

1 Nonresponse and Sample Weighting in Organizational Surveying

2 John T. Kulas¹, Yang Yang², and & David H. Robinson³

3 ¹ Montclair State University

4 ² China Select

5 ³ St. Cloud State University

6 Author Note

7

8 Correspondence concerning this article should be addressed to John T. Kulas,
9 Dickson Hall #250, Montclair State University, Montclair, NJ, 07043. E-mail:
10 kulasj@montclair.edu

11

Abstract

12 Post-stratification weighting is a common procedure used in public opinion polling
13 applications to correct demographic constituency differences between samples and
14 populations. Although common practice in public opinion polling, this form of data
15 remediation is only recently emergent as a procedure of interest in organizational surveying
16 applications. The current paper induces survey nonresponse via data simulation across
17 fictional constituent groups (e.g., organizational strata) and documents the impact of
18 weighting on the accuracy of sample estimates. Our goal was to evaluate the effectiveness
19 of the weighting algorithm when confronted with *passive* and *active* forms of nonresponse
20 in an effort to: 1) interject this nonresponse taxonomy within the broader weighting
21 domain, while 2) exploring the organizationally-relevant sampling scenarios that are either
22 benefit, “hurt”, or effectively immune to post-stratification weighting. The results confirm
23 that sampling contexts characterized by active nonresponse did benefit from application of
24 sample weights, but only when accompanied by constituency differences in underlying
25 population construct (e.g., surveyed *attitude*) standing. Alternatively, constituent member
26 differences in population attitudes, when characterized by passive forms of nonresponse,
27 exhibited no benefit from weighting (in fact these scenarios are somewhat *hurt* by
28 weighting). The simulations reinforce that, moving forward, it would be prudent for
29 surveyors of all disciplinary backgrounds to mutually attend to the traditional focus of
30 both traditions: public opinion polling (e.g., post-stratification adjustment) and
31 organizational surveying (e.g., *form* of nonresponse).

32

Keywords: Survey methodology, sample weighting, nonresponse, response rate

33 Nonresponse and Sample Weighting in Organizational Surveying

34 Akin to differential variable weighting (for instance: a) construct indicators within

35 an assessment scale [aka factor loadings], or b) predictors within a selection system [aka

36 regression weights]; e.g., per data matrix “columns”), sample weighting alters the

37 proportional contributions of *individual respondents* within a data set (e.g., matrix rows).

38 Some respondents are assigned greater relative impact and others are assigned less. This

39 practice is commonplace in the summary of general population polling data reflecting, for

40 example, elections and politics (e.g., Rivers & Bailey, 2009), prevalence rates of

41 psychological disorders (e.g., Kessler et al., 2009), or feelings of physical safety (e.g., Quine

42 & Morrell, 2008). It is also seemingly in the nascent stages of awareness and application

43 within the organizational surveying domain (see, for example, Kulas et al., 2016; Landers

44 & Behrend, 2015; Tett et al., 2014).

45 We speculate that this form of statistical remediation is gaining some interest in the

46 organizational surveying domain, at least in part, because industrial psychologists are

47 keenly aware that response rates within organizational surveying applications are trending

48 downward (see, for example, Anseel et al., 2010; Rogelberg & Stanton, 2007). With lower

49 response rates, surveyors are confronted with heightened levels of scrutiny because,

50 historically, a locally realized high response rate has been interpreted as a positive

51 indicator of data quality - if not from the survey specialists themselves, at least from client

52 stakeholders (e.g., Anseel et al., 2010; Cycyota & Harrison, 2002, 2006; Frohlich, 2002).

53 The orientation of this paper, however, is that although response rate is a

54 commonly referenced proxy of survey quality, it is not response rate but rather sample

55 *representativeness* that should be the primary focus of concern for survey specialists (see,

56 for example, Cook et al., 2000; Krosnick, 1999). Representativeness can of course be “hurt”

57 by low response rates, but the relationship between these two survey concepts is by no

58 means exact (e.g., Curtin et al., 2000; Keeter et al., 2006; Kulas et al., 2017). Stated

59 differently, a low response rate is neither a sufficient nor even necessary condition for
60 sample misrepresentation.¹

61 In the context of any survey application, sample misrepresentation ultimately refers
62 to a discrepancy between estimated sample statistics and population parameters. Ideally,
63 such discrepancies arise from completely random sources (in which case resulting error is
64 less likely to be reasonably characterized as *bias*). In reality, however, discrepancies are not
65 only driven by purely random causes. There are several broader sampling methodology
66 factors that may be systematically driving the relative under- or over-selection of a
67 population segment (see, for example, Kulas et al., 2016), but the most commonly cited
68 contributor within the organizational sciences is non-response (e.g., invited individuals
69 simply either forget [e.g., passive nonresponse] or consciously choose not to participate in
70 the survey process [e.g., active nonresponse], see, for example, Rogelberg et al., 2000). Our
71 presentation also focuses on this non-response contributor to sample misrepresentation, but
72 only because we aim to: 1) integrate the organizational non-response and
73 post-stratification weighting literatures, while also 2) highlighting the associations and
74 dissociations between response rate and bias (although we note here that the current
75 presentation and procedure also inform other sampling methodological sources of
76 misrepresentation than non-response).²

¹ There are indisputable benefits associated with higher response rates (such as greater statistical *power*), although this benefit does not stem directly from response rate, but rather its correlate - larger *n*. Our presentation reflects the fact that greater power (and/or, relatedly, smaller confidence intervals) may in fact introduce a *false sense* of methodological superiority when the sample misrepresents the population. Primarily for this reason, we stress that the sampling concepts of response rate, sample size, and power need to be fully disentangled from the methodological concept of representativeness, and the importance of this dissociation drives the central theme of the current paper.

² Frequently presented as a separate consideration, *measurement error* is an additional contributor to sample misrepresentation. The current focus is on deviations from a perfect sampling methodology as opposed to deviations from an ideal psychometric methodology. We do however note that future

77 Nonresponse in Organizational Surveying

78 Within the organizational surveying domain, it is not uncommon for response rate
79 to be referenced as a proxy for survey data quality (see, for example, Baruch & Holtom,
80 2008; Fan & Yan, 2010; Pedersen & Nielsen, 2016). Baruch (1999), for example, states
81 that, “... to have dependable, valid, and reliable results, we need a high RR from a wide
82 representation of the whole population under study” and that, “The level of RR is an
83 important, sometimes crucial, issue in relation to the validity of a paper’s results” (p. 422).
84 Fan and Yan (2010) similarly state that response rate is, “... the most widely used and
85 commonly computed statistic to indicate the quality of surveys” (p. 132). Pedersen and
86 Nielsen (2016) claim that a high survey response rate, “... diminishes sampling bias
87 concerns and promotes the validity of survey-based research findings” (p. 230). The general
88 consensus seems to be that there are three major (negative) consequences of low response
89 rates, including (a) yielding smaller sample size, which negatively impacts statistical power
90 and confidence intervals, (b) reducing the credibility of survey data, and (c) generating
91 biased samples that impair the generalizability of survey results (Biemer & Lyberg, 2003;
92 Luong & Rogelberg, 1998; Rogelberg et al., 2000).

93 To the likely frustration of those who associate response rate with survey data
94 quality, organizational survey response rates have been declining for decades. Baruch
95 (1999), for example, summarized response rates of 175 studies published in five leading
96 management and behavioral sciences journals in 1975, 1985, and 1995. His results revealed
97 an average response rate (across time periods) of 55.6% ($SD = 19.7\%$), but also a trend
98 within which response rates declined steadily from 64.4% to 55.7% to 48.4% over the three
99 time points. Nine years later, Baruch and Holtom (2008) conducted a follow-up study of
100 1,607 studies published in 17 disciplinary-relevant journals in 2000 and 2005 but found no

advancement of current representations of survey error would benefit from a unified perspective that encompasses error arising from both sources: measurement and sampling strategy.

101 substantial differences in response rates compared to those in 1995, suggesting that the
102 declining trend had perhaps reached a lower asymptote. However, a different approach
103 with similar goals (Anseel et al., 2010) analyzed 2,037 survey projects published in 12
104 journals in Industrial and Organizational Psychology, Management, and Marketing from
105 1995 to 2008 and did note a slight decline (overall $M = 52.3\%$) when controlling for the use
106 of response enhancing techniques.³

107 ***Form of Nonresponse***

108 Although high response rates are generally pursued as a desirable goal within
109 organizational surveying applications, there has also been a broad acknowledgement that
110 not all forms of nonresponse should be considered equally worrisome. Rogelberg et al.
111 (2003), for example, propose a distinction between *active* and *passive* nonrespondents
112 based on intent and (in)action. According to Rogelberg et al. (2003), active
113 nonrespondents are those who intentionally refuse to participate in surveys, while passive
114 nonrespondents are those who fail to respond to surveys due to reasons such as forgetting
115 or misplacing invitations. Passive nonrespondents are thought to be similar to respondents
116 in both attitude (Rogelberg et al., 2003) as well as organizational citizenship behaviors
117 [OCBs; Spitzmüller et al. (2007)], whereas active nonrespondents have been shown to
118 exhibit significantly lower organizational commitment and satisfaction, higher intention to
119 leave, lower conscientiousness, and lower OCBs than actual respondents (Rogelberg et al.,
120 2000, 2003; Spitzmüller et al., 2007).

121 The more commonly encountered form of organizational nonresponse appears to be
122 passive (Rogelberg et al., 2003; e.g., Rogelberg & Stanton, 2007), although subgroup rates

³ It is possible that the declination has stabilized with mean response rates hovering around 50% after roughly the turn of the millennium ($M = 52.5\%$ for HRM studies from 2009 to 2013, Mellahi & Harris, 2016; $M = 52.0\%$ for management studies from 2000 to 2004, Werner et al., 2007). This stability, if authentic, may again possibly be accounted for by an increased contemporary emphasis on response enhancing strategies (Anseel et al., 2010; Fulton, 2016).

123 may evidence variability - men, for example, have a higher proclivity toward active
124 nonresponse than do women (Luong & Rogelberg, 1998; Rogelberg et al., 2000; Spitzmüller
125 et al., 2007). Additionally, it has been noted that selection of an individual population
126 element into a realized sample is often predictable [because of, for example, an increased
127 likelihood of not responding when dissatisfied or disgruntled; Taris and Schreurs (2007)].
128 The organizational surveying expectation is that, *on average*, roughly 15% of
129 nonrespondents can be expected to be accurately characterized as “active” (Rogelberg et
130 al., 2003; Rogelberg & Stanton, 2007; Werner et al., 2007). It is this second, less frequently
131 anticipated form of nonresponse that also carries the greater corresponding threat of biased
132 sample estimates (see, for example, Kulas et al., 2017; Rogelberg & Stanton, 2007).

133 **Sample Weighting - a Brief Overview**

134 Within public opinion polling contexts, when realized sample constituencies (e.g.,
135 44% male - by tradition from *carefully-constructed* and *randomly sampled* data frames)⁴
136 are compared against census estimates of population parameters (e.g., 49% male), weights
137 are applied to the realized sample in an effort to remediate the relative proportional under-
138 or over-sampling. This is because, if the broader populations from which the under- or
139 over-represented groups are sampled differ along surveyed dimensions (e.g., males, within
140 the population, are *less likely to vote for Candidate X* than are women), then unweighted
141 aggregate statistics (of, for example, projected voting results) will misrepresent the true

⁴ These important sampling elements are very carefully attended to within public opinion polling contexts. Conversely, within organizational surveying traditions, these considerations are not commonly acknowledged, at least explicitly. The weighting procedure presented in the current manuscript remediates bias regardless of full methodological consideration of sampling context, but is dependent on accurate “census” population constituency estimates (and, as the results highlight, the presence of an active nonrespondent group). For the interested reader, an acknowledgement of the broader methodological sampling scenario, and the additional potential sources of error, facilitates a much deeper appreciation and understanding of the benefits and potential pitfalls of sample weighting.

142 population parameter. This remedial application of sample weights should also be
 143 considered an option for organizational researchers pursuing answers to similar survey
 144 questions such as: “What is the mood of the employees?” This is because focused queries
 145 such as this are deceptively complex - implicit in the question is a focus not on survey
 146 respondents, but rather the broader employee population. Acknowledging this implicit
 147 target is important, because the next step (after gauging the mood of the surveyed
 148 respondents) is *doing something* about it. Weighting is one remedial option for
 149 organizational surveyors to plausibly transition from, “What do the survey results say”? to
 150 “What do the employees feel”?

151 **Procedural application**

152 *Proportional weights* are the form of weights most directly relevant to organizational
 153 surveying applications that traditionally focus on nonresponse as the primary contributor
 154 to sample misrepresentation. These weights are ratios of the proportion of a population
 155 within a given stratum to the proportion of the sample within that same stratum:

$$\text{proportional weight}(\pi_k) = \frac{(N_k/N)}{(n_k/n)} \quad (1)$$

156 Over-sampling of elements of a stratum (k) results in proportional weights less than
 157 one, while under-sampling (relative to the population) results in proportional weights
 158 greater than one. The common procedure for weight estimation *when more than one*
 159 *stratum is specified* is an iterative process that may be referred to by multiple synonymous
 160 terms: *rim weighting*, *iterative proportional fitting*, or *raking* (see, for example, Deming &
 161 Stephan, 1940). Regardless of label, the procedure guides the surveyor to:

- 162 1) Determine proportional weights for all levels within one stratum, and then assign
 163 these weights to cases.
 164 2) Determine proportional weights for a second group (ratio of population percent to

165 *current* sample percent [the current sample percent will be affected by the step 1
166 weighting procedure]). Multiply previous (step 1) weights by the proportional
167 weights for this second stratum and assign these new weights to cases.

- 168 3) Determine proportional weights for a third stratum (which will once again require
169 re-inspection of the *current* sample percent). Multiply the previous step 2 weights by
170 the third stratum proportional weights and assign to cases.
- 171 4) Repeat steps 1, 2, and 3 (or more if more than three groups/strata are considered) in
172 sequence until the weighted sample characteristics closely match the population
173 characteristics.

174 Possible strata relevant for organizational survey weighting include: branch, full-,
175 part-, or flex-time status, functional area, gender, geographic location, hierarchy, salaried
176 status, subsidiary, tenure, work shift, or any other groupings especially deemed suspect to
177 possess a relatively disproportionate number of active nonrespondents (through application
178 of forecasting strategies such as those advocated by, for example, Rogelberg and Stanton,
179 2007). Each of these strata may of course also be the targeted focus of survey results
180 feedback, but when *aggregating* results across (or even within) strata, a consideration of the
181 impact of nonresponse may yield more accurate survey estimates. The explicit goal is a
182 closer approximation of sample characteristics to population parameters via statistical
183 remediation, and drives the current paper's focus on the interplay of four survey concepts
184 (distribution of attitude within the larger population, response rate, nonresponse form, and
185 remedial weighting):

186 after chatting with Yang (10/31/19) these need to be clarified a bit - reading
187 11/3 they make sense but need to be read very carefully. Check with Yang on
188 1/26 Skype. 2/1 revisions seem ok to Kulas. Three moving parts: underlying
189 attitudinal distributions, response rate, and form of nonresponse <- perhaps we
190 should make these variables more explicit prior to the procedure/results...

191 *Research question 1:* What role does overall response *rate* play in sample

192 misrepresentation? **[make sure this is reflected in results]**

193 *Research question 2:* What role does nonresponse *form* (passive versus active) play

194 in sample misrepresentation? **currently in paper as figures 1-3**

195 *Research question 3:* What impact does the application of weights have on both

196 biased (e.g., misrepresentative) and unbiased sample estimates?

197 *Research question 4:* What is the role of response rate and form in the *effectiveness*

198 of weighting? **[perhaps David can derive/find a proof to parallel our results?]**

199 We view these questions as being analogous to similar questions asked and answered

200 with differential variable weighting within the applied Psychology discipline. Just as, for

201 example, there has been debate regarding the merits of differential versus unit variable

202 weighting in a selection context (e.g., Wainer, 1976) or simple composite score aggregate

203 (Bobko et al., 2007), we propose that a similar consideration is appropriate with persons,

204 and therefore compare and contrast unit- versus variable-sample element weighting via

205 carefully controlled data simulation.

206 **Methods**

207 We address our research questions via data simulation within the broad fictional

208 context of organizational surveying (assessing, for example, attitudinal estimates of

209 employee satisfaction, engagement, or organizational commitment). We began the

210 simulations by establishing “populations”, each consisting of 10,000 respondents

211 characterized by demographic categorizations across gender (male and female) and

212 department (A and B). We therefore had four demographic groups (male-A, male-B,

213 female-A, and female-B). For these population respondents, we generated scaled continuous

214 responses (real numbers) ranging from values of 1 to 5, reflecting averaged aggregate scale

215 scores from a multi-item survey with a typical $1 \rightarrow 5$ Likert-type or graphic rating scale

216 response format.

217 In order to represent different proportions of relative constituency (for example,
218 more females than males or more department A workers than department B), we iterated
219 population characteristics at marginal levels (gender and department) starting at 20% (and
220 80%) with increments and corresponding decrements of 20%. For example, if males
221 accounted for 20% of the simulated population, then females were 80%; also if respondents
222 in Department A represented 60% of a population, then 40% were in Department B.

223 Marginal constituencies were therefore specified at all combinations (across the two
224 variables) of 20% and 80%, 40% and 60%, 60% and 40%, and 80% and 20%. This resulted
225 in population *cell* constituencies (e.g., men in department A) as low as 400 and as high as
226 6,400.

227 Additionally, each of these cell populations was characterized by an attitude
228 distribution in one of three different possible forms: normal, positively skewed, or
229 negatively skewed. These distributional forms were specified in an attempt to model
230 similarities and discrepancies in construct standing (e.g., commitment, satisfaction, or
231 engagement) across respondent groupings. The normal distribution exhibited, on average,
232 a mean of 3.0 whereas the skewed distributions were characterized by average means of 2.0
233 and 4.0, respectively. In total, eight crossings of distributional type across employee
234 categorization were specified (Table 1 presents the combinations of these distributions).

235 Note that these eight conditions are not exhaustive across our four cell groupings - we
236 specified combinations that we expected to be most informative across our passive to active
237 nonresponse continuum (reflected in Table 1's "anticipated bias" column).

238 Individual attitudes were randomly sampled from population distributions at the
239 cell level (e.g., Department A Males) without replacement. Response rates
240 (methodologically these could also be conceptualized as *sampling* rates) were controlled at
241 the marginal level using 10% increments ranging from 60% to 90%, and these were fully
242 iterated. Our cell-level response rates therefore ranged from 36% to 81% - a range of rates
243 chosen because they are, according to the organizational surveying literature, reasonable

Table 1*Attitudinal Distribution Conditions Specified in Current Paper*

| Condition | Distributional Shape | mu | Male | | Female | | Anticipated Bias |
|-----------|----------------------|----|--------|--------|--------|--------|------------------|
| | | | Dept A | Dept B | Dept A | Dept B | |
| 1 | Normal | 3 | X | X | X | X | None |
| | Positive Skew | 2 | | | | | |
| | Negative Skew | 4 | | | | | |
| 2 | Normal | 3 | | | | | None |
| | Positive Skew | 2 | X | X | X | X | |
| | Negative Skew | 4 | | | | | |
| 3 | Normal | 3 | | | | | None |
| | Positive Skew | 2 | | | | | |
| | Negative Skew | 4 | X | X | X | X | |
| 4 | Normal | 3 | X | X | X | | Moderate |
| | Positive Skew | 2 | | | | | |
| | Negative Skew | 4 | | | | X | |
| 5 | Normal | 3 | X | X | | | Moderate/High |
| | Positive Skew | 2 | | | X | X | |
| | Negative Skew | 4 | | | | | |
| 6 | Normal | 3 | | X | X | | Moderate/High |
| | Positive Skew | 2 | X | | | | |
| | Negative Skew | 4 | | | | X | |
| 7 | Normal | 3 | | | | | High |
| | Positive Skew | 2 | X | | X | | |
| | Negative Skew | 4 | | X | | X | |
| 8 | Normal | 3 | | | | | High |
| | Positive Skew | 2 | X | X | | | |
| | Negative Skew | 4 | | | X | X | |

²⁴⁴ expectations (e.g., Mellahi & Harris, 2016; Werner et al., 2007). We therefore investigated
²⁴⁵ error within the aggregate mean (e.g., grand mean or total sample mean) attributable to
²⁴⁶ different likelihoods of sample inclusion from constituent groups of different relative size
²⁴⁷ and representing populations of different attitudinal distribution, but at response rates
²⁴⁸ reasonably expected to exist in real-world organizational surveying contexts.

²⁴⁹ It should be noted here that there are several collective patterns of response that
²⁵⁰ are intended to represent sampling scenarios exhibiting *passive* nonresponse, regardless of
²⁵¹ absolute response rate: all subgroups exhibiting the same response rate (e.g., 36%, 36%,
²⁵² 36%, and 36%). All other combinations of response rate are intended operationalizations of
²⁵³ active forms of nonresponse (e.g., *not* reasonably characterized as missing at random,
²⁵⁴ NMAR), although the degree to which a sampling scenario should be reasonably
²⁵⁵ characterized as exhibiting active nonresponse is intended to be incremental across
²⁵⁶ iterations.

Table 2

Example Summarized Response Rate Conditions Represented in Figures 2 through 5

| Example Response Rates (Any Combination) | | | | | | | Number of Conditions | Form (and degree) of Nonresponse |
|--|-------------|---------------|---------------|----------|-----|---------|----------------------|----------------------------------|
| Male Dept A | Male Dept B | Female Dept A | Female Dept B | SD Index | 256 | Passive | | |
| 36% | 36% | 36% | 36% | .000 | 256 | Passive | | |
| 36% | 36% | 42% | 42% | .034 | 128 | | | |
| 48% | 48% | 54% | 54% | .035 | 64 | | | |
| 42% | 42% | 49% | 49% | .040 | 192 | | | |
| 48% | 48% | 56% | 56% | .046 | 128 | | | |
| 56% | 56% | 64% | 64% | .047 | 64 | | | |
| 54% | 54% | 63% | 63% | .051 | 128 | | | |
| 63% | 63% | 72% | 72% | .052 | 64 | | | |
| 36% | 42% | 42% | 49% | .053 | 64 | | | |
| 42% | 48% | 49% | 56% | .057 | 128 | | | |
| 49% | 56% | 56% | 64% | .061 | 64 | | | |
| 48% | 54% | 56% | 63% | .062 | 128 | | | |
| 56% | 63% | 64% | 72% | .066 | 128 | | | |
| 36% | 36% | 48% | 48% | .069 | 128 | | | |
| 64% | 72% | 72% | 81% | .069 | 64 | | | |
| 42% | 42% | 56% | 56% | .081 | 128 | | | |

Table 2 continued

Example Response Rates (Any Combination)

| Male Dept A | Male Dept B | Female Dept A | Female Dept B | SD Index | Number of Conditions | Form (and degree) of Nonresponse |
|-------------|-------------|---------------|---------------|----------|----------------------|----------------------------------|
| 36% | 42% | 48% | 56% | .085 | 128 | |
| 42% | 49% | 54% | 63% | .089 | 128 | |
| 48% | 48% | 64% | 64% | .092 | 128 | |
| 42% | 48% | 56% | 64% | .096 | 128 | |
| 49% | 56% | 63% | 72% | .098 | 128 | |
| 36% | 36% | 54% | 54% | .104 | 192 | |
| 48% | 54% | 64% | 72% | .106 | 128 | |
| 56% | 63% | 72% | 81% | .109 | 128 | |
| 36% | 48% | 48% | 64% | .115 | 64 | |
| 36% | 42% | 54% | 63% | .120 | 128 | |
| 42% | 42% | 63% | 63% | .121 | 64 | |
| 42% | 54% | 56% | 72% | .123 | 128 | |
| 49% | 63% | 63% | 81% | .131 | 64 | |
| 42% | 48% | 63% | 72% | .137 | 128 | |
| 48% | 48% | 72% | 72% | .139 | 64 | |
| 36% | 48% | 54% | 72% | .150 | 128 | |

Table 2 continued

| Example Response Rates (Any Combination) | | | | | | |
|--|-------------|---------------|---------------|----------|----------------------|----------------------------------|
| Male Dept A | Male Dept B | Female Dept A | Female Dept B | SD Index | Number of Conditions | Form (and degree) of Nonresponse |
| 48% | 54% | 72% | 81% | .154 | 128 | |
| 54% | 54% | 81% | 81% | .156 | 64 | |
| 42% | 54% | 63% | 81% | .164 | 128 | |
| 36% | 54% | 54% | 81% | .186 | 64 | Active |

In an attempt to capture this “degree of active nonresponse”, we calculated a simple index of response rate discrepancy (SD; presented in Table 2). The “least” active nonresponse scenarios are characterized by two subgroups with identical response rates and two having a slightly different response rate (e.g., Dept A Males = 36%, Dept A Females = 36%, Dept B Males = 42%, and Dept B Females = 42%; see the second row of Table 2, the SD index = .034)⁵. Also here note that three of our eight Table 1 conditions represent scenarios where the presence of active nonrespondents is not expected to result in bias (e.g., regardless of patterns of nonresponse, the unweighted sample mean is expected to yield an unbiased estimate of the population mean). These are Table 1 conditions one through three, where attitudinal distributions are of *the same form* across groups, regardless of any individual group response rate discrepancy from others’.

These operationalizations of passive and active forms of nonresponse differ from other investigations with similar-minded approaches. Kulas et al. (2017), for example, directly tie probabilities of sample inclusion to an individual’s held attitude (the likelihood of sample inclusion is fully dependent on the population member’s attitude). With the current investigation, conversely, the probability of sample inclusion is dependent only on group membership (with some of these groups occasionally being characterized by unique attitude distributional forms). Essentially, Kulas et al. (2017) operationalize active nonresponse at the person-level whereas the current paper does so at the group level. This may be a more practical operationalization, as organizational surveyors are more likely to have an inclination of a group’s collective attitude or likelihood to respond (e.g., night shift

⁵ This method of simplifying the presentation of our response rate conditions is fully orthogonal to population constituency and distributional form. That is, the amount of bias present in a sample estimate is expected to be quite different for Condition 7 with response rates of 48%, 48%, 72%, 72% versus 48%, 72%, 48%, 72%, even though the crude response rate index (SD = 0.139) is the same for both scenarios. There is additional information within these simulations (the effect of a *combination* of response rate and population form on degree of bias) that is therefore not captured via this simple SD index.

278 workers, machine operators) than they are of any one individual employee.

279 **Results**

280 *Research question 1:* What role does overall response *rate* play in sample
281 misrepresentation? [make sure this is reflected in results]

282 A couple paragraphs to answer RQ1

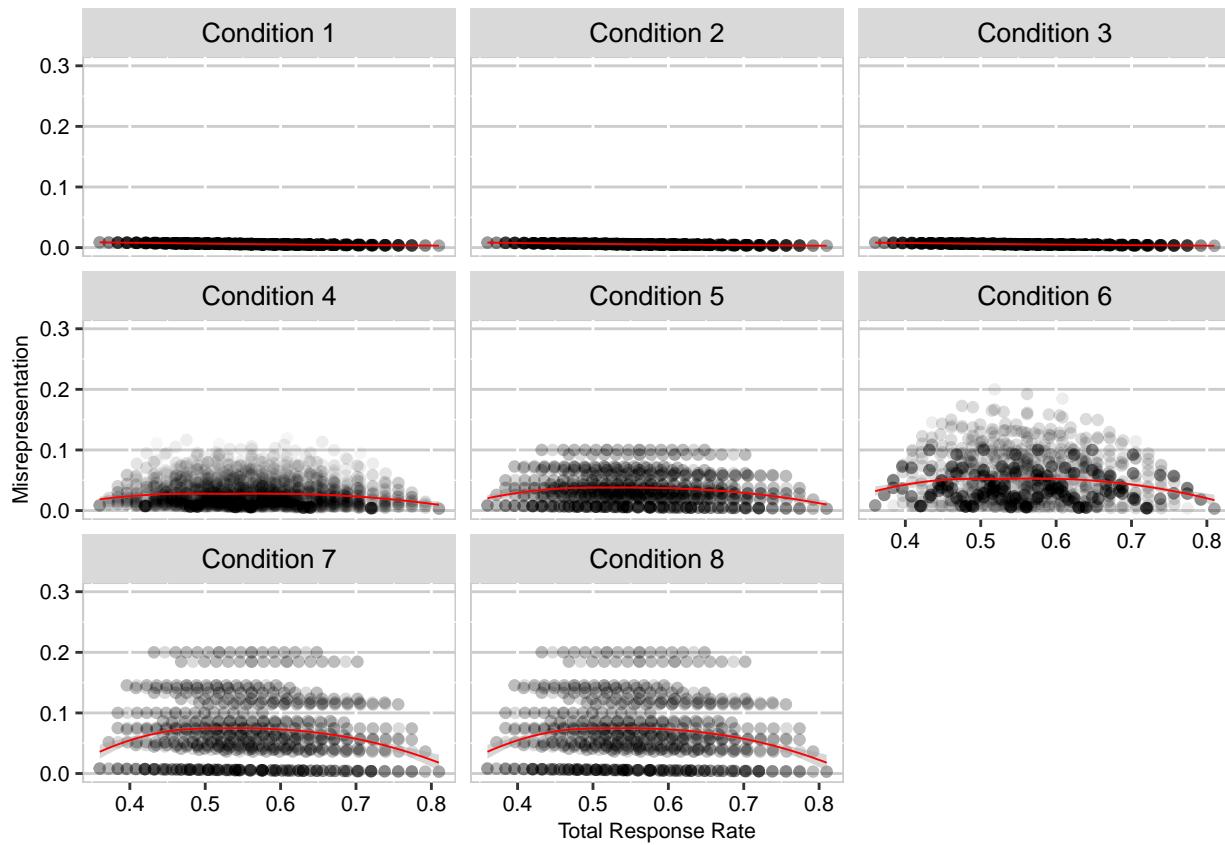
283 Have to operationalize “sample misrepresentation” first

284 The following is RQ2:

285 In total, we generated 327.68 million samples (4,096 unique combinations of
286 response rate and population constituency across gender and department, simulated 10,000
287 times each across our eight Table 1 conditions). Each of these samples was comprised of,
288 on average, $n = r$ format(mean(combo[“NS”]), big.mark=“,”, digits=0,scientific=FALSE),
289 collectively representing an experiment-wide n of 1.8432 trillion. For each individual
290 simulation, weights were applied iteratively to the data at the two marginal (variable)
291 levels via raking, and were estimated via the *anesrake* package (Pasek, 2016) in *R* version
292 3.31 (R Core Team, 2017). We were most interested in comparing the extent to which
293 unweighted (aggregated responses without raking) and weighted (aggregated weighted
294 responses) sample means approximated the population means across our controlled
295 specifications of response rate, nonresponse form, and attitudinal distribution (population
296 means were taken from each iteration, as the simulations specified a new population at
297 each iteration). The “effectiveness” of weighting was evaluated by calculating the
298 discrepancies between the population and both weighted and unweighted sample means as
299 well as the averaged deviations of these discrepancies from the population mean
300 (discrepancy in the “mean” of the means is bias, dispersion about the “mean” of the means
301 is error). If the average weighted sample mean was closer to the true population mean,
302 relative to the unweighted one, then the weighting was deemed beneficial.

303 Add a couple of paragraphs here to answer research questions 1(a) and 1(b)

304 Correlation coefficient needed.[Yang to calculate 2/1]



305 To partially address the second limitation, discrepancy between population

306 constituency and sampling proportions was additionally estimated via Cattell's profile
 307 similarity index [r_p ; Cattell et al. (1966)]. r_p is sensitive to discrepancies in profile shape
 308 (pattern across profile components), elevation (average component score), and scatter (sum
 309 of individual components' deviation from the elevation estimate. Figure 3 demonstrates the
 310 pattern of unweighted sample mean deviation (from the population parameter) when this
 311 index is taken into consideration. edits....gain demonstrate these relationships across the
 312 attitudinal form conditions, being grouped by underlying distributions thought to be
 313 susceptible to bias (Conditions 3 through 8) as well as those thought to be relatively
 314 immune to bias (Conditions 1 through 3; aka those sampling situations in which weighting
 315 is unnecessary).

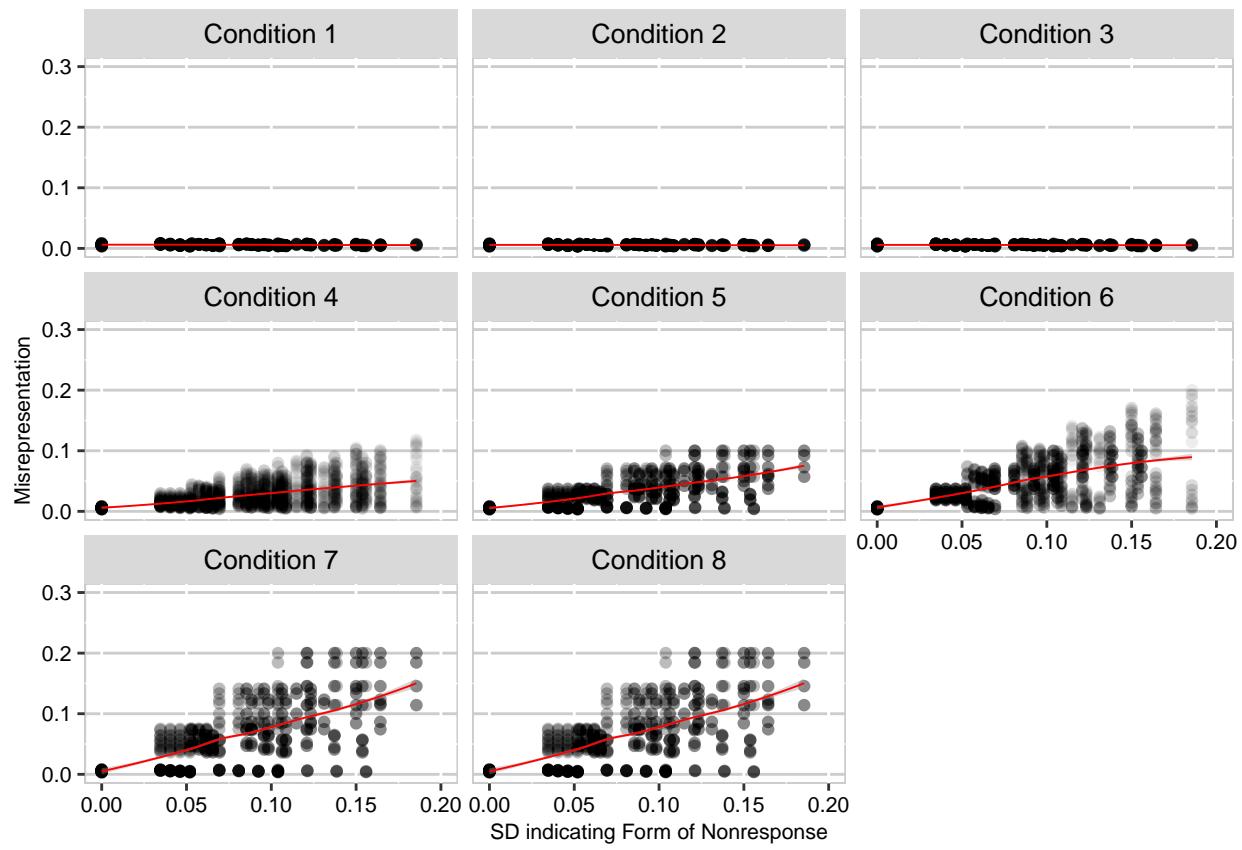


Figure 1

Relationship between nonresponse form and misrepresentation.

317 The plurality of our findings are presented visually, and they focus on the overall

318 mean (e.g., the average rating across all sample members). Figure 1 provides a broad

319 summary of the results across the eight different attitudinal distribution conditions,

320 presenting the average absolute discrepancy from the population mean within each broad

321 condition. Conditions one through three demonstrate that, on average, the unweighted

322 sample mean provides a good (unbiased) estimate of the population mean when the

323 distributional form is held constant across constituent groups (e.g., the distributions of

324 attitudes are of similar functional forms and locations for all constituent groups). This is

325 regardless of form or extent of nonresponse. Additionally, weighting remediates deviations

326 about the true mean in all five attitudinally discrepant conditions, even when considerable

327 error exists in the unweighted estimate (e.g., the rightmost bars in Figure 1).

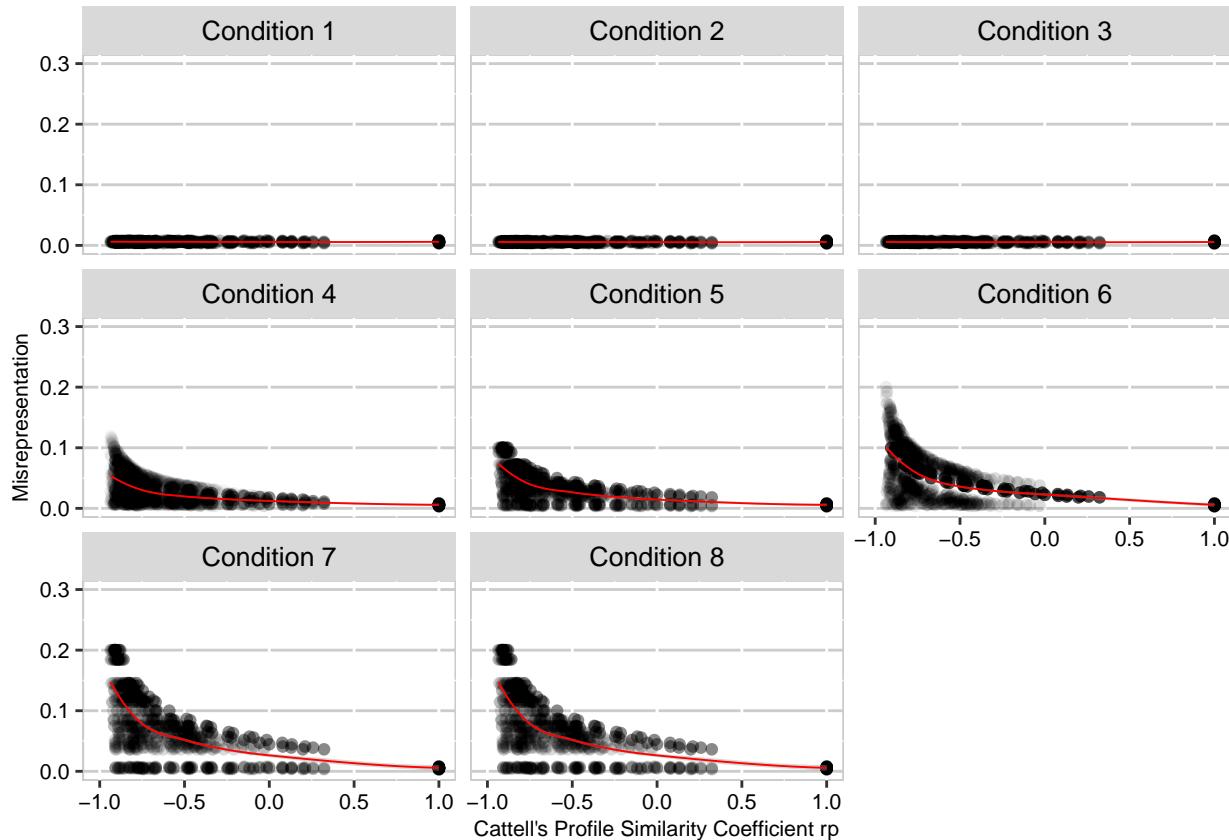


Figure 2

Relationship between sample representativeness and misrepresentation.

328

The Role of Response Rate

329 In terms of explaining the very little error that did emerge within the passive
 330 nonresponse conditions, this error was entirely attributable to response rate (See Figure 2).
 331 The nature of the exact relationship was slightly nonlinear, being fit with quadratic
 332 functions within each condition (collapsing across conditions did exhibit slight within-array
 333 differences [which would affect the statistically perfect relationship]).

334 **Need to Recall Research Questions in appropriate sections**

335 Figure 3 demonstrates how the weighting algorithm operated across conditions one
 336 through three taking form of nonresponse into consideration (along the x-axis, with passive
 337 nonresponse occupying the left of the figure and active nonresponse scenarios occupying

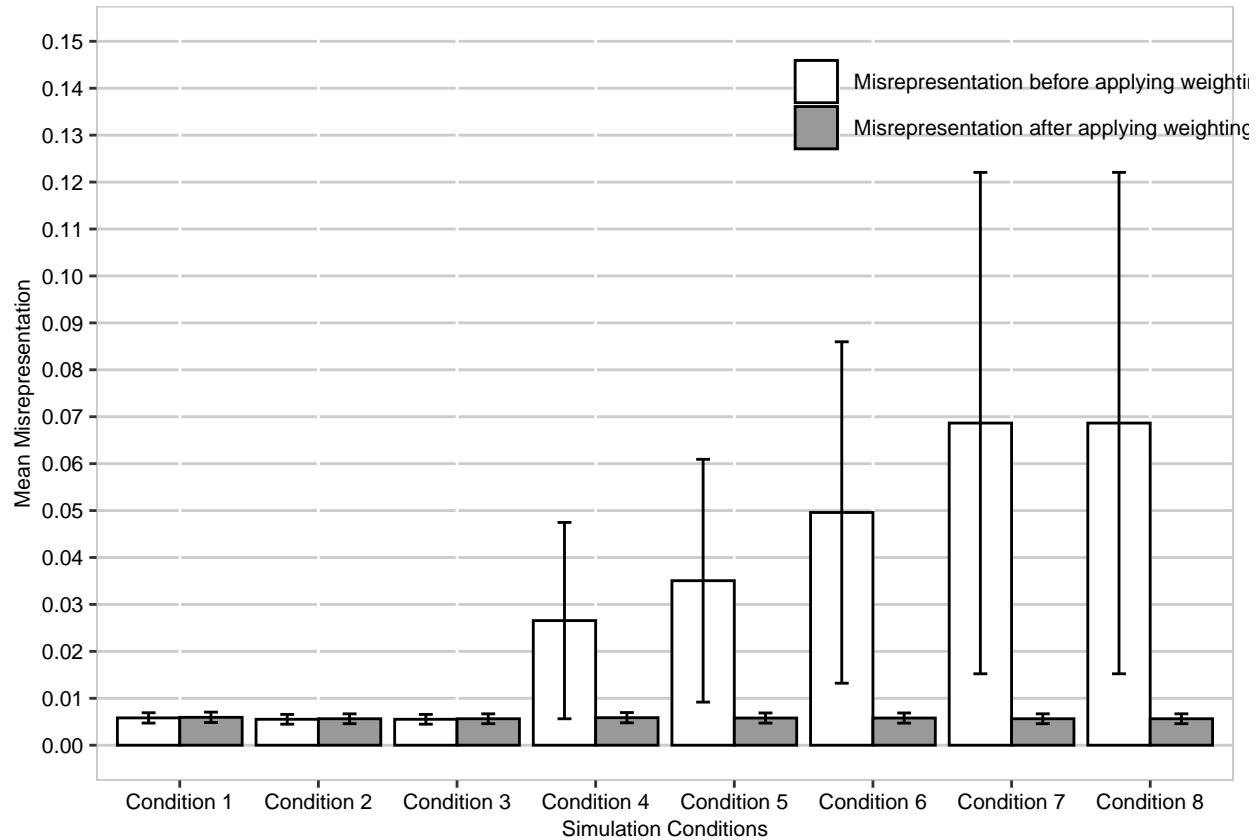
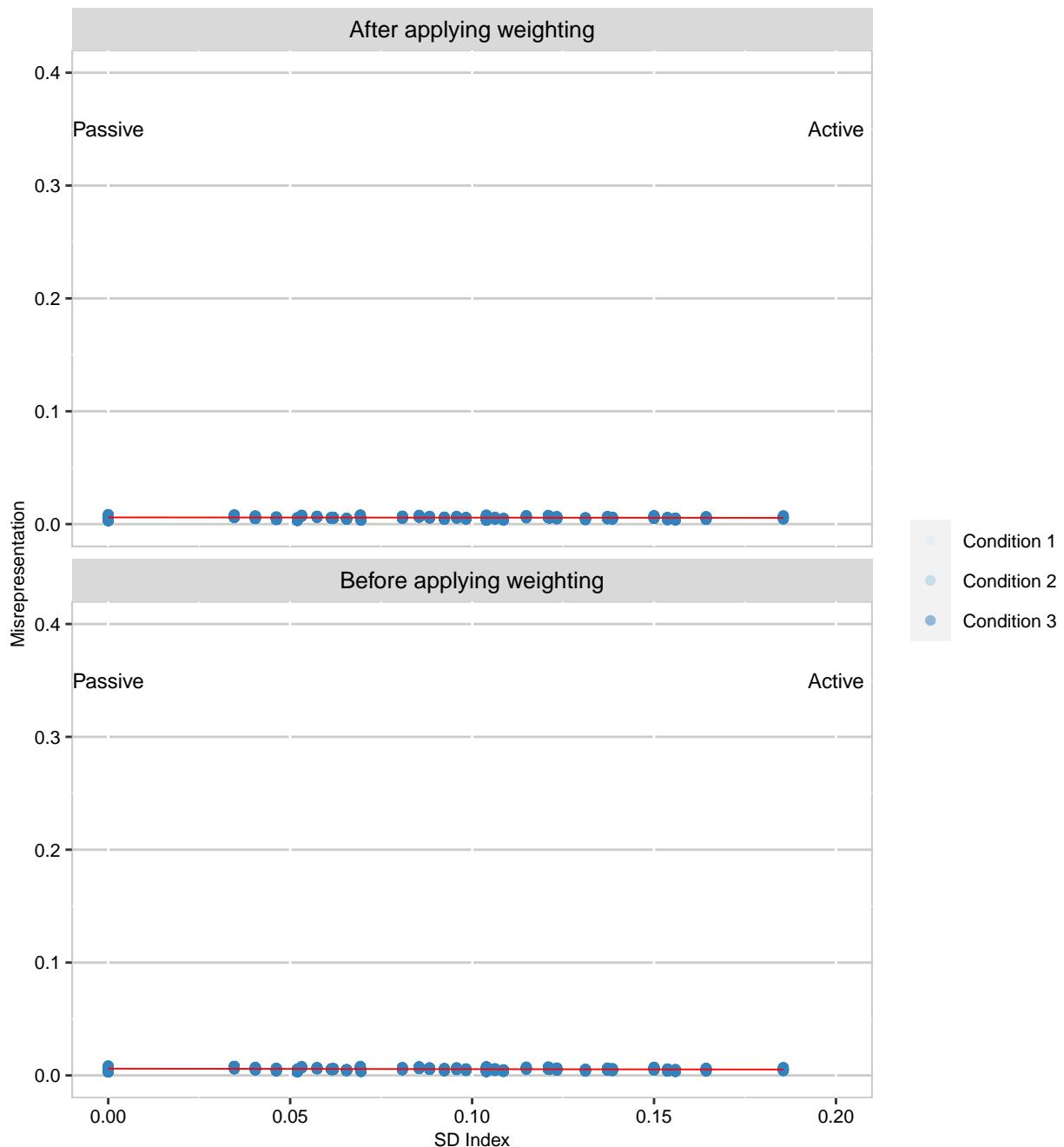


Figure 3

Average absolute discrepancy (unweighted in white and weighted in grey) across the eight attitudinal conditions.

the right). There is a very slight amount of error in the unweighted sample mean with active nonresponse, as well as a systematic pattern of heteroskedasticity across the “passive to active” continuum (studentized Breusch-Pagan = 565.42 [unweighted], 496.67 [weighted], p 's < .001). Weighting always corrects this slight amount of error. Figure 3 demonstrates a more pronounced *form of* nonresponse association when underlying attitudinal distributions evidence group differences, and in these scenarios, active nonresponse is shown to have a fairly large effect on error within the sample estimate (and, again, predictable heteroskedasticity paralleling the SD index, Breusch-Pagan = 3177.2 [unweighted]; 832.91 [weighted], p 's < .001). Weighting again corrects the sample estimate.

**Figure 4**

Presence (or lack) of error in unweighted and weighted sample estimates across passive and active forms of nonresponse (Conditions 1 through 3).

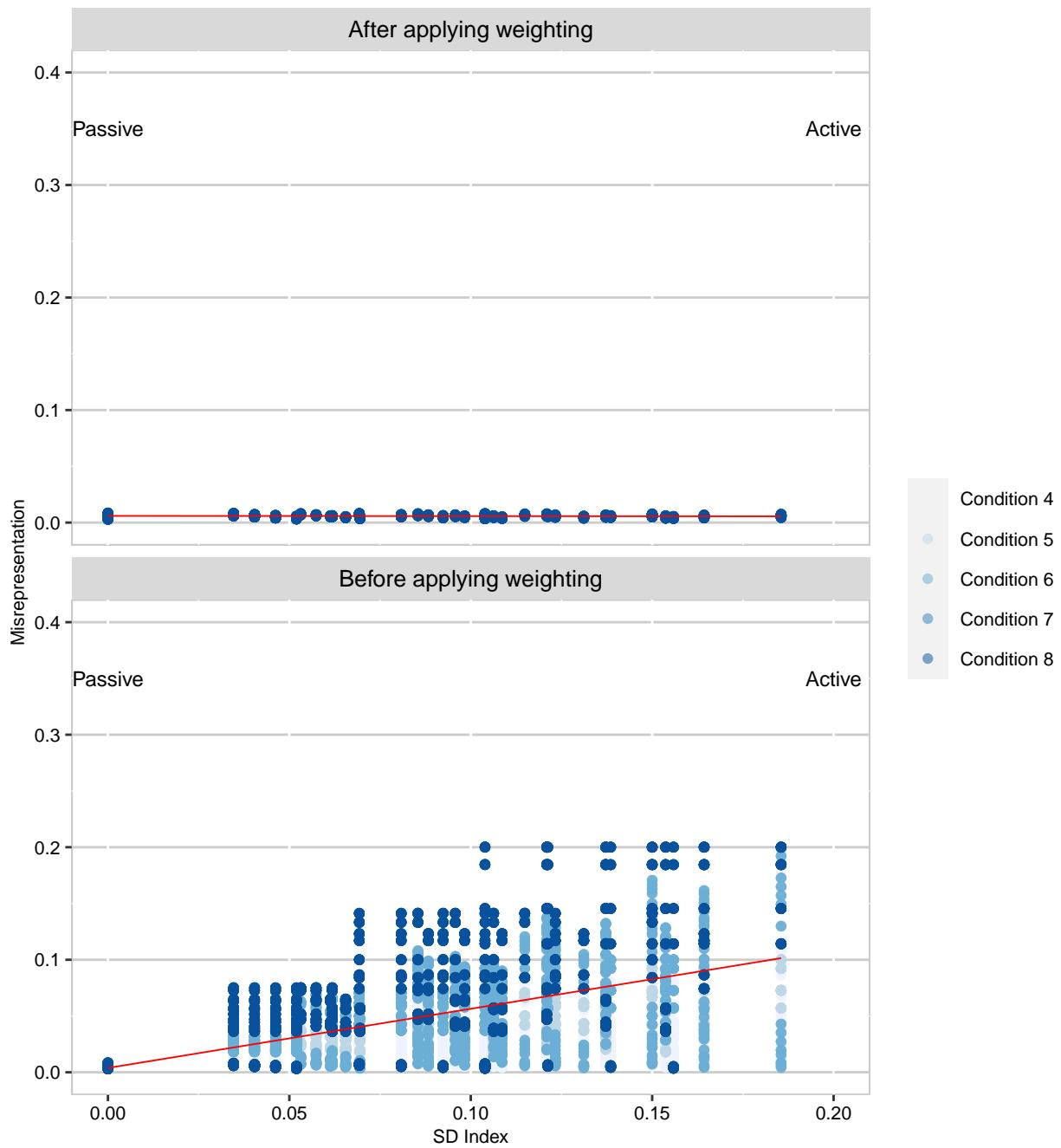


Figure 5

Presence (or lack) of error in unweighted and weighted sample estimates across passive and active forms of nonresponse (Conditions 4 through 8).

347 It should be noted regarding the above-mentioned “heteroskedasticity” that there

348 are active nonresponse scenarios in which no error is found (see, for example, the lower

349 right-hand portion of Figure 3 where values appear all along the passive-active abscissa).

350 These situations are ones within which the response rates “parallel” the distributional

351 form. For example, in Condition Eight, the distributional forms were: Positive Skew_{Male_A},

352 Positive Skew_{Male_B}, Negative Skew_{Female_A}, Negative Skew_{Female_B}. In the most extreme

353 cases of active nonresponse, response rates that fully parallel distributional patterns (e.g.,

354 20%_{Male_A}, 20%_{Male_B}, 80%_{Female_A}, 80%_{Female_B}) result in no error in the population mean

355 approximation (average discrepancy = .0003, SD = .0002). Alternatively, when the

356 response rates are inverted, (e.g., 20%_{Male_A}, 80%_{Male_B}, 20%_{Female_A}, 80%_{Female_B}), there

357 is substantial error in approximation (average discrepancy = .51, SD = .14). **this is an**

358 **old number - why are our new numbers so low? (see, for example, the y-axis**

359 **on Figure 1) - YANG? (11/17/18)** Again, it is not merely response rate or form that

360 is associated with biased sample estimates, but rather the nature of response rate relative

361 to existing attitudinal differences.

362 To further elaborate this point, consider, for example, Condition 4. Here, three

363 groups are characterized by similar distributions of attitudes (normally distributed) and

364 one, Females from Department B, is characterized by negatively skewed attitudes. The

365 greatest unweighted error here arises from sampling scenarios in which there are many

366 Department B females (e.g., in our specifications, 6,400) and fewer males and Department

367 A females⁶, but the Department B females exhibit a much lower response rate (e.g., 20%)

368 than do other groups, who respond at a high rate (e.g., 80%). That is, it is not merely

369 response rate, but response rate within these identifiable groups, and whether or not those

⁶ Because of the “marginal” versus “cell” specifications of population constituencies, our most extreme example here is necessarily 400 Department A males, 1,600 Department B males, and 1,600 Department A females. This was a decision based on keeping the population N’s at 10,000 and certainly more extreme population constituency combinations could be examined in future like-minded explorations.

370 response rate differences parallel underlying attitudinal differences.

371 Although the *patterns* of unweighted sample mean discrepancies differed across
372 conditions, all eight conditions exhibited similar omnibus effect (weighting ameliorating
373 error wherever it arose [in the unweighted statistic]).

374 To partially address the second limitation, discrepancy between population
375 constituency and sampling proportions was additionally estimated via Cattell's profile
376 similarity index [r_p ; Cattell (1949); Cattell et al. (1966)]. r_p is sensitive to discrepancies in
377 profile shape (pattern across profile components), elevation (average component score), and
378 scatter (sum of individual components' deviation from the elevation estimate). Figure 3
379 demonstrates the pattern of unweighted sample mean deviation (from the population
380 parameter) when this index is taken into consideration. *edits... .gain* demonstrate these
381 relationships across the attitudinal form conditions, being grouped by underlying
382 distributions thought to be susceptible to bias (Conditions 3 through 8) as well as those
383 thought to be relatively immune to bias (Conditions 1 through 3; aka those sampling
384 situations in which weighting is unnecessary).

385 Summary

386 Collectively the results highlight three aspects of weighting: 1) our simulations are
387 comprehensive, iterating through all possible combinations of response rates - those
388 paralleling population distributions, those inversely mirroring population distributions, and
389 those "orthogonal to" population distributions, 2) the "SD" operationalization of passive to
390 active forms of nonresponse is a bit crude and insensitive to specific combinations of
391 response rates expected to manifest or not manifest in bias, and 3) substantial bias may be
392 present in the unweighted estimate even with only small proportions of active non-response
393 (e.g., only one or two groups exhibiting slightly different response rates, with the resulting
394 discrepancy [population versus sample mean] being quite large).

395 Mean square error is our second index for sample quality. It is a well-known

396 mathematical theorem that the application of weights increases (random) errors of
397 precision, which was also empirically true in the current study. For each condition in our
398 simulations, we calculated the standard deviations of 40.96 million unweighted and 40.96
399 million weighted samples means (4,096 possible population-sample combinations by 10,000
400 iterations), which yielded eight empirically-estimated standard errors of unweighted and
401 weighted sample means. Figure XXX <- need to readd this visually presents these
402 standard errors in eight pairs of bars, demonstrating that the standard error of weighted
403 sample means (red bar) tended to be 16% to 18% larger than that of unweighted sample
404 means (grey bar) regardless of condition. These errors highlight the caveat that weighting
405 should only be applied in the active nonresponse case (e.g., although the aggregate effect of
406 weighting with passive nonresponse is error-minimizing, any one sampling condition is
407 *more likely* to result in greater deviation from the population parameter when weighting is
408 applied the passive nonresponse data).

409 In summary, as an aggregate across sampling events, weighting always corrects
410 sample bias, when it is present in the unweighted estimate. However, the standard errors
411 suggest that for any *one* sampling event in the absence of bias, the likelihood that the
412 sample mean approximates the *mean* of sample means is (slightly) greater for the
413 unweighted estimate. When bias is present, however, (in the unweighted estimate) there is
414 obviously no advantage to “being closer” to this biased mean of means. That is, under
415 some circumstances, the mean of unweighted sample means does not center on the
416 population mean. The implications of this seem quite obvious: Weighting should only be
417 applied if bias is anticipated in the sample estimate. This may seem to be a picayune
418 recommendation, but we note here that this advocation is not heeded in public opinion
419 polling applications, where the computation and application of weights are default
420 procedures (CITES? - perhaps AAPOR standards or personal communication with polling
421 agencies such as Gallop).

422 Can we look at the “crossing point?” (e.g., when MSE becomes

423 excessive) - David?

424

Discussion

425 We view nonresponse as a serious problem that should be addressed via repeated
426 attempts to survey particularly reluctant or hard-to-reach respondents particularly because
427 nonresponse may be reasonably expected to be greatest in groups that are most unsatisfied
428 [e.g., it may be typical for individuals representing these groups to have their responses
429 diluted; see, for example, Taris and Schreurs (2007)]. However, several researchers have
430 noted potentially misplaced relative emphasis on survey response rates, with Cook et al.
431 (2000), Krosnick (1999), and Visser et al. (1996) articulating the point that
432 representativeness of the sample is more important than response rate. We also believe
433 that the goal in organizational surveying should be representativeness not exhaustiveness.
434 Krosnick (1999) specifically comments that, even when probability sampling is employed,
435 response rate does not necessarily implicate either good or poor sample representativeness.
436 One aim of this paper is to reinforce this primary ‘representativeness’ orientation to those
437 who may be otherwise inclined to focus on response rate as a sufficient index of quality
438 (and propose sample weighting as a practice that can adjust for lack of representativeness).

439 With the above in mind, we set out to answer two fairly simple questions: What
440 impact does the application of weights have on the quality of sample estimates, and what
441 role does nonresponse play? Our answers are that: 1) weighting “always” helps, as long as
442 you capture the proper strata (which of course we were able to do via controlled
443 simulation), but also 2) response rate impact *depends* on relationship between response
444 rate and the underlying distribution of attitudes. conditions 1 through 3 as well as all
445 other conditions are occasionally immune to response rate influence, depending on whether
446 the pattern of nonresponse parallels the pattern of attitudinal distribution differences or
447 not). Active forms of nonresponse can harm the unweighted sample estimate, but only
448 when the pattern of active nonresponse is accompanied by differing distributions of

449 attitudes within the active nonrespondent “populations” [this would appear to be a
450 reasonable expectation based on the literature; e.g., Rogelberg et al. (2000); Rogelberg et
451 al. (2003); Spitzmüller et al. (2007)]. Although the weighted mean proved an unbiased
452 estimate of the population mean across all simulations, in circumstances where no bias
453 existed in the unweighted estimate, the trade-off between bias-correction and random error
454 of precision (e.g., standard error) also needs to be acknowledged.

455 It should be noted that the organizational surveying categorization of passive versus
456 active parallels the broader statistical focus on data that is missing at random or
457 completely at random (MAR or MCAR, see for example, Heitjan & Basu, 1996) versus
458 data not missing at random [non-MCAR, see for example,]. Imputation is the common
459 remediation for data MAR or MCAR whereas non-MCAR solutions may involve strategies
460 such as latent variable estimation procedures (Muthén et al., 1987). In the context of
461 surveying, we are similarly proposing a bifurcation of remediation methods - no
462 remediation with passive nonresponse and post-stratification weighting with active.

463 Previous presentations have noted that bias is sometimes associated with
464 nonresponse and othertimes it is not - this research has not been explicit in the specific
465 conditions that moderate this association, however. The current paper does make this
466 association explicit. It is not merely the form of nonresponse that determines whether or
467 not bias occurs, but also the underlying distributions that the response probabilities are
468 applied to. Some distributional patterns are immune to the biasing effects of active
469 nonresponse (see, for example, Conditions 1 through 3). Some patterns of active
470 nonresponse also result in no bias even when distributional patterns deviate substantially
471 (see, for example, Condition 8 where a 20%, 20%, 80%, 80% response rate pattern exhibits
472 no error). The target therefore should not be merely form of nonresponse but also
473 underlying attitudes. Regardless, however, weighting always remediates the error when it
474 occurs (and does not add error where it is absent).

475 The current findings are of course qualified by the uniqueness of our simulations,
476 most notably our ability to fully capture the correct population parameters (e.g., because
477 these were “created” by us, we were also able to identify these strata as the nonresponse
478 contributors). Even in the extreme conditions (e.g., a small “population” with a
479 correspondingly low response rate; see, for example, the lower-left hand corner of Figure 2),
480 the weighting algorithm was able to provide a bias correction. This is undoubtedly
481 attributable to our random sampling procedure (instead of, for example, sampling
482 conditionally from the population distributions), but here we do note that the raking
483 procedure is applied at the “margins” (e.g., variable level, not interaction level), although
484 our introduction of a biasing element is at the cell (interaction) level.

485 It has been stated that active nonresponse is relatively harmless unless the actively
486 nonrespondent group is relatively large [cites below]. The current study, however, suggests
487 that post-data-collection remediation. There may also be some important implications here
488 regarding sample (and population) size. Because organizational surveyors likely interface
489 with organizations of varying sizes (perhaps some of which are small- or medium-sized), the
490 implications of our simulations particularly in the small population conditions, were
491 highlighted. Findings specific to these conditions were: XXX, XXX, XXX.

492 There is of course no need to restrict weighting protocols to demographic groups -
493 organizational surveyors have a rich tradition of attending to drivers of nonresponse (see,
494 for example, Rogelberg & Stanton, 2007). Groupings of any sort can be the basis of
495 weighting (for example, pre-survey probing might assign probabilities of nonresponse, and
496 these probabilities can be retained post-administration as weighting guides).

497 It should also be pointed out that although the active nonrespondent group seems
498 to be a great concern, it will not seriously bias the results unless the proportion of active
499 nonrespondents is higher than 15% (Rogelberg et al., 2003; Rogelberg & Stanton, 2007;
500 Werner et al., 2007). “In this study we found that the active nonrespondent group was

501 relatively small (approximately 15%), but consistent in size with research conducted by ”
502 (Rogelberg et al., 2003, pp. 1110–1111). “Furthermore, consistent with Roth (1994) who
503 stated that when missingness is not random (as we found for active nonrespondents),
504 meaningful bias will only be introduced if the group is relatively large (which was not the
505 case in this study).” (Rogelberg et al., 2003, p. 1112).

506 “If the results show that the active nonrespondent group comprises a low proportion
507 of the population, fewer concerns for bias arise. If the proportion of active respondents is
508 greater than 15% of the group of individuals included in the interviews or focus groups
509 (this has been the average rate in other studies), generalizability may be compromised.”
510 (Rogelberg & Stanton, 2007, p. 201) * I believe there is an error here. The author want to
511 say that if the proportion of active nonrespondents is greater than 15% of the group .

512 “It has been suggested that it takes a response rate of 85% to conclude that
513 nonresponse error is not a threat (Dooely & Lindner, 2003). We agree that researchers
514 should provide both empirical and theoretical evidence refuting nonresponse bias whenever
515 the response rate is less than 85%.” (Werner et al., 2007, p. 293).

516 Note here however, the seeming disconnect between the reports of 15% active
517 nonresponse and declining response rates (trending toward 50%). Certainly with
518 decreasing overall response rates, the likely reasons would appear to be more active than
519 passive (e.g., it is difficult to entertain the idea that potential respondents are more likely
520 to forget to respond today than they were 40 years ago).

521 Integration of IT/IS systems within HR functions hopefully assists the
522 (un)likelihood that organizatioal population frames are either deficient or
523 contaminated, although we note that this possibility (frame misspecification) is
524 much more plausible within organziations that do not have updated or
525 integrated HR IT/IS systems (perhaps, ironically, *smaller* organizations).

526 Future Directions

527 A very practical implication of this study is that future organizational researchers
528 may find more success implementing strategic sampling strategies as opposed to (or in
529 addition to) pursuing response enhancement. That is, as a field, organizational researchers
530 have been focused on response-enhancing strategies that minimize the presence of
531 nonresponse. The current findings suggest that more careful adherence to random sampling
532 from carefully constructed population frames may provide a different route to the same
533 end-goal of sample representativeness.

534 Experimental methods within the psychological discipline have long been criticized
535 for heavy reliance on samples of convenience (for instance, student samples). Very little
536 progress has been made regarding the application of appropriate population sampling
537 procedures in experimentation. Certain non-experimental procedures (most notably
538 organizational surveying) hold paradoxical advantage over experimental procedures
539 primarily in this arena of sampling - particularly in consideration of population coverage,
540 which refers to the percent of a population that is reachable by the sampling procedure
541 (e.g., postal, intra-office, or internet invitation) and likelihood of having access to
542 population parameter estimates (e.g., strata constituencies). There is a rich tradition and
543 literature of public opinion polling procedures and techniques from which to draw. These
544 procedures, however, only hold advantage if the non-experimental methodologist
545 acknowledges the criticality of sample representativeness. The current paper provides one
546 corrective technique (post-stratification weighting) as an important focus for the
547 organizational surveyor who shares this primary interest in maximizing sample
548 representativeness.

549 We note the above “advantage” held by organizational surveyors because extensions
550 of the current protocol include investigating how inaccurate census estimates (and/or
551 grabbing the “wrong” group) affects the quality of sample estimates. That is, in our

controlled simulations, we were able to know population constituencies, because they were set by us! In real-world applications, there is likely more error between the population estimate and actual population constituency. Similarly, if the association between attitude and group membership were to be controlled, there may be conditions identified whereby weighting loses its efficacy (e.g., low “correlations” between attitude and group membership). Future simulations should test boundary conditions for this type of error, identifying at what point inaccuracy in the population constituency estimate appreciably degrades the weighting procedure. Furthermore, it was demonstrated here that, when bias exists, weighting corrects it. Weighting also, however, results in a larger mean square error (MSE; expected spread of sample estimates around the population parameter). Feasibly then, there is a point at which the decreased bias is accompanied by an unacceptably inflated MSE. At which point does this occur? This is another fertile area for future exploration.

Most potential issues with weighting are addressed through careful consideration of the appropriate strata to take under consideration as well as ultimate level of aggregation (what group constitutes the population of interest or focus of feedback; e.g., regional, functional, or organizational?). We recommend the surveyor especially considers groups that might have issues of active forms of nonresponse and collect those demographics so weighting is an option. It is particularly in these contexts of ‘unsatisfied’ employees being less likely to respond to surveys that pre-stratification consideration becomes critical (for instance, if there is an inclination that attitudes may differ across, for example, night versus day shift workers, it is important that shift be measured and incorporated as a stratum prior to survey administration).

For Condition 5 (for example, low/high response rates with minority/majority population constituencies). The lower-right to upper-left diagonal reflects response rates that parallel population constituencies. The patterns across these stressors were consistent, with the weighted sample means (red dots) providing unbiased estimates of the population

579 parameter, whereas the unweighted sample means (grey dots) tended to yield unbiased
580 estimates when the population constituencies were roughly equal (e.g., close to 50%/50%).

581 Figure 3 drills down this information further by extracting unweighted and weighted
582 estimates in one specific marginal population parameter combination (here, 60% males and
583 40% females; 40% in department A and 60% in department B). In doing so, the population
584 parameters were in control and sample parameters were set free (see dotted red rectangle
585 in Figure 2). Therefore, Figure 3 was then arranged in a fashion that allowed further
586 investigation into the interactive effect of marginal sample parameters (gender on the
587 x-axis and department on the y-axis) on the effectiveness of post-stratification weighting
588 reflected by the pattern of grey and red dots. **Huh? - find old version or delete**

References

- Anseel, F., Lievens, F., Schollaert, E., & Choragwicka, B. (2010). Response rates in organizational science, 1995–2008: A meta-analytic review and guidelines for survey researchers. *Journal of Business and Psychology*, 25(3), 335–349.
- Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. <https://github.com/crsh/papaja>
- Baruch, Y. (1999). Response rate in academic studies—a comparative analysis. *Human Relations*, 52(4), 421–438.
- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8), 1139–1160.
- Biemer, P. P., & Lyberg, L. E. (2003). *Introduction to survey quality* (Vol. 335). John Wiley & Sons.
- Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, 10(4), 689–709.
- Cattell, R. B. (1949). R p and other coefficients of pattern similarity. *Psychometrika*, 14(4), 279–298.
- Cattell, R. B., Coulter, M. A., & Tsujioka, B. (1966). The taxonometric recognition of types and functional emergents. *Handbook of Multivariate Experimental Psychology*, 288–329.
- Cook, C., Heath, F., & Thompson, R. L. (2000). A meta-analysis of response rates in web-or internet-based surveys. *Educational and Psychological Measurement*, 60(6), 821–836.
- Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly*, 64(4), 413–428.
- Cycyota, C. S., & Harrison, D. A. (2002). Enhancing survey response rates at the executive level: Are employee-or consumer-level techniques effective? *Journal of*

- 616 Management, 28(2), 151–176.
- 617 Cycyota, C. S., & Harrison, D. A. (2006). What (not) to expect when surveying
618 executives: A meta-analysis of top manager response rates and techniques over
619 time. *Organizational Research Methods*, 9(2), 133–160.
- 620 Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a
621 sampled frequency table when the expected marginal totals are known. *The
622 Annals of Mathematical Statistics*, 11(4), 427–444.
- 623 Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A
624 systematic review. *Computers in Human Behavior*.
- 625 Frohlich, M. T. (2002). Techniques for improving response rates in OM survey
626 research. *Journal of Operations Management*, 20(1), 53–62.
- 627 Fulton, B. R. (2016). Organizations and survey research: Implementing response
628 enhancing strategies and conducting nonresponse analyses. *Sociological Methods
629 & Research*, 0049124115626169.
- 630 Heitjan, D. F., & Basu, S. (1996). Distinguishing “missing at random” and “missing
631 completely at random.” *The American Statistician*, 50(3), 207–213.
- 632 Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the
633 impact of growing nonresponse on estimates from a national RDD telephone
634 survey. *International Journal of Public Opinion Quarterly*, 70(5), 759–779.
- 635 Kessler, R. C., Avenevoli, S., Costello, E. J., Green, J. G., Gruber, M. J., Heeringa,
636 S., Merikangas, K. R., Pennell, B.-E., Sampson, N. A., & Zaslavsky, A. M.
637 (2009). National comorbidity survey replication adolescent supplement (NCS-a):
638 II. Overview and design. *Journal of the American Academy of Child &
639 Adolescent Psychiatry*, 48(4), 380–385.
- 640 Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1),
641 537–567.
- 642 Kulas, J. T., Robinson, D. H., Kellar, D. Z., & Smith, J. A. (2017). Nonresponse in

- 643 organizational surveying: Attitudinal distribution form and conditional response
644 probabilities' impact on patterns of bias. *Public Opinion Quarterly*, 81(2),
645 401–421.
- 646 Kulas, J. T., Robinson, D. H., Smith, J. A., & Kellar, D. Z. (2016).
647 Post-stratification weighting in organizational surveys: A cross-disciplinary
648 tutorial. *Human Resource Management*.
- 649 Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary
650 distinctions between organizational, mechanical turk, and other convenience
651 samples. *Industrial and Organizational Psychology*, 8(2), 142–164.
- 652 Luong, A., & Rogelberg, S. G. (1998). How to increase your survey response rate.
653 *The Industrial-Organizational Psychologist*, 36(1), 61–65.
- 654 Mellahi, K., & Harris, L. C. (2016). Response rates in business and management
655 research: An overview of current practice and suggestions for future direction.
656 *British Journal of Management*, 27(2), 426–437.
- 657 Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with
658 data that are not missing completely at random. *Psychometrika*, 52(3), 431–462.
- 659 Pasek, J. (2016). *Anesrake: ANES raking implementation*.
- 660 Pedersen, M. J., & Nielsen, C. V. ek. (2016). Improving survey response rates in
661 online panels: Effects of low-cost incentives and cost-free text appeal
662 interventions. *Social Science Computer Review*, 34(2), 229–243.
- 663 Quine, S., & Morrell, S. (2008). Feeling safe in one's neighbourhood: Variation by
664 location among older australians. *The Australian Journal of Rural Health*, 16,
665 115–116.
- 666 R Core Team. (2017). *R: A language and environment for statistical computing*. R
667 Foundation for Statistical Computing. <https://www.R-project.org/>
- 668 Rivers, D., & Bailey, D. (2009). Inference from matched samples in the 2008 US
669 national elections. *Proceedings of the Joint Statistical Meetings*, 1, 627–639.

- 670 Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., &
671 Knight, W. E. (2003). Profiling active and passive nonrespondents to an
672 organizational survey. *Journal of Applied Psychology, 88*(6), 1104.
- 673 Rogelberg, S. G., Luong, A., Sederburg, M. E., & Cristol, D. S. (2000). Employee
674 attitude surveys: Examining the attitudes of noncompliant employees. *Journal*
675 *of Applied Psychology, 85*(2), 284.
- 676 Rogelberg, S. G., & Stanton, J. M. (2007). *Introduction: Understanding and dealing*
677 *with organizational survey nonresponse*. Sage Publications Sage CA: Los
678 Angeles, CA.
- 679 Spitzmüller, C., Glenn, D. M., Sutton, M. M., Barr, C. D., & Rogelberg, S. G.
680 (2007). Survey nonrespondents as bad soldiers: Examining the relationship
681 between organizational citizenship and survey response behavior. *International*
682 *Journal of Selection and Assessment, 15*(4), 449–459.
- 683 Taris, T. W., & Schreurs, P. J. (2007). How may nonresponse affect findings in
684 organizational surveys? The tendency-to-the-positive effect. *International*
685 *Journal of Stress Management, 14*(3), 249.
- 686 Tett, R., Brown, C., & Walser, B. (2014). The 2011 SIOP graduate program
687 benchmarking survey part 7: Theses, dissertations, and performance
688 expectations. *The Industrial-Organizational Psychologist, 51*(4), 62–73.
- 689 Visser, P. S., Krosnick, J. A., Marquette, J., & Curtin, M. (1996). Mail surveys for
690 election forecasting? An evaluation of the columbus dispatch poll. *Public*
691 *Opinion Quarterly, 60*(2), 181–227.
- 692 Wainer, H. (1976). Estimating coefficients in linear models: It don't make no
693 nevermind. *Psychological Bulletin, 83*(2), 213.
- 694 Werner, S., Praxedes, M., & Kim, H.-G. (2007). The reporting of nonresponse
695 analyses in survey research. *Organizational Research Methods, 10*(2), 287–295.