Nonresponse and Sample Weighting in Organizational Surveying

Nonresponse and Sample Weighting in Organizational Surveying

In order to represent different proportions of relative constituency (for example, more females than males or more department A workers than department B), we iterated population characteristics at marginal levels (gender and department) starting at 20% (and 80%) with increments and corresponding decriments of 20%. For example, if males accounted for 20% of the simulated population, then females were 80%; also if respondents in Department A represented 60% of a population, then 40% were in Department B. Marginal constituencies were therefore realized at all combinations (across the two variables) of 20% and 80%, 40% and 60%, 60% and 40%, and 80% and 20%. This resulted in population *cell* constituencies (e.g., Male.A, Female.A, Male.B, Female.B) as low as 400 and as high as 6,400 - see Figure **??** for further clarification of our "cell" and "margin" terminology and variable specification.

Each population cell was characterized by an attitudinal distribution in one of three different possible forms: normal, positively skewed, or negatively skewed. These distributional forms were specified in an attempt to model similarities and discrepancies in construct standing (e.g., commitment, satisfaction, or engagement) across respondent groupings. The normal distribution exhibited, on average, a mean of 3.0 whereas the skewed distributions were characterized by average means of 2.0 and 4.0, respectively. In total, eight crossings of distributional type across employee categorization were specified (Table 1 presents the combinations of these distributions). Note that these eight conditions are not exhaustive of all possible combinations of constituent groups and attitudinal distribution - we limited the simulations to combinations that we projected to collectively be most efficiently informative.

Individual attitudes were randomly sampled from population distributions at the cell level (e.g., Male.A) without replacement. These response rates (methodologically these could alternatively be conceptualized as *sampling* rates) were specified at 10% increments ranging from 60% to 90%, and these were fully iterated across each of our four marginal groups (Males, Females, Departments A and B). Our cell-level response rates therefore ranged from 36% to 81% - a range of rates specified because they are approximations of reasonable expectations according to the organizational surveying literature (e.g., Mellahi & Harris, 2016; Werner et al., 2007). We therefore investigated error within the aggregate mean (e.g., grand mean aka total sample mean) attributable to different likelihoods of sample inclusion from constituent groups of different relative size and representing populations of different attitudinal distribution, but at response rates reasonably expected to exist in real-world organizational surveying contexts.

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., male.a = 36%, female.a = 36%, male.b = 42%, and female.b[1] = 42%; see the second row of Table 2, the SD index = .034)[2]. 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

[1] "Lowercase" specification of simulation strata indicates sample constituencies (e.g., male.b) whereas uppercase implicates population (e.g., Male.B).

[2] This method of simplifying the presentation of our response rate conditions is fully independent of consideration of 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.

rate discrepancy from others'.

Results

Role of nonresponse form

The systematic patterns of heteroskedasticity of the Figure 2 scatterplots should also be noted. There are *active nonresponse* scenarios in which no error is present (see, for example, the lower right-hand portions of conditions 4 through 8 where discrepancy estimates of "0" persist at multiple points along the passive-active x-axis). These circumstances are simulated conditions within which the response rates "parallel" the *population distributional form*. For example, in Condition Eight, the distributional forms across populations were: $PositiveSkew_{Male(A)}$, $PositiveSkew_{Male(B)}$, $NegativeSkew_{Female(A)}$, $NegativeSkew_{Female(B)}$. Response rates that "mirror" distributional patterns in extreme cases of active nonresponse (e.g., SD = .156; $54\%_{\text{Male(A)}}$, $54\%_{\text{Male(B)}}$, $81\%_{\text{Female(A)}}$, $81\%_{\text{Female(B)}}$) result in effectively zero error in the population mean approximation (average discrepancy = 0.00, *SD* = 0.00). Alternatively, when the response rates are inverted for the SD=.156 cases, (e.g., $54\%_{\text{Male\_A}}$, $81\%_{\text{Male\_B}}$, $54\%_{\text{Female\_A}}$, $81\%_{\text{Female\_B}}$), there is substantial error in approximation (average discrepancy = 0.16, SD = 0.03). Here, it is not merely response rate or form that is associated with biased sample estimates, but rather the nature of response rate relative to existing attitudinal differences.[3] See Figure 5 for placeholder explanation.

To further expand upon this *attitudinal form/pattern of nonresponse* interplay, the discrepancies between population constituency and sampling proportions were additionally evaluated through the lens of Cattell's profile similarity index (r_p, Cattell, 1949; Cattell et al., 1966). r_p is sensitive to discrepancies in profile shape (pattern across profile components),

[3] Don't think this is correct - maybe frame: "sometimes the active non-response is non-troublesome - when it fully parallels the distributional proportions (?)" ← still confusing. Looked at with Yang 3/1/24 and still confused - maybe leave in for reviewers to note and question.

elevation (average component score), and scatter (sum of individual components' deviation from the elevation estimate. Here, the profile similarity index references the relationship between the response rates (NEED YANG TO VERIFY - THINK THIS IS SSmale;SSfemale;SSdepta;SSdeptb from `combo` object) and sample sizes (cellrate.ma;cellrate.mb;cellrate.fa;cellrate.gb) across experimental *cells*. For example, VERIFY BEFORE CLARIFYING HERE. Figure 3 demonstrates the pattern of unweighted sample mean deviation (from the population parameter) when this index is taken into consideration. Specifically, Figure 3 demonstrates a more pronounced *form of* nonresponse association when underlying attitudinal distributions evidence group differences (e.g., incrementally across the 8 specified conditions), and in these scenarios, active nonresponse is shown to have a fairly large effect on error within the sample estimate (as well as systematically increasing degrees of heteroskedasticity paralleling the Cattell index; omnibus Breusch-Pagan [across conditions] = 3177.2, $p < .001$). The curvilinear nature of these functions was estimated via hierarchical polynomial regression (excluding conditions 1, 2, and 3), with misrepresentation exhibiting a linear association across condition ($R^2 = 0.15$, $p < .001$) as well as incrementally across the Cattell index ($\Delta R^2 = 0.24$, $p < .001$), and also exhibiting an incremental polynomial effect ($\Delta R^2 = 0.07$, $p < .001$).

To further elaborate this point, consider, for example, Condition 4 as presented in Table 1. Here, three groups are characterized by similar distributions of attitudes (normally distributed) and one, Female.B, is characterized by negatively skewed attitudes. The greatest unweighted error here arises from sampling scenarios in which there are many Female.B (e.g., in our specifications, 6,400) and fewer males and Department A females[4], but the female.b exhibit a much lower response rate (e.g., 20%) than do other groups, who

[4] Because of the "marginal" versus "cell" specifications of population constituencies, our most extreme example here necessarily results in 400 Male.A's, 1,600 Male.B's, and 1,600 Female.A's. 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.

respond at a high rate (e.g., 80%). That is, it is not merely response rate, but response rate within these identifiable groups, and whether or not those response rate differences parallel underlying attitudinal differences that drives sample misrepresentation.

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* https://github.com/crsh/papaja

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.

Mellahi, K., & Harris, L. C. (2016). Response rates in business and management research: An overview of current practice and suggestions for future direction. *British Journal of Management*, *27*(2), 426–437.

Werner, S., Praxedes, M., & Kim, H.-G. (2007). The reporting of nonresponse analyses in survey research. *Organizational Research Methods*, *10*(2), 287–295.

Table 1

Attitudinal Distribution Conditions Specified in Current Paper

			Male		Female		
Condition	Distributional Shape	mu	Dept A	Dept B	Dept A	Dept B	Bias Susceptibility
1	Normal	3	X	X	X	X	Low
	Positive Skew	2					
	Negative Skew	4					
2	Normal	3					Low
	Positive Skew	2	X	X	X	X	
	Negative Skew	4					
3	Normal	3					Low
	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	

Table 2

Example Summarized Response Rate Conditions Represented in Figures 2 through 5

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%	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

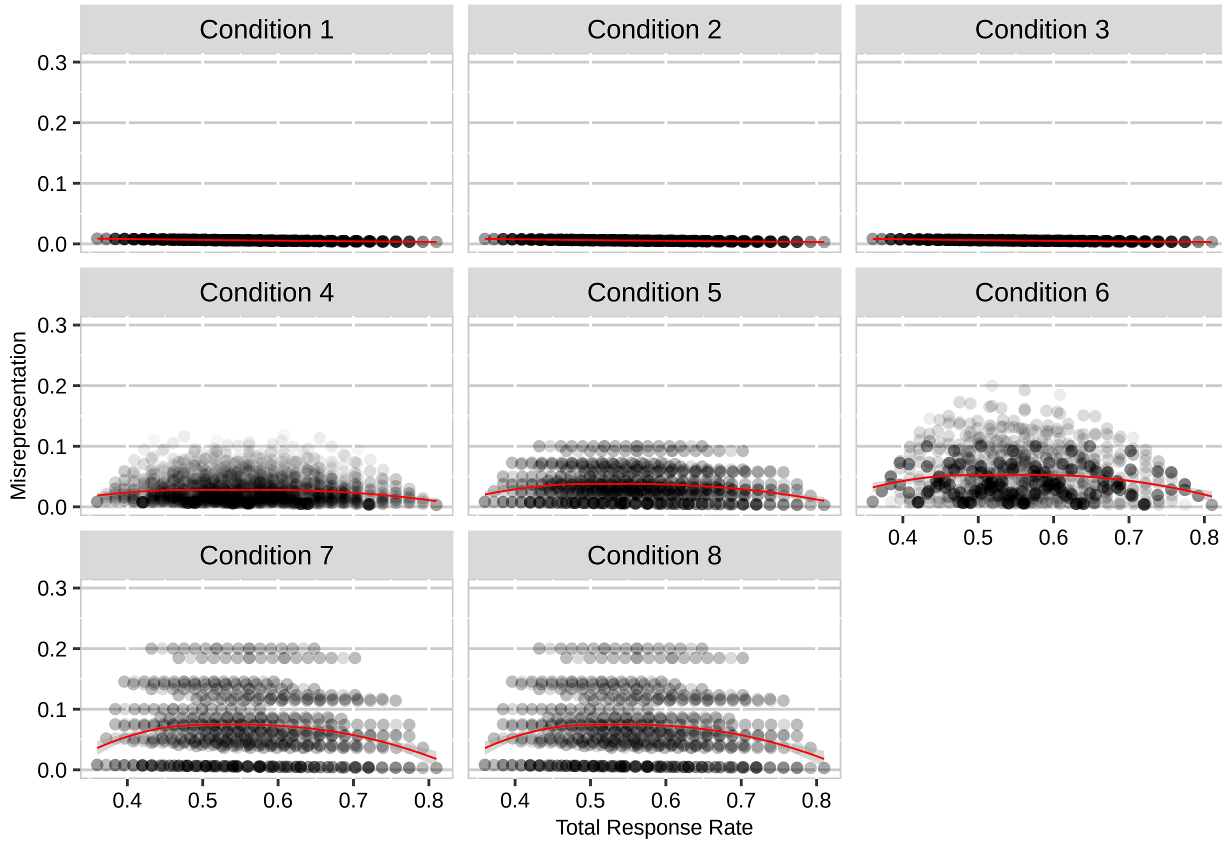

Figure 1

Relationship between total response rate and misrepresentation.

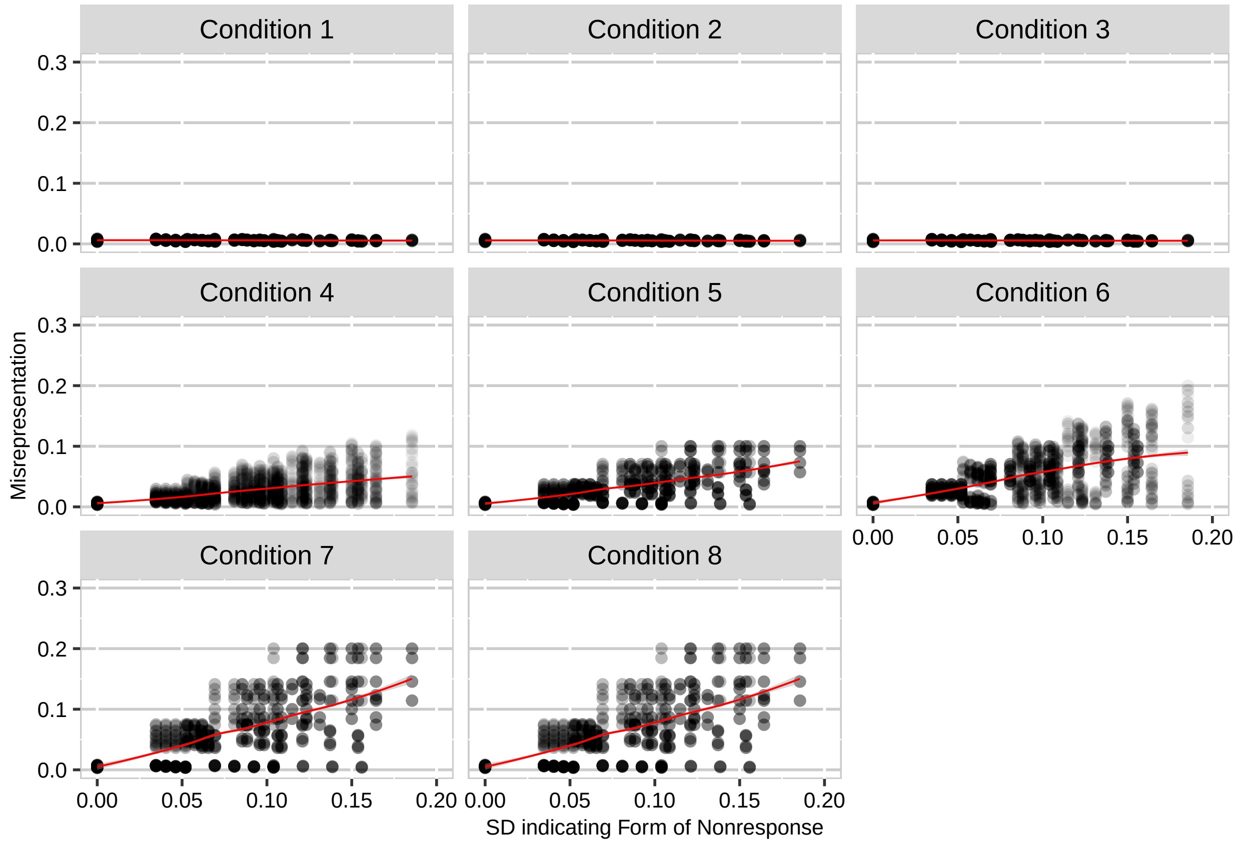

Figure 2

Relationship between nonresponse form and misrepresentation.

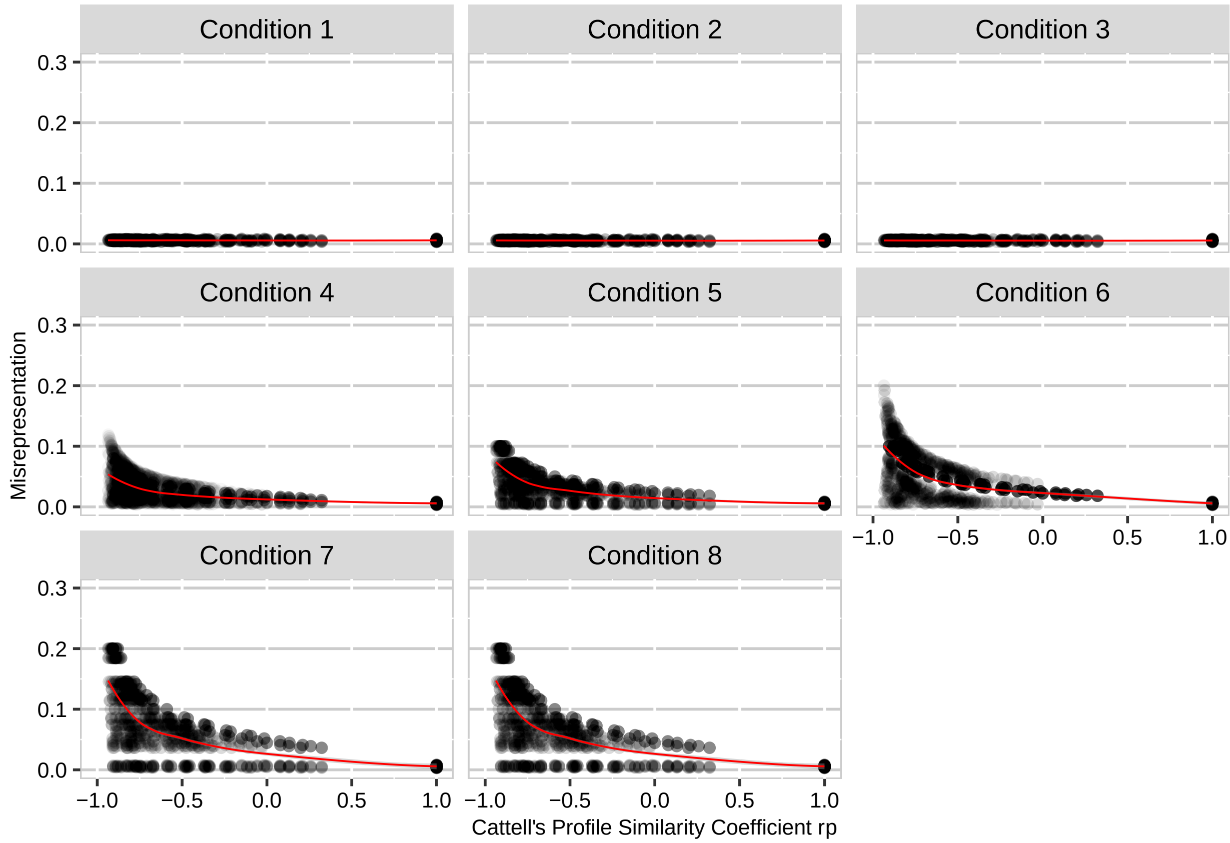

Figure 3

Effect of subgroup sampling rate match with distributional form on population misrepresentation.

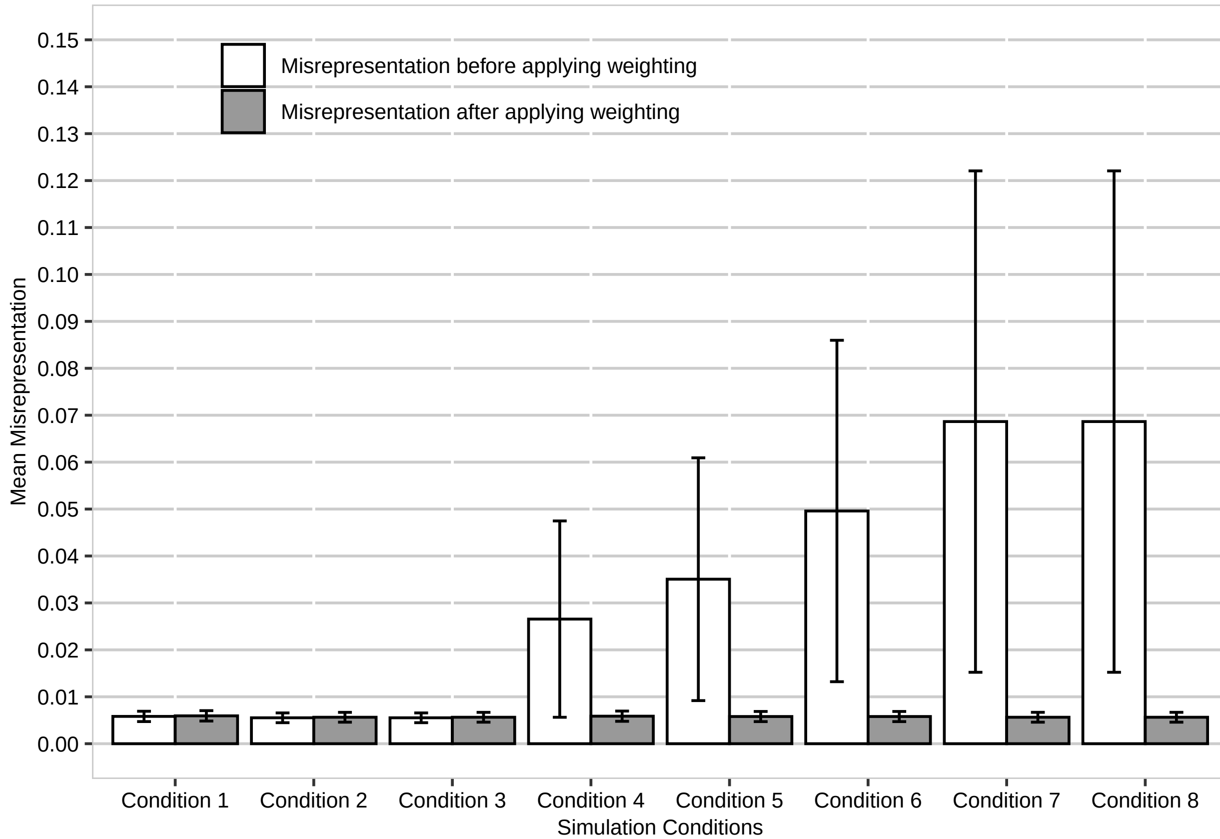

Figure 4

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

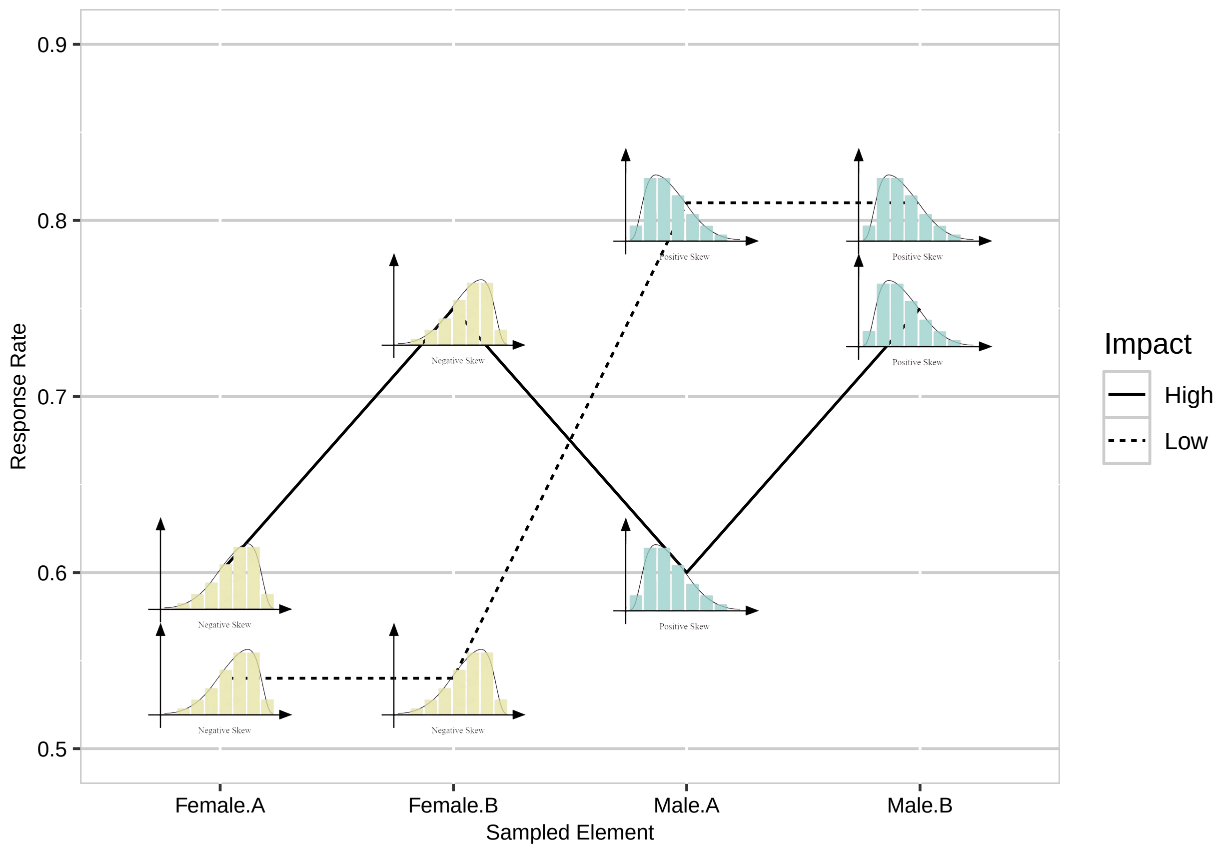

Figure 5

Allocation of response rates relative to underlying distributional form and its impact on population misrepresentation (need to think through hi/lo given Dr Robinsons thoughts)