
Office	US Sen.	Compt.	Lt. Gov.[1]	Lt. Gov.[2]	RR Comm.	At. Gen.
k	8	4	4	4	4	3
$p_{\mathcal{B}}$	1/280	1/6	1/24	1/6	1/4	1/6
$\hat{\tau}_{\mathcal{B}}$	2	45	10	56	82	43
b	1	44	9	55 (50)	81 (72)	42
$\hat{\tau}_{\mathcal{B}} - b \equiv n$	1	1	1	1** (6*)	1** (10*)	1
$\gamma(b; N, p_{\mathcal{B}})$	.784	.739	.429	.992 (.948)	.998 (.953)	.619
p-value	.216	.261	.571	.008**(.052*)	.002**(.047*)	.381
$\hat{\tau}_{\tilde{\mathcal{B}}}$	0	51	4			47
$\tilde{b}$	-	50	3			46
$\hat{\tau}_{\tilde{\mathcal{B}}} - \tilde{b} \equiv n$	-	1*	1			1
$\gamma(\tilde{b}; N, p_{\mathcal{B}})$	-	.948	.008			.835
p-value	-	.052*	.992			.165
$\hat{\tau}_{\mathcal{B}} + \hat{\tau}_{\tilde{\mathcal{B}}}$	2	96	14			90
c	1	94	13			89
$\hat{\tau}_{\mathcal{B}} + \hat{\tau}_{\tilde{\mathcal{B}}} - c \equiv n$	1	2*	1			1
$\gamma(c; N, 2p_{\mathcal{B}})$	.481	.958	.048			.856
p-value	.519	.042*	.952			.144
$\bar{T}_{\mathbf{s}^*}$	1.471	-.016	.196	.384	.451	-.086
z-value	1.582	-.099	1.061	2.326	3.156	-.949
p-value	.057	.539	.144	.010**	.001**	.829
$\bar{T}_{-\mathbf{s}^*}$	-1.471	.016	-.196			.086
z-value	-1.582	.099	-1.061			.949
p-value	.943	.461	.856			.171
$\tilde{T}_{\mathbf{s}^*}$	208.55	7.478	8.278			2.102
z-value	-.088	1.867	-.113			1.129
p-value	.535	.031*	.545			.129

Table 1: Lower confidence bounds and significance levels based on the rank-compatibility (RC) statistic $\hat{\tau}_{\mathcal{B}}$ (and also $\hat{\tau}_{\tilde{\mathcal{B}}}$ for the bi-directional PC's) and significance levels based on the linear concordance (LC) statistic $\bar{T}_{\mathbf{s}^*}$ (and also $\bar{T}_{-\mathbf{s}^*}$ and $\tilde{T}_{\mathbf{s}^*}$ for the bi-directional PC's. Here, Lt.Gov.[1] (Lt.Gov.[2]) refers to the partitionings $\mathcal{B}^1$ and $\tilde{\mathcal{B}}^1$ ($\mathcal{B}^2$ and $\tilde{\mathcal{B}}^2$) - cf. §5.3.

Cleaver, Cope, Mapp, and Vega were minor candidates not meriting classification in either group. In the notation of §3.2, the PC is represented by the partitioning $\mathcal{B}$ and its opposite $\tilde{\mathcal{B}}$:

$$\mathcal{B} \equiv (B_1, B_2, B_3) = \{3\}, \{1, 2, 4, 8\}, \ \{5, 6, 7\},$$
$$\tilde{\mathcal{B}} \equiv (\tilde{B}_1, \tilde{B}_2, \tilde{B}_3) = \{5, 6, 7\}, \ \{1, 2, 4, 8\}, \ \{3\}.$$

The observed numbers of ballot orderings compatible with $\mathcal{B}$ and $\tilde{\mathcal{B}}$ (cf. (4), (10)) were small: $\hat{\tau}_{\mathcal{B}} = 2$ and $\hat{\tau}_{\tilde{\mathcal{B}}} = 0$, respectively, out of 243 reporting counties. Here $p_{\mathcal{B}} = \frac{1!4!3!}{8!} = \frac{1}{280}$ and from (13) with $b = 1$, $\gamma(1; 243, 1/280) = .78$, so the RC statistic $\hat{\tau}_{\mathcal{B}} - 1 \equiv 1$ provides a 78% lower confidence bound for $\tau_{\mathcal{B}}$ (cf. (17)), the number of counties where bias favoring $\mathcal{B}$ is present. In the language of hypothesis testing, this yields a non-significant p-value of 1-.78=.22 for the alternative hypothesis that bias favoring $\mathcal{B}$ is present in one or more counties. Trivially, the p-value for $\tilde{\mathcal{B}}$ is also non-significant.

Also, $\gamma(1; 243, 1/140) = .48$, so from (19) with $c = 1$, $\hat{\tau}_{\mathcal{B}} + \hat{\tau}_{\tilde{\mathcal{B}}} - 1 \equiv 1$ is a 48% lower confidence bound for $\tau_{\mathcal{B}} + \tau_{\mathcal{B}}$, corresponding to a non-significant p-value of 1-.48=.52 for the alternative hypothesis that bias favoring either $\mathcal{B}$ or $\tilde{\mathcal{B}}$ is present in one or more counties.

Next consider the LC statistic $\bar{T}_{\mathbf{s}}$ (37) for testing $\breve{H}_1$ (35) vs. the one-sided alternative $\breve{A}'_{1,\mathbf{s}}$ (36). From (57) the maximizing centered score vector derived from $\mathcal{B}$ is

$$\mathbf{s}^* = (-3.5, -1, -1, -1, -1, 2.5, 2.5, 2.5), \qquad (\|\mathbf{s}^*\|^2 = 35),$$

from which we find $\bar{T}_{\mathbf{s}^*;\text{obs}} = 1.47$, corresponding to a nearly significant p-value (40) of .06 for testing $\breve{H}_1$ (35) vs. the one-sided alternative $\breve{A}'_{1,\mathbf{s}^*}$ (36), evidence for the presence of bias for the PC in one or more of the 243 reporting counties. As expected, the LC statistic is more sensitive than the RC statistic in this case, yielding a p-value (.06) which is notably smaller than the RC statistic's non-significant p-value of .22.

Lastly, if we consider the LC statistic $\tilde{T}_{\mathbf{s}^*}$ (48) for testing $\breve{H}_2$ (45) vs. the two-sided alternative $\breve{A}'_{2,\mathbf{s}^*}$ (46), its observed value is $\tilde{T}_{\mathbf{s}^*;\text{obs}} = 208.55$, corresponding to a non-significant p-value (53) of .54. Throughout the state, this race was seen as a foregone conclusion, which the final vote confirmed; so it is unsurprising to find no evidence of manipulation.

5.2. State Comptroller. Again the PC is bi-directional political ideology on the Republican establishment-Tea Party spectrum. There were $k = 4$

candidates: in alphabetical order Hegar $\equiv 1$, Hilderbran $\equiv 2$, Medina $\equiv 3$, and Torres $\equiv 4$. Hilderbran and Torres were establishment state legislators, Hegar and Medina were explicitly identified with the Tea Party. In the notation of §3.2, this PC is represented by the opposite partitionings

$$\mathcal{B} \equiv (B_1, B_2) = (\{2,4\},\ \{1,3\}),$$
$$\tilde{\mathcal{B}} \equiv (\tilde{B}_1, \tilde{B}_2) = (\{1,3\},\ \{2,4\}).$$

The observed numbers of ballot orderings compatible with $\mathcal{B}$ and $\tilde{\mathcal{B}}$ were much larger than in §5.1: $\hat{\tau}_{\mathcal{B}} = 45$ and $\hat{\tau}_{\tilde{\mathcal{B}}} = 51$, respectively, out of 245 counties. Here $p_{\mathcal{B}} = \frac{2!2!}{4!} = \frac{1}{6}$ and $\gamma(44; 245, 1/6) = .74$, so the RC statistic $\hat{\tau}_{\mathcal{B}} - b = \hat{\tau}_{\mathcal{B}} - 44 = 1$ is a 74% lower confidence bound for $\tau_{\mathcal{B}}$, the number of counties for which bias favoring $\mathcal{B}$ is present. This yields a p-value of 1-.74=.26 for the alternative hypothesis that bias favoring $\mathcal{B}$ is present in one or more of the 245 reporting counties, which is not statistically significant. However, $\gamma(50; 245, 1/6) = .95$, so with $\tilde{b} = 50$, $\hat{\tau}_{\tilde{\mathcal{B}}} - 50 = 1$ is a 95% lower confidence bound for $\tau_{\tilde{\mathcal{B}}}$, corresponding to a p-value of 1-.95=.05, which lends support to the opposite alternative hypothesis that bias favoring $\tilde{\mathcal{B}}$ is present in one or more of these 245 counties.

Stronger evidence for bias emerges when testing the bi-directional alternative. For this test, $\gamma(94; 245, 1/3) = .96$, so from (19) with $c = 94$, $\hat{\tau}_{\mathcal{B}} + \hat{\tau}_{\tilde{\mathcal{B}}} - 94 \equiv 2$ is a 96% lower confidence bound for $\tau_{\mathcal{B}} + \tau_{\mathcal{B}}$. A significant p-value .04 is obtained for the alternative hypothesis that bias favoring either $\mathcal{B}$ or $\tilde{\mathcal{B}}$ is present in two or more of the 245 reporting counties.

Next consider the LC statistic $\bar{T}_{\mathbf{s}}$ for testing $\breve{H}_1$ vs. the one-sided alternative $\breve{A}'_{1,\mathbf{s}}$. From (57) the maximizing centered score vector for $\mathcal{B}$ is

$$\mathbf{s}^* = (-1, -1, 1, 1), \qquad (\|\mathbf{s}^*\|^2 = 4), \tag{54}$$

from which we find $\bar{T}_{\mathbf{s}^*;\mathrm{obs}} = -.02$, corresponding to a non-significant p-value of .54 for testing $\breve{H}_1$ vs. the one-sided alternative $\breve{A}'_{1,\mathbf{s}^*}$, and a non-significant p-value of 1-.54=.46 for testing $\breve{H}_1$ vs. $\breve{A}'_{1,-\mathbf{s}^*}$. Note that the LC statistic is less sensitive than the RC statistic in these two cases. This is explained by the fact that when the PC is bi-directional and biases in both directions are present, their effects cancel in $\bar{T}_{\mathbf{s}^*}$ and $\bar{T}_{\mathbf{s}^*}$.

However, as for the RC tests, the bi-directional alternative yields clarity. The observed value of the LC statistic $\tilde{T}_{\mathbf{s}^*}$ for testing $\breve{H}_2$ vs. the two-sided alternative $\breve{A}'_{2,\mathbf{s}^*}$ is $\tilde{T}_{\mathbf{s}^*;\text{obs}} = 7.48$, corresponding to a significant p-value (53) of .03. This provides statistical evidence of the presence of bias in ballot-order assignment, favoring either the establishment or Tea Party candidates, in one or more of the 245 counties. The LC statistic $\tilde{T}_{\mathbf{s}^*}$ is more sensitive than the RC statistic $\hat{\tau}_{\mathcal{B}} + \hat{\tau}_{\tilde{\mathcal{B}}}$ (p-value .04) in this case.

5.3. Lieutenant Governor. There were $k = 4$ candidates: Dewhurst $\equiv 1$, Patrick $\equiv 2$, Patterson $\equiv 3$, Staples $\equiv 4$. Dewhurst, the establishment candidate, was the incumbent, hence very well known. Patrick, the Tea Party challenger, was also high-profile. Patterson and Staples were between the far-right Patrick and the center-right Dewhurst ideologically, with Staples to the right of Patterson. We consider two preference criteria: PC^1 = political ideology on the establishment-Tea Party spectrum (bi-directional) and PC^2 = public prominence (uni-directional). The associated partitionings are

$$\mathcal{B}^1 \equiv (B_1^1, B_2^1, B_3^1, B_4^1) = (\{1\},\ \{3\},\ \{4\},\ \{2\}),$$
$$\tilde{\mathcal{B}}^1 \equiv (\tilde{B}_1^1, \tilde{B}_2^1, \tilde{B}_3^1, \tilde{B}_4^1) = (\{2\},\ \{4\},\ \{3\},\ \{1\});$$
$$\mathcal{B}^2 \equiv (B_1^2, B_2^2) = (\{1,2\},\ \{3,4\}).$$

From Table 1, no significant evidence of bias for or against PC^1 in ballot-order assignment is present. However, both the RC and LC statistics show highly significant evidence of bias for PC^2, as follows:

First, $\gamma(55; 245, 1/6) = .99$, so with $b = 55$, $\hat{\tau}_{\mathcal{B}} - 55 = 1$ is a 99% lower confidence bound for $\tau_{\mathcal{B}}$, corresponding to a highly significant p-value of 1-.99=.01, which strongly supports the alternative hypothesis that bias for $\mathcal{B}^2$ is present in one or more of the 245 reporting counties. Furthermore, $\gamma(50; 245, 1/6) = .95$, so with $b = 50$, $\hat{\tau}_{\mathcal{B}} - 50 = 6$ is a 95% lower confidence bound for $\tau_{\mathcal{B}}$. A significant p-value of .05 is obtained for the alternative hypothesis that bias for $\mathcal{B}^2$ is present in six or more of these 245 counties.

Next consider the LC statistic $\bar{T}_{\mathbf{s}}$. From (57) the maximizing centered score vector $\mathbf{s}^*$ derived from $\mathcal{B}^2$ again is given in (54), from which we find $\bar{T}_{\mathbf{s}^*;\text{obs}} = .38$, corresponding to a strongly significant p-value of .01 for testing $\breve{H}_1$ vs. the one-sided alternative $\breve{A}'_{1,\mathbf{s}^*}$. In this case the LC test's p-value is only slightly higher than that of the RC test.

Note that the analyses of the three races thus far considered support the viability of all three general hypotheses considered: the null (of uniformity), uni-directional bias, and bi-directional bias. Furthermore, on the whole the

LC test tends to yield lower p-values. The apparently greater sensitivity of the LC test is seen in the next race discussed as well.

5.4. Railroad Commissioner. There were $k = 4$ candidates: Berger $\equiv 1$, Boyuls $\equiv 2$, Christian $\equiv 3$, Sitton $\equiv 4$. Christian was widely shunned, so we propose a uni-directional PC specified by the partitioning

$$\mathcal{B} \equiv (B_1, B_2) = \{1, 2, 4\}, \{3\}.$$

Both the RC and LC statistics show highly significant evidence of bias for $\mathcal{B}$ in ballot-order assignments, as follows:

Because $\gamma(81; 244, 1/4) = .998$, $\hat{\tau}_{\mathcal{B}} - 81 = 1$ is a 99.8% lower confidence bound for $\tau_{\mathcal{B}}$, corresponding to a highly significant p-value of 1-.998=.002, which strongly supports the alternative that bias favoring $\mathcal{B}$ is present in one or more of the 244 reporting counties. Furthermore, $\gamma(72; 244, 1/4) = .95$, so with $b = 72$, $\hat{\tau}_{\mathcal{B}} - 72 = 10$ is a 95% lower confidence bound for $\tau_{\mathcal{B}}$. A significant p-value .05 is obtained for the alternative hypothesis that bias favoring $\mathcal{B}$ is present in ten or more of these 244 counties.

Next consider the LC statistic $\bar{T}_{\mathbf{s}}$. From (57) the maximizing centered score vector derived from $\mathcal{B}$ is

$$\mathbf{s}^* = (-.5, -.5, -.5, 1.5), \qquad (\|\mathbf{s}^*\|^2 = 3),$$

from which we obtain $\bar{T}_{\mathbf{s}^*;\text{obs}} = .45$, corresponding to a highly significant p-value of .001 for testing $\breve{H}_1$ vs. the one-sided alternative $\breve{A}'_{1,\mathbf{s}^*}$.

5.5. Attorney General. There were $k = 3$ candidates: Branch $\equiv 1$, Paxton $\equiv 2$, Smitherman $\equiv 3$. Branch was the establishment candidate, Paxton was the Tea Party candidate, and Smitherman was a non-factor. Again the PC is taken to be the establishment-Tea Party spectrum, which is bi-directional and specified by the opposite partitionings

$$\mathcal{B} \equiv (B_1, B_2, B_3) = (\{1\},\ \{3\},\ \{2\}),$$
$$\tilde{\mathcal{B}} \equiv (\tilde{B}_1, \tilde{B}_2, \tilde{B}_3) = (\{2\},\ \{3\},\ \{1\}).$$

Although four of the six p-values in the final column of Table 1 are somewhat small (.17, .14, .17, .13), we do not find conclusive evidence for bias in ballot-order assignments for this race.

6. Concluding remarks.

The ballot-order effect is an example of a more general psychological phenomenon, the *primacy effect* (cf. Murdock, 1962), in which the first-listed of a set of options tends to be chosen more frequently. Other in which the primacy effect may affect outcomes include funding approval processes, such as those at the NIH and NSF, and college admissions decisions. In such scenarios it is often possible to identify various preference criteria (e.g, social, religious, or economic, as well as political) that might disfavor certain groups, so any preliminary order of appearance for the items under consideration (applicants, candidates, etc.) should be determined without prejudice. The procedures in this paper for testing uniform randomness vs. specific alternatives are directly applicable in such cases.

The application of our methods to ballot orders from the 2014 Texas Republican Primaries suggested that the LC method may have somewhat more power to detect departures from uniformity than the RC method. Additionally, in Appendix C, we compare the statistical power of these tests to the rank test used by Ho and Imai (2008) to test for departures from uniformity over permutations of 26 items, a test that does not utilize *a priori* information. We find that our methods have superior power, as anticipated.

Lastly, our treatment assumes throughout that the preference criterion (PC) is uni-dimensional and that agents' preferences are extreme, i.e. that they most strongly prefer items at the far ends of the PC spectrum. If instead the PC is multi-dimensional (e.g., Treier and Hillygus (2009)), or if there exist biased agents with moderate rather than extreme preferences, further bias-detection procedures will be needed. Both these topics are under investigation.

Appendix A: The maximizing linear scores $\tilde{s}_j^*$.

As in Remark 4.1, consider score vectors $\mathbf{s}$ that conform to a specified partitioning $\mathcal{B}$ of the items $1, \dots, k$, so $\mathbf{s}$ satisfies (43) with $\sum_{j=1}^{r} \tilde{s}_j k_j = 0$. Thus by (4), if $\pi \in \mathcal{P}(\mathcal{B})$ then

$$\begin{aligned}\langle \pi, \mathbf{s} \rangle &= \sum\nolimits_{j=1}^{r} \tilde{s}_j \sum\nolimits_{i=1}^{k_j} \pi_{K_j+i} \\ &= \sum\nolimits_{j=1}^{r} \tilde{s}_j \sum\nolimits_{i=1}^{k_j} (K_j + i) \\ &= \sum\nolimits_{j=1}^{r} \tilde{s}_j \left(K_j + \frac{k_j+1}{2} \right) k_j.\end{aligned}$$

Therefore, if $\pi \in \mathcal{P}(\mathcal{B})$ then

$$\begin{aligned}\frac{\langle \pi, \mathbf{s} \rangle}{\|\mathbf{s}\|} &= \left[\sum \tilde{s}_j \left(K_j + \frac{k_j+1}{2} \right) k_j \right] \left(\sum \tilde{s}_j^2 k_j \right)^{-1/2} \\ &\leq \left[\sum \left(K_j + \frac{k_j+1}{2} \right)^2 k_j \right]^{1/2} \qquad (55)\end{aligned}$$

by the Cauchy-Schwartz inequality. Equality holds in (55) if and only if

$$\tilde{s}_j = a \left(K_j + \frac{k_j+1}{2} \right) + b, \quad j = 1, \ldots, r, \qquad (56)$$

for some scalars a, b; without loss of generality we can take $a = 1$. In this case the condition $\sum \tilde{s}_j k_j = 0$ requires that

$$\tilde{s}_j = (K_j - \tilde{K}) + \frac{1}{2}(k_j - \tilde{k}) \equiv \tilde{s}_j^*, \qquad (57)$$

where

$$\tilde{K} = \frac{1}{k} \sum K_j k_j, \quad \tilde{k} = \frac{1}{k} \sum k_j^2. \qquad (58)$$

As stated in Remark 4.1, these maximizing scores $\tilde{s}_j^*$ are just the ranks averaged over ties, then centered to sum to 0.

Appendix B: Variance of $[\langle\Pi, \mathbf{s}\rangle]^2$.

Assume that the random permutation Π satisfies H_2, that is, fourth-order exchangeability, and the scores are centered ($\sum s_i = 0$). It is convenient also to center Π at 0, so define

$$\Omega \equiv (\Omega_1, \ldots, \Omega_k) = (\Pi_1 - \tfrac{k+1}{2}, \ldots, \Pi_k - \tfrac{k+1}{2}).$$

Then Ω is also fourth-order exchangeable, $\sum \Omega_i = 0$, $\sum \Pi_i s_i = \sum \Omega_i s_i$, and

$$\omega_2 \equiv \sum_i \Omega_i^2 = \sum_i (i - \tfrac{k+1}{2})^2 = \frac{k(k^2-1)}{12}, \tag{59}$$

$$\omega_4 \equiv \sum_i \Omega_i^4 = \sum_i (i - \tfrac{k+1}{2})^4 = \frac{k(k^2-1)(3k^2-7)}{240}. \tag{60}$$

Also define $\sigma_2 = \sum s_i^2 \equiv \|\mathbf{s}\|^2$ and $\sigma_4 = \sum s_i^4$.

The fourth moment

$$\mathrm{E}\big([\langle\Pi, \mathbf{s}\rangle]^4\big) \equiv \mathrm{E}\Big(\Big[\sum \Pi_i s_i\Big]^4\Big) = \mathrm{E}\Big(\Big[\sum \Omega_i s_i\Big]^4\Big)$$

can be deduced from the fourth moment of the generalized rank correlation coefficient

$$r_{\mathbf{s}} \equiv \frac{\sum \Omega_i s_i}{\sqrt{\omega_2}\sqrt{\sigma_2}}, \tag{61}$$

as developed by M. G. Kendall and his colleagues. In our notation, eqn. (26.45) in Stuart and Ord (1987) becomes

$$\mathrm{E}\big(r_{\mathbf{s}}^4\big) = \frac{3}{(k^2-1)}\left[1 + \frac{(k-2)(k-3)}{3k(k-1)^2}\left(\frac{k_4}{k_2^2}\right)\left(\frac{k_4'}{(k_2')^2}\right)\right], \tag{62}$$

where (cf. Stuart and Ord (1987), eqns. (12.29) and (10.3)),

$$k_2 \equiv k_2(\Omega) = \frac{\omega_2}{k-1},$$
$$k_4 \equiv k_4(\Omega) = \frac{k^2}{(k-1)(k-2)(k-2)}\left[\frac{k+1}{k}\omega_4 - 3\left(\frac{k-1}{k^2}\right)\omega_2^2\right],$$

and $k_2' \equiv k_2(\mathbf{s})$ and $k_4' \equiv k_4(\mathbf{s})$ are defined similarly with (ω_2, ω_4) replaced by (σ_2, σ_4). From (59), (60), and some algebra, we find that

$$\frac{k_4}{k_2^2} = -\frac{6}{5},$$
$$\frac{k_4'}{(k_2')^2} = \frac{k-1}{(k-2)(k-3)}\left[k(k+1)\frac{\sigma_4}{\sigma_2^2} - 3(k-1)\right].$$

From this, (62), and more algebra we obtain

$$\text{(63)} \qquad \mathrm{E}\left(r_{\mathbf{s}}^4\right) = \frac{3[(k-1)(5k+6)\sigma_2^2 - 2k(k+1)\sigma_4]}{5k(k-1)(k^2-1)\sigma_2^2}.$$

Thus from (61) and (59),

$$\text{(64)} \qquad \mathrm{E}\left(\left[\sum \Omega_i s_i\right]^4\right) = \frac{k(k+1)[(k-1)(5k+6)\sigma_2^2 - 2k(k+1)\sigma_4]}{240}.$$

Finally, because $\mathrm{E}\left(\sum \Omega_i s_i\right) = 0$, and $\|\mathbf{s}\|^2 = \sigma_2$, it follows from (38) that

$$\mathrm{E}\left(\left[\sum \Omega_i s_i\right]^2\right) = \mathrm{Var}\left(\sum \Omega_i s_i\right) = \frac{k(k+1)}{12}\sigma_2,$$

so

$$\begin{aligned}
\mathrm{Var}\left(\left[\sum \Pi_i s_i\right]^2\right) &= \mathrm{Var}\left(\left[\sum \Omega_i s_i\right]^2\right) \\
&= \mathrm{E}\left(\left[\sum \Omega_i s_i\right]^4\right) - \left[\mathrm{E}\left(\left[\sum \Omega_i s_i\right]^2\right)\right]^2 \\
\text{(65)} \qquad &= \frac{k(k+1)}{360}\left[(5k^2-k-9)\sigma_2^2 - 3k(k+1)\sigma_4\right].
\end{aligned}$$

Note: To check the accuracy of (64), consider the special case of centered linear ranks, i.e., $s_i = i - \frac{k+1}{2}$, where $\sigma_2 = \omega_2$ and $\sigma_4 = \omega_4$. Then verify that (64) reduces to the well-known relation (cf. Kendall (1962), eqn. (5.27))

$$\text{(66)} \qquad \mathrm{E}\left(\left[\langle \Pi, \mathbf{s}\rangle\right]^4\right) = \frac{k^3(k+1)^3(k-1)(25k^3 - 38k^2 - 35k + 72)}{172800}.$$

Appendix C: Statistical Power.

In this appendix, we perform a simulation study to compare the statistical power of our rank-compatibility and linear concordance tests to the rank test used by Ho and Imai (2008).[13] This rank test was chosen as the standard of comparison because it is approximately asymptotically valid at a given α level (the null distribution must be computed using Monte Carlo) and because it could plausibly demonstrate some power at smaller sample sizes, though we note that since it does not incorporate *a priori* information we expect it to be less sensitive to departures from uniformity than our methods. We present simulation results only for the uni-directional bias setting, since these results make clear the superiority of our methods and we are aware of no tests other than the RC and LC tests developed in sections 3 and 4 that are targeted to the small-sample, bi-directional bias setting.

This study uses data generated under the model in (5), i.e. the model in which agents either generate an ordering uniformly at random (are unbiased) or generate an ordering uniformly at random from the set of orderings compatible with a given partitioning $\mathcal{B}$ (are biased). Clearly, this model is unlikely to hold exactly in practice, and future work should examine the power of these methods under more complex, realistic departures from full uniformity.

The simulation parameters were motivated by the ballot order context, in which the number of counties in a state (the number of agents) is at most 254, a ballot typically has from 2 to 8 candidates per race, and we hypothesize in general that while prejudicial bias may exist in some counties, most agents will follow the law and randomize the ballot appropriately. Accordingly, we perform simulations for three sample sizes–$N = 50$, 100, and 500–which are modest relative to the selected k values of 3, 6, and 8 (if the ratio N/k is very large, rank tests or chi-square goodness of fit tests are suitable and the additional sensitivity of our methods is unhelpful; if it is too small, any sort of testing is hopelessly underpowered). We consider three values for the proportion of truly biased agents ρ: 1%, 5%, and 20%. We consider various plausible configurations of the PC, designed using the principle that in practice there are likely to be a small number of items of interest that we expect to generate preference, that these items can be grouped into one or two

[13] If R_{ij} is the rank of item i in the jth ordering, the test statistic is $\frac{1}{k(k-1)} \sum_{i \neq i'} |\frac{1}{n} \sum_{j=1}^{n} (R_{ij} - R_{i'j})|$.

blocks, and that the remaining items are more or less indistinguishable and can be grouped into a single large block. Size $\alpha = 0.05$ tests are performed throughout. For a given percentage of biased agents ρ, when the equation $\tau/N = \rho$ cannot be solved for an integral τ, we randomly select τ in each replication to be either $\lfloor \rho N \rfloor$ or $\lceil \rho N \rceil$ such that the expected value of τ is ρN.

	$\rho = 0.01$			$\rho = 0.05$			$\rho = 0.2$		
$\mathcal{B}$	N=50	100	500	N=50	100	500	N=50	100	500
$\{1\}, \{2,3\}$	0.05	0.05	0.09	0.12	0.16	0.46	0.58	0.89	1.00
$\{1\}, \{2,\ldots,6\}$	0.04	0.05	0.13	0.12	0.25	0.80	0.91	1.00	1.00
$\{1\}, \{2,3\}, \{4,5,6\}$	0.03	0.08	0.41	0.36	1.00	1.00	1.00	1.00	1.00
$\{1,2\}, \{3,\ldots,8\}$	0.06	0.06	0.23	0.37	0.68	1.00	1.00	1.00	1.00
$\{1,2\}, \{3,4\}, \{5,\ldots,8\}$	0.07	0.22	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: Power for the rank-compatibility (RC) test, uni-directional preferences

Note that the probability of an ordering concordant with the fifth partitioning $\{1,2\}, \{3,4\}, \{5,\ldots,8\}$ appearing under the null hypothesis of uniformity is extremely small: $\frac{2!2!4!}{8!} = .002$. This causes a sharp increase in power for the RC test as N increases, because the binomial density under such a small probability is sharply decreasing in N when N is such that the density is neither close to its maximum nor close to zero.

	$\rho = 0.01$			$\rho = 0.05$			$\rho = 0.2$		
$\mathcal{B}$	N=50	100	500	N=50	100	500	N=50	100	500
$\{1\}, \{2,3\}$	0.06	0.06	0.07	0.10	0.14	0.39	0.57	0.80	1.00
$\{1\}, \{2,\ldots,6\}$	0.05	0.07	0.10	0.11	0.16	0.49	0.71	0.93	1.00
$\{1\}, \{2,3\}, \{4,5,6\}$	0.07	0.07	0.11	0.17	0.26	0.77	0.92	0.99	1.00
$\{1,2\}, \{3,\ldots,8\}$	0.06	0.07	0.12	0.20	0.23	0.74	0.90	0.99	1.00
$\{1,2\}, \{3,4\}, \{5,\ldots,8\}$	0.08	0.08	0.13	0.23	0.35	0.87	0.98	1.00	1.00

Table 3: Power for the linear concordance (LC) test, uni-directional preferences

	$\rho = 0.01$			$\rho = 0.05$			$\rho = 0.2$		
$\mathcal{B}$	N=50	100	500	N=50	100	500	N=50	100	500
$\{1\}, \{2, 3\}$	0.03	0.06	0.05	0.05	0.06	0.12	0.28	0.50	1.00
$\{1\}, \{2, \dots, 6\}$	0.05	0.04	0.04	0.06	0.06	0.15	0.14	0.47	1.00
$\{1\}, \{2, 3\}, \{4, 5, 6\}$	0.07	0.05	0.04	0.05	0.06	0.27	0.38	0.88	1.00
$\{1, 2\}, \{3, \dots, 8\}$	0.03	0.03	0.05	0.05	0.06	0.20	0.33	0.72	1.00
$\{1, 2\}, \{3, 4\}, \{5, \dots, 8\}$	0.06	0.05	0.05	0.04	0.06	0.33	0.61	0.94	1.00

Table 4: Power for Ho and Imai's Rank Test, uni-directional preferences

In the uni-dimensional setting, the rank-compatibility test appears to modestly outperform the linear concordance test, which vastly outperforms the rank test over a fairly wide range of parameter choices. These results are unsurprising. Because the data are generated under the model that motivates the rank-compatibility test, we expect the RC test to be most powerful, and it is. The LC test is sensitive to local departures from this model, and we expect it to pay for this sensitivity somewhat with power under the model. The rank test of Ho and Imai, which is a standard test for uniformity when no *a priori* information is available, is only remotely competitive with the RC and LC tests when the sample size and fraction of prejudicially biased agents are so large that the testing problem is not difficult. Thus, significant gains in power can be made for small samples or sparsely occurring bias by incorporating *a priori* information.

Appendix D: Size of Fisher's Exact Test.

As noted in the introduction, the assumptions required for Fisher's Exact Test (FET) to provide valid p-values are not satisfied in this setting because each observation contributes k *dependent* counts to the cross-tabulation. Intuition suggests that since the counts in the table are dependent, they contain less information than independent counts would, and thus FET will be anti-conservative (overly confident in its decision to reject the null hypothesis, because it fails to account for the dependency). A simple simulation study–see Table 5, below–confirms this hypothesis: rejection rates for FET are multiple times higher than the nominal size of the test for cross-tabulated order data generated under the null hypothesis of full uniformity.

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
k	N=50	100	500	N=50	100	500	N=50	100	500
3	0.06	0.06	0.06	0.15	0.17	0.17	0.24	0.25	0.26
4	0.07	0.07	0.04	0.16	0.17	0.17	0.25	0.27	0.27
5	0.07	0.07	0.06	0.18	0.20	0.18	0.27	0.30	0.29
6	0.05	0.06	0.05	0.17	0.17	0.19	0.29	0.28	0.28

Table 5: Rejection rates under the null hypothesis for Fisher's Exact Test

References

Darcy, R., and McAllister, I. (1990). Ballot position effects. *Electoral Studies* 9(1), 5-17.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications, Vol. II,* Wiley & Sons, New York.

Grant, D. (2017). The ballot order effect is huge: evidence from Texas. *Public Choice* 172(3-4), 421-442.

Ho, D. E. and Imai, K. (2008). Estimating causal effects of ballot order from a randomized natural experiment: the California alphabet lottery. *Public Opinion Quarterly* 72(2), 216-240.

Kendall, M. G. (1962). *Rank Correlation Methods, 3rd Edition,* Griffin & Co., London.

Knuth, D. E. (1981) *The Art of Computer Programming; Volume 2: Seminumeral Algorithms.* Addison-Wesley, Boston.

Krosnick, J., Miller, J. and Tichy, M. (2004). An unrecognized need for ballot reform: the effects of candidate name order on election outcomes. In *Rethinking the Vote: the Politics and Prospects of American Election Reform* 51-73. Oxford U. Press, Oxford.

Meredith, M. and Salant, Y. (2013). On the causes and consequences of ballot order effects. *Political Behavior* 35(1), 175-197.

Murdock , B. B. (1962). The serial position effect of free recall. *J. Experimental Psychology* 64.5, p.482.

Silvapulle, M. and Sen, P.K. (2005). *Constrained Statistical Inference,* Wiley & Sons, New Jersey.

Stuart, A. and J. K. Ord (1987). *Kendall's Advanced Theory of Statistics, Vol. 1, 5th Edition,* Oxford, New York.

Treier, S. and Hillygus, D. S. (2009). The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly* 73(4), 679-703.