Political Analysis (2025), **0**, 1–22
doi:10.1017/pan.2024.33

ARTICLE

Addressing Measurement Errors in Ranking Questions for the Social Sciences

Yuki Atsusaka[1] and Seo-young Silvia Kim[2]

[1] Assistant Professor, Hobby School of Public Affairs, University of Houston, Houston, TX, USA; Email: atsusaka@uh.edu;
[2] Assistant Professor, Department of Political Science and International Relations, Seoul National University, Seoul, South Korea

Corresponding author: Seo-young Silvia Kim; Email: sy.silvia.kim@gmail.com

(Received 14 April 2024; revised 2 September 2024; accepted 8 September 2024)

Abstract

Social scientists often use ranking questions to study people's opinions and preferences. However, little is understood about the general nature of measurement errors in such questions, let alone their statistical consequences and what researchers can do about them. We introduce a statistical framework to improve ranking data analysis by addressing measurement errors in ranking questions. First, we characterize measurement errors from random responses—arbitrary and meaningless responses based on a wide range of random patterns. We then quantify bias due to random responses, show that the bias may change our conclusion in any direction, and clarify why item order randomization alone does not solve the statistical issue. Next, we introduce our methodology based on two key design-based considerations: item order randomization and the addition of an "anchor" ranking question with known correct answers. They allow researchers to (1) learn about the direction of the bias and (2) estimate the proportion of random responses, enabling our bias-corrected estimators. We illustrate our methods by studying the relative importance of people's partisan identity compared to their racial, gender, and religious identities in American politics. We find that about 30% of respondents offered random responses and that these responses may affect our substantive conclusions.

Keywords: measurement error; ranking data; ranking questions; response bias; survey design

Edited by: Jeff Gill

1. Introduction

In the social sciences, researchers often use ranking questions to study public opinion and preferences on various topics (Alvo and Yu 2014; Marden 1996). Using ranking questions, for example, scholars of American politics study media framing (Nelson, Clawson, and Oxley 1997), representation (Costa 2021; Tate 2004), blame attribution (Malhotra and Kuo 2008), political values (Ciuk 2016), and redistricting (Kaufman, King, and Komisarchik 2021). Comparative politics researchers also study nationalism (Miles and Rochefort 1991), post-materialism (Inglehart and Abramson 1994), candidate selection (Jankowski and Rehmert 2022), and ethnic identity (McMurry 2022), whereas international relations works examine foreign aid (Dietrich 2016), economic coercion (Gueorguiev, McDowell, and Steinberg 2020), and sexual violence (Agerberg and Kreft 2023). In addition to research, ranking questions are used in actual and polling of elections with ordinal ballots, such as ranked-choice voting (RCV), single transferable vote, Borda count, and Coombs rule (Atsusaka 2025; Shugart and Taagepera 2017).

Despite the wide usage of ranking, relatively little has been discussed and understood about the general nature of measurement errors in ranking questions. For example, only 3 out of 28 studies that use ranking questions published in the *American Political Science Review*, the *American Journal of Political Science*, and the *Journal of Politics*, 2012–2023, mention potential measurement errors. Meanwhile, some methodological studies have examined measurement issues, but focus only on *specific* aspects of ranking questions, such as item-order effects (Krosnick and Alwin 1987; Malhotra 2009; Serenko and Bontis 2013), question-order effects (Tranter and Western 2010), their advantage in eliciting relative preferences compared to other questions (Alwin and Krosnick 1985; Dillman, Smyth, and Christian 2014; Kaufman *et al.* 2021; Krosnick 1999; Krosnick and Alwin 1988; McCarty and Shrum 2000), and measurement issues by format (Blasius 2012; Genter, Trejo, and Nichols 2022; Smyth, Olson, and Burke 2018), or by devices and layouts (Revilla and Couper 2018).[1]

In this paper, we introduce a general statistical framework for understanding measurement errors in ranking questions based on random responses—rankings based on arbitrary patterns independent of respondents' underlying preferences. With the framework, we clarify what ranking-based quantities researchers can study, what random responses look like in different formats, and why using observed data alone may induce measurement errors. Moreover, we propose simple design-based methods to correct the bias with respect to various quantities of interest due to measurement errors. Using *item order randomization*, we learn about the direction of the bias due to random responses. Leveraging *anchor questions*—auxiliary ranking questions whose correct answers are *ex-ante* known to researchers—we estimate the proportion of random responses. The two pieces of information allow our bias-corrected estimators to estimate many quantities of interest in nonparametric and parametric analyses. In contrast to existing studies, the proposed framework is general and encompasses measurement issues from various sources discussed for ranking questions while also contributing to the growing scholarship on design-based methods to address measurement errors in survey research and beyond (Atsusaka and Stevenson 2023; Berinsky *et al.* 2024; Clayton *et al.* 2023; Horowitz and Manski 1995; Kane and Barabas 2019; Kane, Velez, and Barabas 2023; Tyler, Grimmer, and Westwood 2024).

At first glance, the problem of random responses seems resolvable by randomizing the order of items, and some of the most well-intended studies adopt this strategy.[2] However, we show that randomization *alone* does not remove the bias. Instead, randomization makes random responses follow a uniform distribution. Although this is significantly better than having no randomization, in which measurement errors have an unpredictable direction, the bias still remains under randomization—the distribution of observed rankings is now pulled towards a uniform distribution (i.e., indifference among items). This way, even under randomization, random responses can mask otherwise salient ranked preferences among respondents and "dilute" empirical results. Thus, understanding and overcoming the limitation of randomization has important implications for research using ranking questions.

Measurement errors in ranking questions can also have implications for electoral institutions and democratic representation. Recently, observations of improper ranked ballots in RCV have also been discussed in light of its growing adoption in American elections (Alvarez, Hall, and Levin 2018; Neely and Cook 2008; Neely and McDaniel 2015). For example, Atkeson *et al.* (2024) study voter confusion in RCV, arguing that voting errors may emerge due to the complexity of ballots and the lack of information on candidates. Cormack (2024) examines over-voting—ranking the same candidate more than once, finding their prevalence are higher in areas with lower education and income levels. Similarly, Pettigrew and Radley (2023) classify ballot-marking errors and ballot rejections, concluding that, on average, about 4.7% of voters make at least one type of error. Furthermore, Atsusaka (2025) analyzes ballot order effects—the effect of candidate order on the ballot on voters' entire candidate rankings, showing that about 0.6%–3.0% of voters may provide ranked ballots based on "donkey voting" (ranking candidates in the order they appear on the grid-style ballot). Thus, understanding measurement errors and their solutions may help assess the quality and validity of elections under RCV.

[1] For more discussions on ranking questions, see Appendix E of the Supplementary Material.

[2] For example, Costa (2021, 354) writes, "the [ranking] list of issues respondents could choose from was order randomized." Malhotra and Margalit (2014, 1002) note, "[t]he order in which the six traits were presented was fully randomized to ensure that primacy effects did not bias the rankings." Rathbun and Pomeroy (2022, 678) write, "[a]ttributes are listed randomly to avoid order effects." More recently, Pradel *et al.* (2024, 7) "used a randomized presentation of the concepts [to be ranked] to avoid response ordering effects."

https://doi.org/10.1017/pan.2024.33 Published online by Cambridge University Press

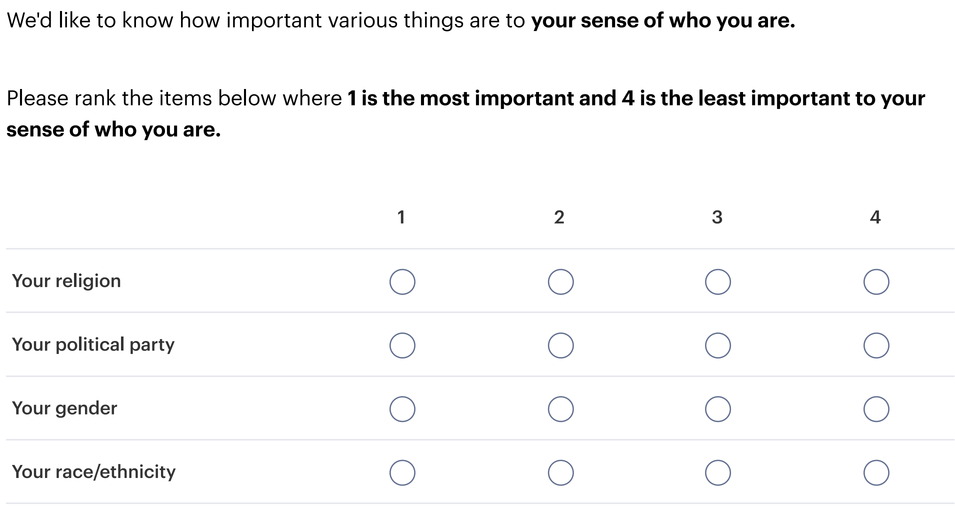


Figure 1. Ranking question to measure relative partisanship.

This paper is organized as follows. In Section 2, we introduce our motivating application of measuring and analyzing the relative importance of different identities. In Section 3, we define measurement errors from random responses and introduce our statistical framework for understanding the consequences of measurement errors. Section 4 introduces our design-based methodology to correct the resulting bias. In Section 5, we illustrate our methods with our empirical application and show that about 30% of respondents offered random responses, which can affect our conclusion. Section 6 provides extended analyses of our methods by comparing our methods to alternative designs and addressing a wider population of interest. In Section 7, we discuss our work's limitations and future directions. Our methods will be available through the `R` package `rankingQ`.

2. Motivating Application: Relative Partisanship

Partisan identity has been one of the most influential variables in modern American politics (Green, Palmquist, and Schickler 2002; Huddy and Bankert 2017; West and Iyengar 2022). Although many works stress the centrality of partisan identity in political and social behavior, relatively little is understood about the *relative* importance of partisan identity (or any identity) compared to other core social identities relevant to politics (for exceptions, see Lee 2009; Spry 2017; Setzler and Yanus 2018). For example, Spry (2021, 434) notes that typical questions in identity and politics ask "respondents to report their closeness to one group at a time, [but] not to multiple groups within the same measure," and as a result, the conventional approach may miss "an opportunity to measure how close a respondent feels to one group category relative to other categories."

Using ranking questions, we seek to measure and analyze the multidimensionality of people's identities and what we call *relative partisanship*—the relative importance of partisan identity.[3] We obtain a representative sample of American adults through YouGov (N = 1,082)[4] and ask respondents to rank four sources of their identities, including their (a) political party, (b) race, (c) gender, and (d) religion, according to their relative importance.[5] Figure 1 shows the ranking question used in the

[3] Of course, ranking is not the only way to measure multidimensional concepts. While multiple rating and point-allocation questions can be useful, our study focuses on ranking questions.

[4] We limited our sample to respondents with a computer or tablet device with a sufficiently large screen size. The survey weights take this decision into account.

[5] Researchers can also focus on other items when analyzing this question. For example, future research may study *relative racial identification*, which may have important implications for minority representation (Atsusaka 2021), collective action (Lopez, Alvarez, and Kim 2022), and other phenomena related to racial politics.

survey. In this particular example, religion is the first item on the list, followed by political party, gender, and race/ethnicity. For the reason we describe below, we randomize item order at the respondent level.

Our question follows a long tradition in political science of using ranking questions to study identity. For example, Miles and Rochefort (1991) test a theory of nationalism by examining whether people near the Nigeria–Niger border rank ethnic identity higher than national consciousness and religious affinity. Similarly, McCauley and Posner (2019) study the relative importance of religion in identity using the Cote d'Ivore-Burkina Faso border as a quasi-natural experiment. McMurry (2022) also uses a rank-order question to assess the relative importance of tribe, religion, gender, and nationality in the Philippines. Recently, Hopkins, Kaiser, and Perez (2023, 12) study the relative importance of partisanship among Latinos and Asian Americans compared to "religion, job/occupation, gender, family role, political party, pan-ethnic group, national origin group, and being American." In addition, the Collaborative Multiracial Post-Election Survey (CMPS) also surveys respondents on their relative identities, such as between national origin, race, and being American (e.g., Question `A195_1` in the 2016 wave).

3. Statistical Framework for Measurement Errors

In this section, we introduce a statistical framework for measurement errors in ranking questions. The framework is general and applies to multiple formats of ranking questions, including radio buttons or grid-style, drag and drop, numeric entry, and select box formats. While the framework itself has been considered in existing studies (Horowitz and Manski 1995), we tailor our discussion to ranking questions and propose two design-based methods below.

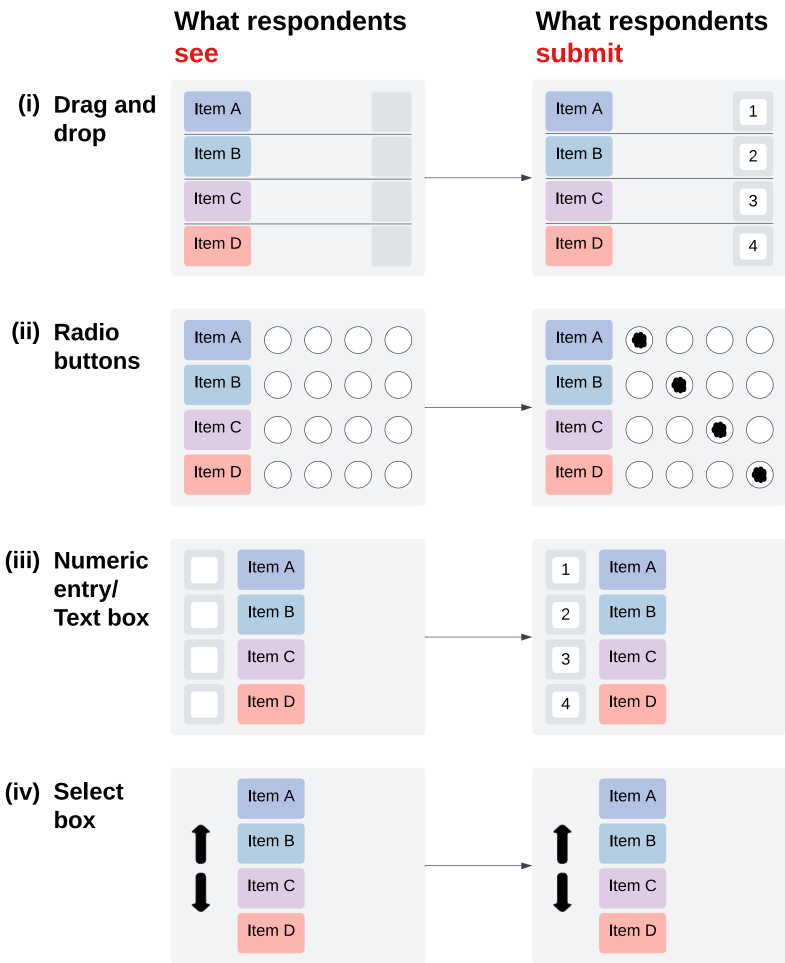
3.1. Setup

Suppose there are J items that respondents rank. Here, we assume that (1) all respondents have well-defined preferences (completeness and transitivity) and (2) respondents rank all items. We also assume that each person has two potential ranking responses: non-random and random responses. First, we define *non-random responses* as responses based on underlying preferences or intentions. Non-random responses do not need to perfectly align with people's underlying preferences as long as they reflect respondents' substantial intent, whether sincere or strategic.

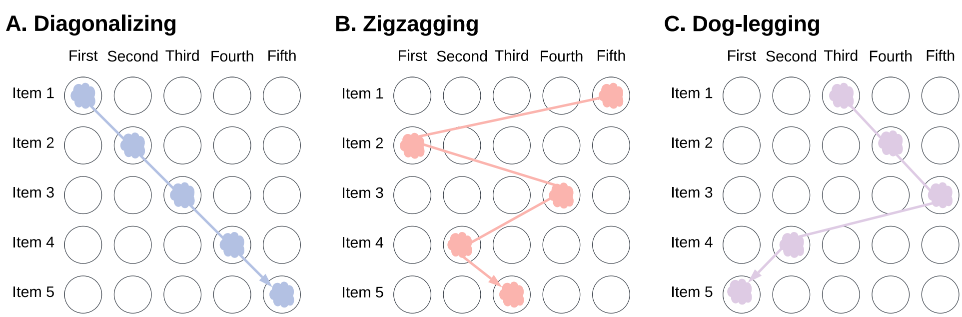
Next, we define *random responses* as meaningless responses that are *independent* of people's preferences—irrelevant to/unreflective of true preferences. Our definition is general, and random responses can take many forms. For example, Figure 2a visualizes what occurs when respondents provide random response $(1,2,3,4)$ in four commonly used ranking question formats. Here, random responders (i) use the same order as a presented list (drag and drop), (ii) draw a diagonal line from the top-left to the bottom-right (radio buttons), (iii) enter in numerical ascension when asked to rank (numeric entry), or (iv) do not reorder a presented item (select box). Many other patterns are also possible. Figure 2b provides three visually intuitive examples of random responses we call diagonalizing, zigzagging, and dog-legging in the radio-button format (for a similar discussion in RCV, see Atsusaka 2025).

Let Y_i^* and e_i be respondent i's $(i = 1,\dots,N)$ non-random and random responses (or errors), respectively. Let z_i be a random variable denoting whether respondent i offers a non-random response $(z_i = 1)$ or otherwise $(z_i = 0)$. We denote respondent i's *observed* response by $Y_i^{\text{obs}} = Y_i^* z_i + e_i(1 - z_i)$. We use a general notation $g(\cdot)$ to represent a ranking-based quantity of interest (QOI).[6]

[6] We assume some linear operator for $g(\cdot)$. The QOI can be some function of non-random rankings, including the probability mass function of unique ranking profiles, $f(Y_i^*)$ (Section 5.2) and the average rank of item j, $\mathbb{E}[Y_{i,j,nr}]$ (Section 5.3), among many others.



(a) Random Response (1, 2, 3, 4) Across Different Formats of Ranking Questions



(b) Different Types of Randomness in Radio-Button Format

Figure 2. Examples of random responses in ranking questions.

https://doi.org/10.1017/pan.2024.33 Published online by Cambridge University Press

Our framework represents the observed ranking data as a mixture of non-random and random responses:

$$\underbrace{g(Y_i^{\text{obs}})}_{\text{raw data}} = \underbrace{g(Y_i^*|z_i = 1)}_{\text{non-random responses}} \times \Pr(z_i = 1) + \underbrace{g(e_i|z_i = 0)}_{\text{random responses}} \times \Pr(z_i = 0) \tag{1}$$

where $\Pr(z_i = 1)$ and $\Pr(z_i = 0)$ are the proportions of non-random and random responses, respectively.

This illustrates why researchers cannot simply use raw data to study their variable of interest—the data contains the information on what they wish to analyze and irrelevant noises that skew their understanding of their target concept. Appendix A.1 of the Supplementary Material formally illustrates the bias from random responses and discusses why it is consequential in empirical studies. The fundamental problem of random responses is that there is no way for researchers to know which part of their data is susceptible to errors (i.e., which respondents offer random responses) and to what extent.

3.2. *Quantities of Interest and Identification Problems*

There are two classes of quantities that interest researchers, differing in terms of the population of interest in which the target quantity is defined. The following distinction relates to the difference between the average engaged response among the engaged and the average engaged response discussed in Tyler *et al.* (2024).

The first quantity is a ranking-based quantity among non-random responses:

$$\theta_z \equiv g(Y_i^*|z_i = 1). \tag{2}$$

This paper mainly studies the identification of θ_z. We highlight that this estimand is only defined among people who offer non-random responses. Rearranging Equation (1), our identification problem becomes

$$\underbrace{g(Y_i^*|z_i = 1)}_{\text{what we wish to study}} = \frac{\overbrace{g(Y_i^{\text{obs}})}^{\text{derived from raw data}} - \overbrace{\Pr(z_i = 0)}^{\text{prop. random resp.}} \overbrace{g(e_i|z_i = 0)}^{\text{random resp.}}}{\underbrace{1 - \Pr(z_i = 0)}_{\text{prop. non-random resp.}}}. \tag{3}$$

The right-hand side contains three quantities: (1) the target quantity based on observed rankings $g(Y_i^{\text{obs}})$, (2) the proportion of random responses $\Pr(z_i = 0)$, and (3) the target quantity based on random responses $g(e_i|z_i = 0)$.

The key problem is that *we only observe the first quantity* $g(Y_i^{\text{obs}})$. Thus, without making any assumptions, the QOI will never be estimated from raw data alone, regardless of how many responses researchers collect. Section 4 discusses how our design-based methods allow us to point-estimate the latter two unknowns.

The second quantity of interest is a ranking-based quantity in the target population from which samples are drawn:

$$\theta \equiv g(Y_i^*) \tag{4}$$

$$= \underbrace{g(Y_i^*|z_i = 1)}_{\theta_z} \Pr(z_i = 1) + \underbrace{g(Y_i^*|z_i = 0)}_{\text{counterfactual}} \Pr(z_i = 0), \tag{5}$$

where, θ becomes closer to θ_z as the probability of error-free responses $\Pr(z_i = 1)$ increases.

Generally, θ is more difficult to identify than θ_z because it requires an additional assumption about the counterfactual quantity $g(Y_i^*|z_i = 0)$—non-random rankings that random respondents would have

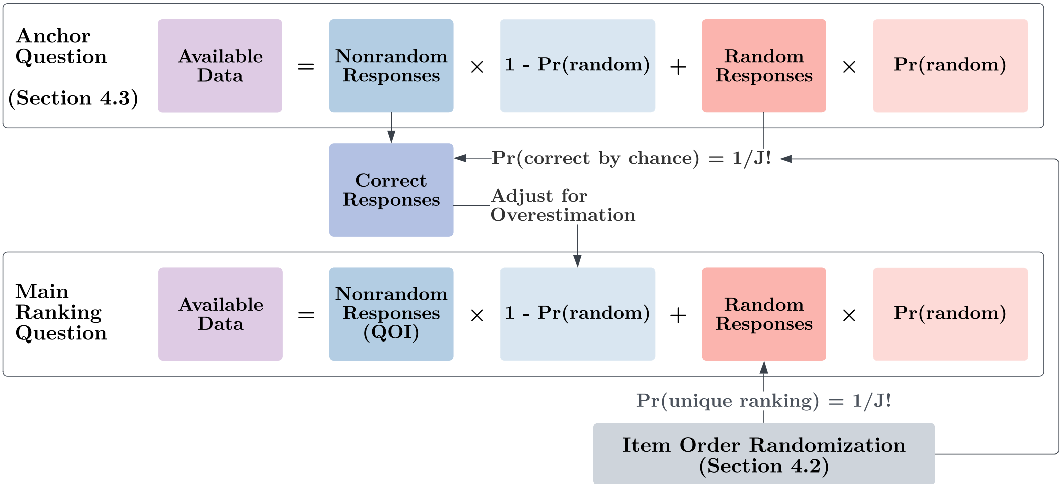


Figure 3. Design-based methods for estimating the proportion and distribution of random responses.

provided had they not responded randomly. In Section 6, we discuss three identification strategies for this population-level quantity.[7]

4. Proposed Methodology

4.1. Overview

Figure 3 summarizes our design-based methodology, which leverages two survey designs: item order randomization and an anchor question.

4.2. Item Order Randomization

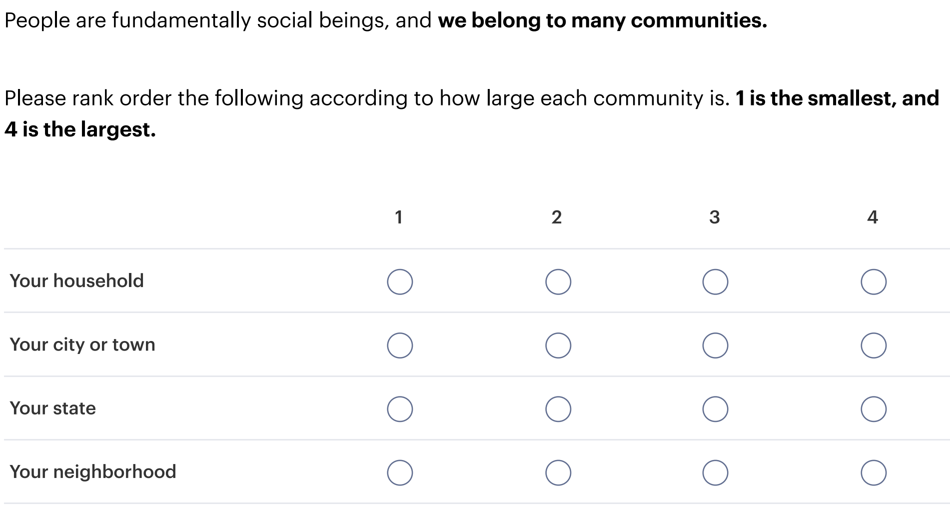
The first design consideration is item order randomization. Many survey platforms support randomization, and some studies use item order randomization with ranking questions (Costa 2021; Malhotra and Margalit 2014; Pradel *et al.* 2024; Rathbun, Rathbun, and Pomeroy 2022). The primary role of item order randomization is to identify the distribution of rankings among random responses and the direction of the bias with respect to our quantities of interest.

Our key theoretical result is that, under item order randomization, the rankings among random responses follow a uniform distribution with probability $\frac{1}{J!}$, where J is the number of items.[8] For example, with three items, random responses will correspond to one of the six profiles in the set $\{123, 132, 213, 231, 312, 321\}$ with probability $\frac{1}{6}$.[9]

[7] While this work focuses on point identification and estimation, other works discuss approaches based on partial identification (Horowitz and Manski 1995; Tyler *et al.* 2024). While these do not directly address ranking questions, future research can extend our framework by drawing from these perspectives.

[8] See Appendix A.2 of the Supplementary Material for proof and details. For a simple, intuitive example, consider a binary question with two choices—YES and NO. Suppose that all respondents offer random responses. Suppose also that 30% and 70% of them pick the first and second option, respectively. When YES always appears first and NO second (i.e., no randomization), the distribution of the two answers is (0.3, 0.7). However, with item order randomization, it becomes (0.5, 0.5) since half the 30% picks YES and the other half NO, while half the 70% selects YES and the other half NO. Our result is an extension of this example to ranking questions.

[9] See Appendix A of the Supplementary Material for technical discussions.


People are fundamentally social beings, and **we belong to many communities.**

Please rank order the following according to how large each community is. **1 is the smallest, and 4 is the largest.**

	1	2	3	4
Your household	○	○	○	○
Your city or town	○	○	○	○
Your state	○	○	○	○
Your neighborhood	○	○	○	○

Figure 4. Example of an anchor ranking question.

Let U_J be a random variable that follows a discrete uniform distribution with J ranking items. Then, our result implies that

$$g(e_i | z_i = 0) = g(U_J). \tag{6}$$

This result is powerful because it holds *regardless* of the pattern of response patterns—item order randomization transforms all random responses into a set of rankings following the uniform distribution.

Our result also clarifies why the bias still remains even after randomization. Integrating Equations 1 and 6, we can show that observed data still contain random responses under randomization as follows:

$$\underbrace{g(Y_i^{\text{obs}})}_{\text{raw data}} = \underbrace{g(Y_i^* | z_i = 1)}_{\text{non-random responses}} \times \Pr(z_i = 1) + \underbrace{g(U_J)}_{\text{random responses}} \times \Pr(z_i = 0). \tag{7}$$

The above equation shows that, under randomization, random responses will pull any estimates towards what researchers may observe when all respondents are indifferent among available items. Thus, even under randomization, random responses still affect researchers' substantive inferences by "diluting" their conclusions.

4.3. Anchor Questions

To identify the proportion of random responses *at the time of the target ranking question*, $\Pr(z_i = 0)$, we propose using an auxiliary ranking question whose "correct answer" is *ex-ante* known to researchers (for a similar idea, see Atsusaka and Stevenson (2023)). We call this an *anchor question* and ask it right before or after the target ranking question.[10] The item order in the anchor question must be randomized just as it must be randomized for the target question. To illustrate, Figure 4 presents the anchor question we included in our survey. In this example, the question asks respondents to rank four communities from the smallest to the largest, and the correct answer is (household, neighborhood, city or town, state).[11]

[10]Note that question order must be randomized.
[11]Cases with multiple correct answers can also be accommodated; see Section 5.1.

Estimating the Proportion of Random Responses in the Anchor Question.
First, we estimate the proportion of random answers in the anchor question by using correct responses. Let c_i be a binary variable indicating whether respondent i offers the correct answer in the anchor question $(c_i = 1)$ or not $(c_i = 0)$. Let $\Pr(z_i^{\text{anc}} = 0)$ and $\Pr(z_i^{\text{anc}} = 1)$ be the proportions of random and non-random responses in the anchor question, respectively.[12] Under item order randomization, we can estimate the proportion of random responses in the anchor question using the following estimator.

Proposition 1. *Unbiased Estimator of the Proportion of Random Responses in the Anchor Question*

$$\widehat{\Pr}(z_i^{anc} = 0) = 1 - \widehat{\Pr}(z_i^{anc} = 1) \tag{8}$$

$$= 1 - \underbrace{\left[\frac{\sum_{i=1}^{N} c_i}{N} - \frac{1}{J!}\right]}_{\text{adjust overestimation}} \underbrace{\left(1 - \frac{1}{J!}\right)^{-1}}_{\text{normalization}}. \tag{9}$$

The proof is in Appendix A.3 of the Supplementary Material. Equation 9 has intuitive interpretations. The second term suggests that the proportion of non-random responses can be estimated from the proportion of correct answers $\frac{\sum_{i=1}^{N} c_i}{N}$ after accounting for the probability that random responses *happen to be correct* (which is $\frac{1}{J!}$ under item order randomization).[13] The third term can then be interpreted as renormalization to ensure that the resulting quantity becomes probability.

Estimating the Proportion of Random Responses in the Target Question.
Next, we estimate the proportion of random responses in the target question using the above result. To allow this extrapolation, we make the following assumption.

Assumption 1 (Constant Proportion of Random Responses). *The proportion of random responses remains constant across the target and anchor questions or* $\Pr(z_i^{anc} = 0) = \Pr(z_i = 0)$.

One key advantage of our approach is that it allows researchers to *design* their anchor questions so that Assumption 1 becomes more plausible—researchers can tailor an anchor question to their target question so that the two ranking questions have similar substantive topics, instruction length, the number and length of items, and locations in the survey (i.e., the anchor should come right before or after the target question). Appendix D of the Supplementary Material offers a practical guide for building anchor questions. Another advantage is that it does not assume individual-level randomness to be constant across the questions (see also Section 6.1).

4.4. Bias-Corrected Estimators

Integrating the above results, we propose two approaches to correct measurement errors. The first strategy is to directly correct the bias with a specific quantity of interest. The second approach is to apply the idea of inverse probability weighting (IPW).

[12] They are not the probability that specific respondent i offers a random response.
[13] As the number of items increases, the impact of $\frac{1}{J!}$ quickly diminishes, and the anchor question will carry the most weight in Equation 9.

Figure 5. Graphical representation of IPW.

4.4.1. Direct Correction

The first approach is to use the following bias-corrected estimator:

$$\widehat{g}(Y_i^*|z_i = 1) = \frac{\overbrace{\widehat{g}(Y_i^{\text{obs}})}^{\text{raw data}} - \overbrace{g(U_J)}^{\text{random resp.}} \overbrace{\left(1 - \left[\frac{\sum_{i=1}^N c_i}{N} - \frac{1}{J!}\right]\left(1 - \frac{1}{J!}\right)^{-1}\right)}^{\text{prop. random resp.}}}{\underbrace{\left[\frac{\sum_{i=1}^N c_i}{N} - \frac{1}{J!}\right]\left(1 - \frac{1}{J!}\right)^{-1}}_{\text{prop. non-random resp.}}}. \tag{10}$$

This estimator is simple and only requires one extra estimation $\frac{\sum_{i=1}^N c_i}{N}$ compared to the naïve estimator $\widehat{g}(Y_i^{\text{obs}})$ while retaining the original sample size. Our proposed estimator has a wider confidence interval than the naïve estimator due to the additional uncertainty around the estimated proportion of correct answers. We use bootstrapping for constructing confidence intervals. Moreover, researchers can include survey weights in $\widehat{g}(Y_i^{\text{obs}})$ as in typical survey data analysis, where survey weights represent the product of the design weight and a poststratification or calibration adjustment.

4.4.2. Inverse Probability Weighting

The second strategy is to leverage the idea of IPW. Under this framework, the problem of measurement errors (Equation 1) can be considered an issue of selection bias. Figure 5 illustrates this idea graphically. Here, due to random responses, relatively popular rankings (e.g., 123) are under-sampled, while relatively unpopular rankings (e.g., 231) are over-sampled compared to their true distribution. A natural solution is to *weight up* a set of rankings that are supposed to be more prevalent and *weight down* a set of rankings that are supposed to be less prevalent than what raw data suggest.

Table 1. Comparison of two bias correction methods.

	Pro	Con	Useful scenario
Direct correction	exact	not flexible	nonparametric analysis
IPW framework	flexible	not exact	parametric analysis

Let $\mathbb{P}(Y_i^*|z_i = 1)$ be the population proportion of respondent i's ranking profile given non-random responses. Let $\mathbb{P}(Y_i^{\text{obs}})$ be the same proportion based on observed data, and $w = \{w_i\}_{i=1}^N$ be a vector of weights for N respondents included in nonparametric or parametric analyses. We propose the following inverse probability weight:

$$w_i = \left[\frac{\mathbb{P}(Y_i^{\text{obs}})}{\mathbb{P}(Y_i^*|z_i = 1)}\right]^{-1}. \tag{11}$$

The weights can be nonparametrically identified via the following plug-in estimator:

$$\widehat{w}_i = \left[\frac{\widehat{\mathbb{P}}(Y_i^{\text{obs}})}{\widehat{\mathbb{P}}(Y_i^*|z_i = 1)}\right]^{-1}. \tag{12}$$

Let $\mathbb{P}(U_J)$ be the uniform distribution with probability $\frac{1}{J!}$. Building on a similar derivation to Equation A.7, the denominator can be unbiasedly estimated with the following estimator:

$$\widehat{\mathbb{P}}(Y_i^*|z_i = 1) = \frac{\widehat{\mathbb{P}}(Y_i^{\text{obs}}) - \mathbb{P}(U_J)\left(1 - \left[\frac{\sum_{i=1}^N c_i}{N} - \frac{1}{J!}\right]\left(1 - \frac{1}{J!}\right)^{-1}\right)}{\left[\frac{\sum_{i=1}^N c_i}{N} - \frac{1}{J!}\right]\left(1 - \frac{1}{J!}\right)^{-1}}. \tag{13}$$

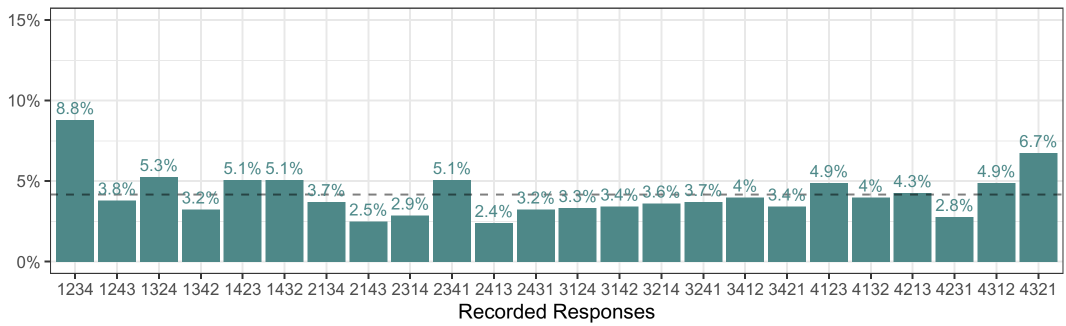
Researchers can also use survey weights in the IPW framework by constructing a new weight $w_i^* = w_i w_i^s$, where w_i^s is respondent i's survey weight.

4.4.3. Methods Selection

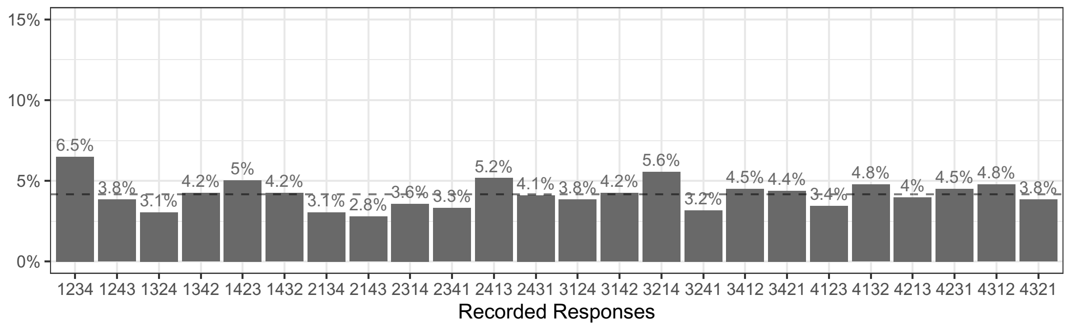
The two approaches complement each other. Table 1 provides a comparison. We recommend that researchers use the direct approach whenever their target quantities are simple and nonparametrically identifiable (e.g., average ranks) as it provides exact bias correction to their QOIs. In contrast, when they wish to perform more complex and parametric analyses, such as running regressions, the IPW framework can be helpful, as it allows researchers to perform any analyses with the bias-correction weights.

4.5. Uniformity Test

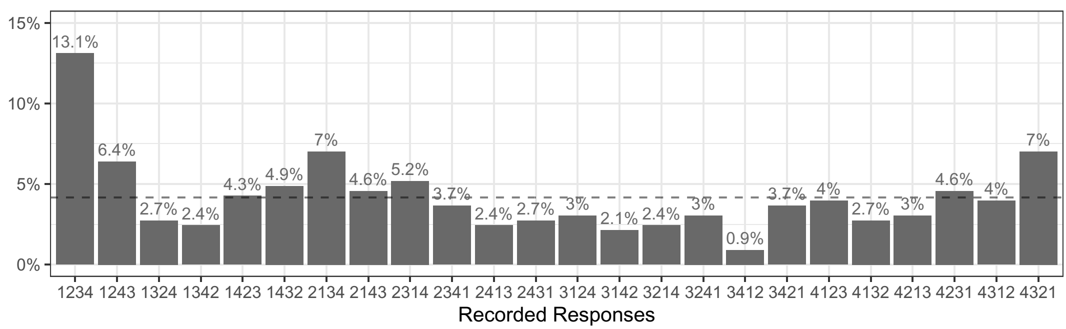
Finally, our framework also allows researchers to detect the presence of random responses without requiring any anchor questions. Appendix B of the Supplementary Material introduces the uniformity test, which shows that *recorded responses* (what respondents submit in Figure 2; see Appendix A.2 of the Supplementary Material) will follow a uniform distribution in the absence of random responses. Conversely, non-uniformity in the test suggests the presence of random responses.



(a) Identity Question, All Respondents



(b) Anchor Question, Respondents with Correct Answers



(c) Anchor Question, Respondents with Incorrect Answers

Figure 6. Visualization of the uniformity test: Distribution over all possible recorded responses in the target and anchor questions. Note: The dashed line represents 1/24 × 100%, to which the distribution should converge in the absence of random responses.

5. Measuring and Analyzing Relative Partisanship

Using our proposed method, we present the analysis of relative partisanship in American politics.[14] We focus on how bias-corrected estimates can differ from unadjusted estimates under different analyses, leaving more detailed analyses for future research. All analyses incorporate survey weights calculated by the polling firm.

Do our data contain random responses? To first address this question, Figure 6a shows the result of the uniformity test applied to our identity question. The figure visualizes the distribution of respondents' recorded responses—the exact patterns they provided with respect to the four items presented in a given

[14]The replication data and code for this article are available in Atsusaka and Kim (2024).

Table 2. Distribution of responses to the anchor question.

Anchor ranking	Count	Percent (%)	Anchor ranking	Count	Percent (%)
1234	**754**	**69.80**	3124	4	0.40
1243	24	2.20	3142	1	0.10
1324	40	3.70	3214	9	0.80
1342	12	1.10	3412	6	0.60
1423	18	1.70	3421	8	0.70
1432	12	1.10	4123	10	0.90
2134	11	1.00	4132	4	0.40
2143	5	0.50	4213	4	0.40
2314	9	0.80	4231	8	0.70
2341	6	0.60	4312	4	0.40
2413	5	0.50	4321	117	10.80
2431	10	0.90			

Note: 1234 corresponds to household → neighborhood → city or town → state, which is coded as the correct anchor response.

order. That is to say, the integers refer to submitted "patterns," as shown in Figure 2b, whose substantive meanings differ across respondents. For example, "1234" means that respondents ranked the four items in the order they appear in the question, regardless of what items were presented to them.

Since there are 4! = 24 possible ways to rank, the proportion (percentage) of recorded responses should converge to 1/24 = 0.042 (4.2%) in the absence of random responses (see A.4 for proof). In contrast, the graph shows clear evidence for non-uniformity—some recorded responses, notably 1234 (8.8%) and 4321 (6.7%), are more likely to occur than they are supposed to under the null (chi-squared test statistic = 68.45, p-value < 0.001), suggesting the presence of random responses in the data.

Checking for uniformity also validates the usage of our anchor question. Figure 6b applies the test to the anchor question only among respondents with correct answers. The result shows a more or less uniform distribution, and the χ^2 test does not reject the null (χ^2 test statistic = 32.70, with p-value of 0.1066).[15] In contrast, Figure 6c visualizes the test among those who offer incorrect anchor responses. It offers clear evidence for non-uniformity, where about 20% of respondents submitted either 1234 or 4321 (χ^2 test statistic = 107.95, p-value < 0.001).

5.1. *Analysis of the Anchor Question*

First, we estimate the proportion of random responses using our anchor question. Table 2 reports the number of each ranking response to the anchor question. We code 1234 (household < neighborhood < city or town < state) as the correct response ($c_i = 1$) and other rankings as the incorrect response ($c_i = 0$). The result shows that the empirical proportion of correct responses is $\frac{\sum_{i=1}^{N} c_i}{N} = \frac{754}{1082} \approx 0.697$. Proposition 1 states that the estimated proportion of random responses can be estimated as $1 - [\frac{754}{1082} - \frac{1}{24}](1 - \frac{1}{24})^{-1} \approx 0.316$. That is, we find that about 31.6% of ranking answers are random responses in the anchor question.[16]

[15] It still shows some non-uniformity because some random responses can pass the anchor question with probability 1/24.

[16] Note that the latter is *not* $1 - \frac{\sum_{i=1}^{N} c_i}{N} \approx 0.303$ since random responses happen to be correct by chance (which is what Proposition 1 accounts for).

By invoking Assumption 1, we estimate that 31.6% of respondents offer random responses in our target ranking question. We believe that this is not an unreasonably high (or low) estimate. For example, Berinsky, Margolis, and Sances (2014), using four different screeners, show that the failure rates for them ranges from 34% to 41%. Relatedly, Atsusaka (2025) finds that about 31% and 37% of respondents in survey experiments offered the same ranking response to two different ranking questions. Moreover, Clayton *et al.* (2023) study measurement errors in conjoint experiments and estimate that about 19.3–27.0% of respondents offered different responses to an identical question that was asked twice.

In some cases, researchers may encounter "debatable cases," in which multiple responses can be considered correct even after carefully designing anchor questions. For example, based on substantive knowledge, some researchers may think that for some people, the size of the community should be ordered as household < neighborhood < state < city or town. In another example, after analyzing data, analysts may find that some ranking responses are more prevalent than other incorrect answers (e.g., 4321 in Table 2). In these cases, analysts can code more than one response as correct answers. Moreover, researchers can also give credit to "partially correct answers," if any, by coding such responses with known probability (e.g., coding an 80% correct response as correct with probability 0.8). Furthermore, it is also possible to use the most conservative (only one correct answer) and most liberal (as many correct answers as possible) coding schemes to make bounds for the resulting estimates.

Note that conservative (liberal) coding leans toward an over-estimation (under-estimation) of the prevalence of random responses in the anchor question. More importantly, the main focus should be on satisfying Assumption 1 when researchers consider different coding schemes for c_i. If anything, we recommend underestimating rather than overestimating the proportion of random responses in the target question because it leads to under-correction of the bias, which guards against inflating Type I errors. For example, if researchers suspect that there are *more* random responses in the anchor than in the main question (e.g., the anchor looks like an attention check, which caused more respondents to answer randomly), it would be better to code *fewer* anchor responses as correct answers should the coding is debatable.

5.2. Summarizing Data with Empirical Distributions

We begin our analysis by describing the distribution of our data. The left panel of Figure 7 presents the distribution of all possible rankings with our methods, with the gray region indicating where party was ranked first. We find that while a great variation in the ranking outcome exists, people rarely rank political party as their first choice. This may provide evidence that relative partisanship is rather low among American adults—a notable finding given the emphasis on partisan identity in American political behavior. We also identify that three rankings/orderings are particularly prevalent, including (gender, race, religion, party), (gender, race, party, religion), and (religion, gender, race, party).

The right panel of Figure 7 visualizes what researchers could have observed had our methods not been applied (but item order randomization is still implemented). Here, many unpopular rankings (e.g., those starting with party) are overrepresented due to random responses. Indeed, the panel leads to a different conclusion that the relative importance of party is as much as that of race. Again, this demonstrates that random rankings, under item order randomization, pull the naïve estimates towards uniformity, where each ranking profile is equally prevalent.

5.3. Understanding Average Patterns

Next, we study the average ranks of the four items as another way to measure relative partisanship. Figure 8 visually compares the results based on our methods and raw data. Overall, we find that the average rank is the lowest for political party, followed by religion, race and ethnicity, and gender. Consistent with our statistical argument (Appendix A.1 of the Supplementary Material), the difference between bias-corrected and unadjusted estimates (and thus bias) is larger when the unknown target parameter is farther away from the average rank based on uniformity (in this case, $\frac{1+4}{2} = 2.5$).

https://doi.org/10.1017/pan.2024.33 Published online by Cambridge University Press

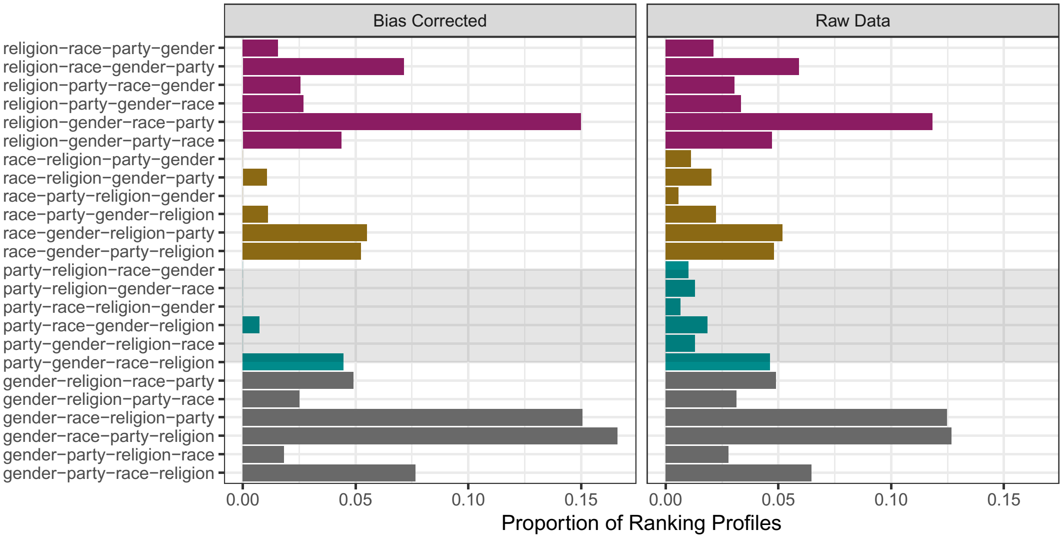


Figure 7. Distributions of identity rankings with bias-corrected and raw data.

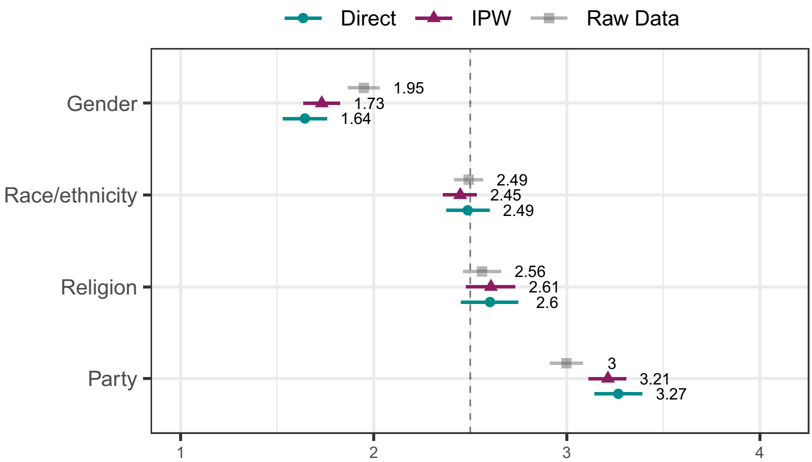


Figure 8. Average ranks with and without bias correction
Note: The dashed line represents the average rank that arises when people are indifferent among the four items.

For party and gender, bias-corrected estimates are statistically significantly different from unadjusted estimates and closer to their bound values (1 and 4). Accordingly, the difference between religion and party is 1.52 (direct) and 1.36 (IPW) times larger in our methods than unadjusted estimates. Similarly, the difference between party and gender is 1.55 (direct) and 1.41 (IPW) times larger in our estimates than unadjusted estimates. In contrast, bias-corrected and raw-data estimates are similar for race and religion. This is consistent with our argument because while bias pulls the estimated average ranks of race and religion towards 2.5, unadjusted estimates of the two items were already close to the value. This illustrates that the magnitude of the bias and the difference between bias-corrected and unadjusted estimates varies not only by the proportion of random responses but also by the values of the target parameters. Thus, researchers should keep in mind that finding a small difference for a particular item after bias correction does not mean that the methods "failed" to address measurement errors.

Researchers can also estimate many other quantities of interest while applying bias correction. For example, our software, `rankingQ`, supports the pairwise ranking probability for items j and

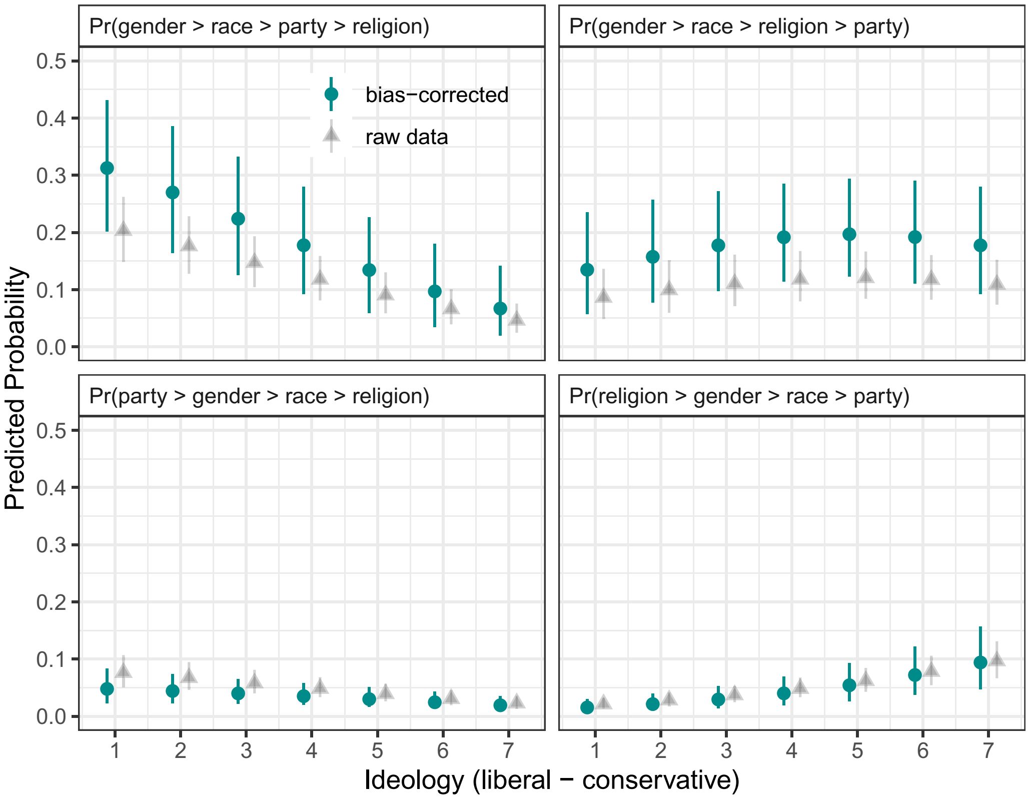


Figure 9. Predicted probabilities with and without bias correction.

j' $\Pr(Y_{ij} < Y_{ij'})$, the top-k ranking probability $\Pr(Y_{ij} \leq k)$, and the marginal ranking probability $\Pr(Y_{ij} = k)$ in addition to the average rank $\mathbb{E}[Y_{ij}]$.

5.4. Regression and Predicted Probability

Moreover, we analyze how respondent characteristics influence their relative partisanship while applying bias correction. To do so, we construct a Plackett–Luce model (also known as rank-order logistic regression) to associate people's ranking choices with their attributes (Alvo and Yu 2014; Train 2003). We regress identity rankings on age, gender, race, education level, ideology, partisanship, and region, while incorporating survey and bias-correction weights via the IPW framework. After estimation, we generate the predicted probabilities that people submit a particular ranking profile with 95% confidence intervals via parametric bootstrapping (Tomz, Wittenberg, and King 2003) over the range of the ideology variable (7-point scale).

Here, we examine how ideology influences relative partisanship among Americans who are 40 years old, white, male, independent, with some college education, and living in the Northeast. We examine four ranking profiles, including the three most prevalent rankings discussed in Section 5.2 and the most prevalent ranking profile starting with party.

Figure 9 presents our results. The top-left panel shows that people are more likely to choose (gender, race, party, religion) as they become more liberal, all things being equal. The first difference in predicted probabilities between the most liberal and conservative Americans with bias correction is roughly

https://doi.org/10.1017/pan.2024.33 Published online by Cambridge University Press

$0.313 - 0.067 = 0.246$, which is almost 1.6 times larger than its unadjusted counterpart of $0.203 - 0.046 = 0.157$. This illustrates how random responses can weaken the association between an independent variable and a target ranking profile.

The top-right panel suggests that ideology only weakly relates to ranking (gender, race, religion, party). While both results show similar patterns, on average, bias-corrected predictions (0.175) are 0.077 points higher than unadjusted predictions (0.109). Thus, without correction, researchers can underestimate the prevalence of the target ranking profile by 1.61 times. Finally, the lower two panels provide examples of relatively similar bias-corrected and unadjusted predictions. Importantly, this does not mean that our methods "did not work." Rather, it illustrates that the nature of bias depends on the target ranking profile, the target independent variable, and the reference values at which other variables are fixed, in addition to the proportion of random responses.

6. Extended Analysis

6.1. Comparison with Alternative Designs

Listwise Deletion.

Some readers may wonder how our methods differ from more traditional solutions relying on attention checks, repeated questions, and so on. More specifically, how is our proposal different from *listwise deletion* based on these alternative design considerations?[17]

Let z_i^* be a binary variable taking 1 if respondent i passes a certain instructional manipulation check and 0 otherwise. For example, $z_i^* = 1$ when respondent i passes an attention check, provides the same answer to the same question asked multiple times, or does not speed through the target question. Researchers then *drop* all respondents who did not pass the test (i.e., delete all i if $z_i^* = 0$) and produce "cleaned" data of size N_c, $\{Y_i^{\text{obs}}(z_i^* = 1)\}_{i=1}^{N_c}$.

When adopting this strategy, researchers often implicitly assume the following.

Assumption 2 (Individually Constant Randomness). *Random responders in the target ranking question are identical to those who fail the instructional manipulation check. Formally, $z_i = z_i^*$ for all $i = 1, \ldots, N$.*

Invoking Assumption 2, listwise deletion identifies θ_z as follows:

$$g(Y_i^{\text{obs}} | z_i^* = 1) = g(Y_i^* | z_i = 1) \quad (14)$$

$$= \theta_z. \quad (15)$$

This way, listwise deletion along with Assumption 2 allows researchers to estimate θ_z directly from the "cleaned" data. In other words, it is assumed that those who failed the test also provided random responses *to the target ranking question*. Importantly, this assumption requires that attention is *stable* for all respondents across the test and the target question. In this sense, Assumption 2 is much stronger than Assumption 1, which requires only the *proportion* of random responses to be the same.

Although the assumption is not directly verifiable, we collected auxiliary information to examine its plausibility in our survey. We find that Assumption 2 is indeed strong; as we show in Appendix C.3 of the Supplementary Material, even between two attention checks, there is very little correlation ($\rho = 0.25$). This is why we propose anchor questions—to approximate the randomness in the target ranking question by using a similar ranking question asked right before or after it.

Alternative Anchors.

Researchers may also wish to try multiple anchor questions and study their effectiveness for pilot studies. To illustrate, we added two additional anchor questions that respectively ask respondents to

[17] Another possibility is to use these alternative designs to estimate the proportion of random responses; we direct researchers towards Appendix C of the Supplementary Material to carefully explore the implications.

(a) alphabetically order four items and (b) order them in the exact order we provide. Appendix C of the Supplementary Material reports the comparison of the empirical results based on them (along with the results using listwise deletion from attention checks and repeated questions).

As highlighted in Sections 4.4 and 5.1, researchers have full control of the choice of and the assessment of the measure to estimate $\Pr(z = 0)$ (that is why our methodology is "design-based"). Some key takeaways are, however, that any anchor questions or instructional/factual manipulation checks need to be pretested and checked for a few sanity measures such as response time, situated adjacent to the main ranking question of interest, and preferably also a ranking format (Appendix C of the Supplementary Material).

6.2. *Identification of* θ

We now propose three identification strategies for $\theta = g(Y_i^*) = g(Y_i^*|z_i = 1)\Pr(z_i = 1) + g(Y_i^*|z_i = 0)\Pr(z_i = 0)$—the ranking-based quantity among *all* people in the target population. The key is to identify $g(Y_i^*|z_i = 0)$, which is a function of counterfactual rankings that random respondents would have provided had they responded non-randomly.

The first approach is to assume that those who provide random responses are indifferent among available items. More specifically, we assume that the counterfactual ranking of J items is a uniformly distributed random variable $\mathcal{U}_J$.

Assumption 3 (Uniform Preference). $Y_i^*(z_i = 0) = U_J$.

This assumption is plausible, for example, when respondents offer random responses because they do not have sufficient information about available options. Here, randomness and preference are correlated. For example, in RCV elections, voters with low education levels may be more likely to provide random responses and have uniform preferences as they have less contextual knowledge to rank multiple candidates.

With Assumption 3, it is straightforward to compute $g(Y_i^*|z_i = 1)$ using a uniform distribution and then estimate θ accordingly. However, our design-based methods provide an even simpler solution. Using item order randomization, we can show that

$$\theta = \underbrace{g(Y_i^*|z_i = 1)}_{\theta_z}\Pr(z_i = 1) + g(Y_i^*|z_i = 0)\Pr(z_i = 0) \tag{16}$$

$$= \underbrace{g(Y_i^*|z_i = 1)}_{\theta_z}\Pr(z_i = 1) + g(U_J)\Pr(z_i = 0) \tag{17}$$

$$= \underbrace{g(Y_i^*|z_i = 1)}_{\theta_z}\Pr(z_i = 1) + g(e_i|z_i = 0)\Pr(z_i = 0) \tag{18}$$

$$= g(Y_i^{\text{obs}}). \tag{19}$$

In other words, θ can be estimated directly from raw data alone.

A second approach is to assume that random respondents would have submitted similar rankings to non-random respondents. More specifically, we assume the following.

Assumption 4 (Contaminated Sampling). $Y_i^* \perp z_i$.

This assumption is plausible, for example, when random responses are based on simple misunderstandings, confusions, or mistakes that prevent respondents from expressing their underlying preferences. We call this assumption *contaminated sampling* building on Horowitz and Manski (1995). With Assumption 4, researchers can identify θ by replacing counterfactual rankings with observed ones as follows:

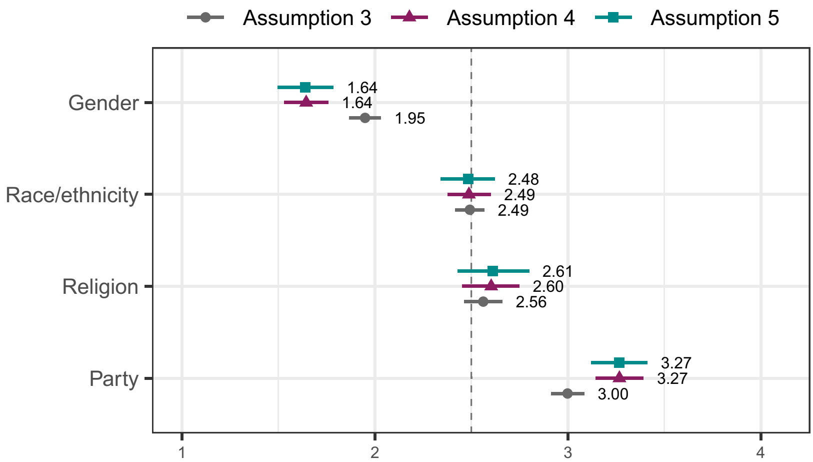


Figure 10. Average ranks in the entire population under different assumptions.
Note: The dashed line represents the average rank that arises when people are indifferent among the four items.

$$\theta = g(Y_i^*|z_i = 1)\Pr(z_i = 1) + g(Y_i^*|z_i = 0)\Pr(z_i = 0) \tag{20}$$
$$= g(Y_i^*|z_i = 1)\Pr(z_i = 1) + g(Y_i^*|z_i = 1)\Pr(z_i = 0) \tag{21}$$
$$= g(Y_i^*|z_i = 1) \tag{22}$$
$$= \theta_z. \tag{23}$$

Assumption 4 is violated whenever there exists a confounder that relates to both randomness and preference. Our final approach is to relax the assumption by conditioning on such a confounder. Let $\mathbf{X}_i$ be a set of covariates that are related to both random responding z_i and preference Y_i^*. We assume the following.

Assumption 5 (Stratified Contaminated Sampling). $Y_i^* \perp z_i|\mathbf{X}_i$.

For simplicity, consider a single confounder. Let $\mathbf{x}$ be a specific covariate value and $\mathcal{X}$ be its sample space. Combined with Equation 23, we propose the following identification strategy via stratification:

$$\theta = \sum_{\mathbf{x}\in\mathcal{X}} \theta(\mathbf{X}_i = \mathbf{x})\Pr(\mathbf{X}_i = \mathbf{x}) \tag{24}$$
$$= \sum_{\mathbf{x}\in\mathcal{X}} \theta_z(\mathbf{X}_i = \mathbf{x})\Pr(\mathbf{X}_i = \mathbf{x}). \tag{25}$$

In other words, we compute the weighted average of θ_z in each distinct category defined by the covariate, where the weight is the proportion of each stratum $\Pr(\mathbf{X}_i = \mathbf{x})$. For example, suppose that strength in partisanship is related to both random responding and identity ranking. Then, researchers can estimate θ by estimating θ_z within groups of people who have reported the same partisan strength and then sum up the estimates while weighting them by the proportions of the groups.

To illustrate the three strategies, Figure 10 presents the estimates of θ (average rank) under the three different assumptions. We use partisan strength (Independent, Weak Partisan, Strong Partisan) to illustrate the stratification approach, which yields similar estimates to the contaminated sampling approach. The uniform preference approach yields estimates closer to $\frac{1+4}{2} = 2.5$ than the other two methods, consistently with its assumption. This way, researchers can extend their inference to θ by leveraging their substantive knowledge about why random responses may occur in their specific application.

7. Concluding Remarks

We introduced a statistical framework to quantify and address measurement errors in ranking survey questions due to random responses. We show that two additional survey designs—item order randomization and a paired anchor ranking question—will help us learn about the direction and magnitude of measurement errors, enabling our bias corrections. Without any corrections, substantial conclusions can be biased in completely unpredictable directions. Even with the current best practice of item order randomization, random responses may conceal otherwise interesting patterns in ranking data. More specifically, we illustrated that measurement errors make the distribution of observed rankings closer to a uniform distribution under randomization, still affecting our inferences.

Using a motivating application that measures relative identities, we show that more than 30% of respondents can fall prey to random responses, and not accounting for this can affect our substantive conclusions. We also show that our methods are valid by showing that recorded responses among respondents who pass the anchor question are close to a uniform distribution, while those who do not show a wildly non-uniform pattern. Our framework provides a heightened understanding of why observed ranking data may be contaminated and what information we require to correct the resulting bias.

Although our current framework focuses on full-ranking questions, it can be extended in several ways for future studies. For example, future research may study how our theoretical results change in more complicated situations that allow partial rankings, top-k rankings, and tie rankings. Moreover, our methods can be extended to other discrete-choice questions, such as binary, multinomial, and ordered-choice questions. In fact, many of our methods, including the uniformity test, randomization, and anchor questions, can be readily applicable to many discrete-choice questions, although such applications may involve unique challenges (e.g., the inability to randomize option order in ordered-response questions). With these future directions, this work contributes to a growing body of design-based methodologies to counter measurement errors in survey research. We hope this work is also informative to election administrators and election science scholars as the number of jurisdictions considering RCV increases.

Acknowledgments. This study was approved by the Institutional Review Board (IRB) at American University (Study ID: IRB-2023-189). We thank Jim Bisbee, Bernie Grofman, Diana Da In Lee, Ryan Moore, Carlisle Rainey, Matt Tyler, and four anonymous reviewers for their feedback. We also thank the attendees of the EPOVB Conference 2023, Japanese Society for Quantitative Political Science Winter Meeting 2024, PolMeth XL, KAIS 2023 Annual Meeting, Seoul National University's Statistics Department workshop, and Harvard University's applied statistics workshop. Our accompanying software `rankingQ` based on R is available at https://github.com/sysilviakim/rankingQ.

Author Contributions. Both authors contributed equally and are listed in alphabetical order.

Data Availability Statement. Replication materials can be found on the Political Analysis Harvard Dataverse for Atsusaka and Kim (2024) at https://doi.org/10.7910/DVN/UCTXEF. A copy of the same code and data can also be accessed at https://github.com/sysilviakim/ranking_error.

Supplementary Material. For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2024.24.

References

Agerberg, M., and A. -K. Kreft. 2023. "Sexual Violence, Gendered Protection and Support for Intervention." *Journal of Peace Research* 60 (5): 853–867.

Alvarez, R. M., T. E. Hall, and I. Levin. 2018. "Low-Information Voting: Evidence from Instant-Runoff Elections." *American Politics Research* 46 (6): 1012–1038.

Alvo, M., and P. L. H. Yu. 2014. *Statistical Methods for Ranking Data*. Springer New York, NY: Springer.

Alwin, D. F., and J. A. Krosnick. 1985. "The Measurement of Values in Surveys: A Comparison of Ratings and Rankings." *Public Opinion Quarterly* 49 (4): 535–552.

Atkeson, L. R., E. McKown-Dawson, J. Santucci, and K. L. Saunders. 2024. "The Impact of Voter Confusion in Ranked Choice Voting." *Social Science Quarterly*. 105 (4): 1029–1041.

Atsusaka, Y. 2021. "A Logical Model for Predicting Minority Representation: Application to Redistricting and Voting Rights Cases." *American Political Science Review* 115 (4): 1210–1225.

Atsusaka, Y. 2025. "Analyzing Ballot Order Effects When Voters Rank Candidates." *Political Analysis*. 33 (1): 64–72.

Atsusaka, Y., and S.-Y. S. Kim. 2024. "Replication Data for: Addressing Measurement Errors in Ranking Questions for the Social Sciences." https://doi.org/10.7910/DVN/UCTXEF.

Atsusaka, Y., and R. T. Stevenson. 2023. "A Bias-Corrected Estimator for the Crosswise Model with Inattentive Respondents." *Political Analysis* 31 (1): 134–148.

Berinsky, A. J., A. Frydman, M. F. Margolis, M. W. Sances, and D. C. Valerio. 2024. "Measuring Attentiveness in Self-Administered Surveys." *Public Opinion Quarterly* 88 (1): 214–241.

Berinsky, A. J., M. F. Margolis, and M. W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58 (3): 739–753.

Blasius, J. 2012. "Comparing Ranking Techniques in Web Surveys." *Field Methods* 24 (4): 382–398.

Ciuk, D. J. 2016. "Americans' Value Preferences Pre- and Post-9/11." *Social Science Quarterly* 97 (2): 407–417.

Clayton, K., Y. Horiuchi, Aaron R. Kaufman, Gary King, and M. Komisarchik. 2023. "Correcting Measurement Error Bias in Conjoint Survey Experiments."

Cormack, L. 2024. "More Choices, More problems? Ranked Choice Voting Errors in New York City." *American Politics Research* 52 (3): 306–319.

Costa, M. 2021. "Ideology, Not Affect: What Americans Want from Political Representation." *American Journal of Political Science* 65 (2): 342–358.

Dietrich, S. 2016. "Donor Political Economies and the Pursuit of Aid Effectiveness." *International Organization* 70 (1): 65–102.

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. John Wiley & Sons.

Genter, S., Y. G. Trejo, and E. Nichols. 2022. "Drag-and-Drop Versus Numeric Entry Options: A Comparison of Survey Ranking Questions in Qualtrics." *Journal of User Experience* 17 (3): 117–130.

Green, D. P., B. Palmquist, and E. Schickler. 2002. *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*. Yale University Press.

Gueorguiev, D., D. McDowell, and D. A. Steinberg. 2020. "The Impact of Economic Coercion on Public Opinion: The Case of US–China Currency Relations." *Journal of Conflict Resolution* 64 (9): 1555–1583.

Hopkins, D. J., C. Kaiser, and E. Perez. 2023. "The Surprising Stability of Asian Americans' and Latinos' Partisan Identities in the Early Trump Era." *The Journal of Politics* 85 (4): 1167–1580.

Horowitz, J. L., and C. F. Manski. 1995. "Identification and Robustness with Contaminated and Corrupted Data." *Econometrica* 63 (2): 281–302.

Huddy, L., and A. Bankert. 2017. "Political Partisanship as a Social Identity." In *Oxford Research Encyclopedia of Politics*. Oxford University Press.

Inglehart, R., and P. R. Abramson. 1994. "Economic Security and Value Change." *American Political Science Review* 88 (2): 336–354.

Jankowski, M., and J. Rehmert. 2022. "Selecting and Ranking Female Candidates under PR: Evidence from a Two-Stage Conjoint Experiment with Party Elites." *European Journal of Political Research*. doi:10.1111/1475-6765.12726.

Kane, J. V., and J. Barabas. 2019. "No Harm in Checking: Using Factual Manipulation Checks to Assess Attentiveness in Experiments." *American Journal of Political Science* 63 (1): 234–249.

Kane, J. V., Y. R. Velez, and J. Barabas. 2023. "Analyze the Attentive and Bypass Bias: Mock Vignette Checks in Survey Experiments." *Political Science Research and Methods* 11 (2): 293–310.

Kaufman, A. R., G. King, and M. Komisarchik. 2021. "How to Measure Legislative District Compactness If You Only Know It When You See It." *American Journal of Political Science* 65 (3): 533–550.

Krosnick, J. A. 1999. "Survey Research." *Annual Review of Psychology* 50 (1): 537–567.

Krosnick, J. A. and D. F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51 (2): 201–219.

Krosnick, J. A, and D. F Alwin. 1988. "A Test of the Form-Resistant Correlation Hypothesis: Ratings, Rankings, and the Measurement of Values." *Public Opinion Quarterly* 52 (4): 526–538.

Lee, T. 2009. "Between Social Theory and Social Science Practice. Toward a New Approach to the Survey Measurement of "Race." In: Abdelal R, Herrera YM, Johnston AI, McDermott R, eds. Measuring Identity: A Guide for Social Scientists. Cambridge University Press; 2009: 113–144.

Lopez, J. R. M. Alvarez, and S.-Y. S. Kim. 2022. "Latinos, Group Identity, and Equal Opportunity on the 2020 California Ballot." *Social Science Quarterly* 103 (7): 1572–1586.

Malhotra, N. 2009. "Order Effects in Complex and Simple Tasks." *Public Opinion Quarterly* 73 (1): 180–198.

Malhotra, N., and A. G. Kuo. 2008. "Attributing Blame: The Public's Response to Hurricane Katrina." *The Journal of Politics* 70 (1): 120–135.

Malhotra, N., and Y. Margalit. 2014. "Expectation Setting and Retrospective Voting." *The Journal of Politics* 76I (4): 1000–1016.

Marden, J. I. 1996. *Analyzing and Modeling Rank Data*. Chapman and Hall/CRC, New York: CRC Press.

McCarty, J. A., and L. J. Shrum. 2000. "The Measurement of Personal Values in Survey Research: A Test of Alternative Rating Procedures." *Public Opinion Quarterly* 64 (3): 271–298.

McCauley, J. F and D. N . Posner. 2019. "The Political Sources of Religious Identification: Evidence from the Burkina Faso–Côte d'Ivoire Border." *British Journal of Political Science* 49 (2): 421–441.

McMurry, N. 2022. "From Recognition to Integration: Indigenous Autonomy, State Authority, and National Identity in the Philippines." *American Political Science Review* 116 (2): 547–563.

Miles, W. F. S., and D. A. Rochefort. 1991. "Nationalism versus Ethnic Identity in Sub-Saharan Africa." *American Political Science Review* 85 (2): 393–403.

Neely, F., and C. Cook. 2008. "Whose Votes Count? Undervotes, Overvotes, and Ranking in San Francisco's Instant-Runoff Elections." *American Politics Research* 36 (4): 530–554.

Neely, F., and J. McDaniel. 2015. "Overvoting and the Equality of Voice Under Instant-Runoff Voting in San Francisco." *California Journal of Politics and Policy* 7 (4).

Nelson, T. E., R. A. Clawson, and Z. M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91 (3): 567–583.

Pettigrew, S., and D. Radley. 2023. "Ballot Marking Errors in Ranked-Choice Voting." DOI: 10.2139/ssrn.4670677

Pradel, F., J. Zilinsky, S. Kosmidis, and Y. Theocharis. 2024. "Toxic Speech and Limited Demand for Content Moderation on Social Media." *American Political Science Review* 118 (4): 1895–1912.

Rathbun, B., N. S. Rathbun, and C. Pomeroy. 2022. "No Fair! Distinguishing Between the Pursuit of Status and Equity in International Relations." *International Studies Quarterly* 66 (1): sqac002.

Rathbun, B. C., and C. Pomeroy. 2022. "See No Evil, Speak No Evil? Morality, Evolutionary Psychology, and the Nature of International Relations." *International Organization* 76 (3): 656–689.

Revilla, M., and M. P. Couper. 2018. "Testing Different Rank Order Question Layouts for PC and Smartphone Respondents." *International Journal of Social Research Methodology* 21 (6): 695–712.

Serenko, A., and N. Bontis. 2013. "First in, Best Dressed: The Presence of Order-Effect Bias in Journal Ranking Surveys." *Journal of Informetrics* 7 (1): 138–144.

Setzler, M., and A. B. Yanus. 2018. "Why Did Women Vote for Donald Trump?" *PS: Political Science & Politics* 51 (3): 523–527.

Shugart, M. S., and R. Taagepera. 2017. *Votes from Seats: Logical Models of Electoral Systems*. Cambridge University Press.

Smyth, J. D., K. Olson, and A. Burke. 2018. "Comparing Survey Ranking Question Formats in Mail Surveys." *International Journal of Market Research* 60 (5), 502–516.

Spry, A. D. 2017. "Identity and Political Attitudes: What Insights can an Alternative Measurement Strategy Provide?" *Newsletter of the Comparative Politics Organized Section of the American Political Science Association* 27 (2).

Spry, A. D. 2021. "Identity Experiments: Design Challenges and Opportunities for Studying Race and Ethnic Politics." In *Advances in Experimental Political Science*, edited by J. N. Druckman and D. P. Green, 431–446. Cambridge University Press.

Tate, K. 2004. *Black Faces in the Mirror: African Americans and Their Representatives in the U.S. Congress. Princeton University Press*. Princeton, New Jersey.

Tomz, M., J. Wittenberg, and G. King. 2003. "Clarify: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* 8 (1): 1–30.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Tranter, B., and M. Western. 2010. "Overstating Value Change: Question Ordering in the Postmaterial Values Index." *European Sociological Review* 26 (5): 571–583.

Tyler, M. J. Grimmer, and S. J. Westwood. 2024. "A Statistical Framework to Engage the Problem of Disengaged Survey Respondents: Measuring Public Support for Partisan Violence." https://doi.org/10.2139/ssrn.4894008

West, E. A., and S. Iyengar. 2022. "Partisanship as a Social Identity: Implications for Polarization." *Political Behavior* 44 (2): 807–838.

Cite this article: Atsusaka, Y. and Kim, S. Silvia (2025). Addressing Measurement Errors in Ranking Questions for the Social Sciences. *Political Analysis*, 1–22. https://doi.org/10.1017/pan.2024.33