Hurdles to Inference: The Demographic Correlates of Survey Breakoff and Shirking

David Fortunato
University of California, San Diego
dfortunato@ucsd.edu

Matthew V. Hibbing
University of California, Merced
mhibbing@ucmerced.edu

Tessa Provins
University of California, Merced
kts33@pitt.edu

August 19, 2021

Abstract

Scholars are all too familiar with the numerous threats to inference inherent in survey research. Some of these dangers can be mitigated with careful attention to sample and survey design, while others such as survey breakoff and inattentive responses are largely out of the control of the researcher. In this paper, we explore the prevalence of shirking among different demographic groups with a survey of 3,256 respondents and develop a means of identifying these response types. We find that younger male respondents with low educational attainment are most likely to complete surveys with low quality answers, while women are more likely to break off before completing the survey. These patterns suggest that breakoff and inattentive responses have different psychological determinants and also imply different hurdles to understanding the survey behavior of individuals that share specific demographic traits. Using our proposed identification strategy, researchers can identify shirking within surveys and take measures to correct sample issues caused by respondent behavior.

In social science research, the path to genuine insight is a perilous one, and every step of the journey is fraught with threats to valid inference. Obtaining a representative sample is an ever-evolving challenge, as changing habits in the population and developments in technology conspire to lead us astray. Even when the original sample is sound, the problem of survey nonresponse or inattentive response can confound our ability to draw meaningful conclusions. If we can successfully get the "right" respondents in our survey, we must still ask the "right" questions, and fifty years of survey research has documented a host of pitfalls associated with the writing of questions and the administration of surveys (DeCastellarnau 2018; Groves et. al. 2009; Saris and Gallhofer 2014; Steinbrecher, Rossmann, and Blumenstiel 2015). Viewed against the backdrop of these many challenges, the amount we have learned about the public's views and behaviors is quite remarkable.

In this paper we focus on one of the most frustrating challenges to survey research: nonresponse and "shirking." Conceptually we define shirking as the errors produced when respondents fail to provide quality response through various satisficing behaviors. In other words, the "errors" are produced through incompleteness or inattentiveness. Even the best designed instrument with an immaculate sample can be undermined if respondents systematically fail to complete the survey thus distorting the final sample. To make matters worse, not every incomplete survey looks incomplete when we consider satisficing behaviors. While some respondents overuse the "don't know" option as a form of item nonrepsonse or give up and clearly drop out of the survey (by hanging up the phone or closing the browser), other respondents continue to respond, but stop giving quality responses. The tendency to "straight-line" by providing identical responses to every question presents a unique challenge to acquisition of a high quality sample because it effectively introduces nonresponse that appears sincere on the surface.

In this short paper, we identify the demographic characteristics associated with three different types of shirking behaviors: breakoff (failure to complete survey), inattentive responses (as manifest in failure to correctly complete very simple questions or tasks, such as one's own age), and straight-lining (provide the same answer for every question). We analyze these behaviors utilizing a large survey (N=3,256) that includes the means to explicitly identify each type within the same sample (previous research has tended to only proxy for inattentiveness), a novel approach that allows us to safely compare the predictors of these types. We find that women are more likely to break off during a survey, while men are more likely to shirk while completing the instrument. In addition, young, less educated respondents are more likely to use straight-line tactics. What this means, of course, is that we are likely to misunderstand the (reported) social behavior of women and men in different ways as a function of these nonresponse tendencies. We illustrate the potential for inferential error with an application to respondents' views of the economy — one of the most central covariates in political science and economics research.

Our approach to identifying different types of survey shirking behaviors and examining the demographic trends in the propensity to use these behaviors will benefit practitioners of survey research in not only political science but other social science fields including economics, psychology, and sociology. While other studies have focused on specific within-survey shirking behaviors, we create a typology that will aid scholars in identifying not only incompleteness but inaccuracy of survey responses. Unlike other studies, leveraging respondents within the same survey gives us the ability to directly compare the effect of these demographic characteristics on different shirking behaviors. Finally, our findings have implications for understanding and, more importantly, developing methods to deal with biased samples caused by certain survey shirking practices.

Survey Shirking

Survey research uses total survey error (TSE) to map out the possible threats that lead to biased inferences and creates a framework for researchers to holistically think about maximizing data quality while balancing the financial and time constraints of designing surveys, obtaining samples, and analyzing responses (Weisberg 2005, Groves et al. 2009). This framework outlines errors to include: sampling error (error from estimating population characteristics from a sample), coverage error (error when members of the population are excluded from the sample), measurement error (error where the observed value differs from the true value), nonresponse error (when respondents in the sample do not respond or incompleteness of responses in sample), and postsurvey error (error produced through data editing, weighting and encoding). Within this framework, we focus on nonresponse errors which are produced through incompleteness or inattentiveness. The later creates a potentially larger concern given some response seem "complete" but lack being quality responses. The error is exacerbated when there are systematic differences in the types of respondents that use different types of nonrepsonse behaviors. Importantly, this is error that can manifest within high-quality, representative samples, using non-biased question construction.

There is some extant research on demographic correlates in survey nonresponse. Past articles have observed correlations between race and gender and the likelihood of an individual breaking off during a survey (Peytchev 2009; Steinbrecher, Rossmann, and Blumenstiel 2015). Unsurprisingly, there is evidence that respondent interest plays a large role in the propensity to break off from a survey (Galesic 2006; Groves, Presser, and Dipko 2004; Krosnick 1991), with cognitive ability, age, and education also quite influential (Conrad, Schober, Dijkstra 2007; Knauper 1999). Obviously, these differences in propensity to break off have major substantive implications. If breakoff were unrelated to individual characteristics it presents a problem in terms of costs and efficiency, but the threat to validity would be minimal. However, if certain types of people are more likely to break off their participation we must account for this risk.

Of course, a respondent does not have to formally terminate their participation in order to stop giving valid responses. As Krosnick (1991) outlines, there are several response patterns (broadly conceptualized as satisficing) that respondents can employ when the cognitive effort of the survey becomes too great. Among these are a tendency to overuse the “don’t know” option (e.g. Krosnick et al. 2002) or various methods of random response, including “straight-lining” (Herzog and Bachman 1981). Existing research has not established whether these strategies are psychologically comparable to break off, but there is reason to think they are not. Straight-lining and excessive use of “don’t know” are understood in terms of respondents providing answers that are “good enough,” but deciding to end the survey halfway is decidedly not “good enough.” If different kinds of people choose to explicitly drop out of surveys than those who choose to satisfice by answering randomly or carelessly, our final survey demographics are likely to adjust for the former, but miss the latter, which may bias our substantive conclusions. More specifically, while breakoff behaviors may unbalance samples, or underrepresent people with certain psychological characteristics, shirking adds noise (which may be random or systematic depending upon instrument design), perturbing our ability to estimate unbiased behavioral correlations (like the role of economic perceptions in vote choice) for specific groups or the entire sample. Moving forward we examine breakoff and other within-survey shirking practices to fill this void in the current understanding and assess the extent to which survey respondents who belong to different genders, racial groups, age groups, education levels, and income levels differ in their propensity to engage in certain types of survey shirking.

Data and analysis

In late January and early February of 2016, we contracted Survey Sampling International[1] to send respondents to a Qualtrics instrument we had built for the purpose of collecting data on political economic perceptions and expectations. Respondents were directed to our instrument randomly from a bank designed for balance to census region, age, and gender. Our goal was a sample of around 2,500 complete, quality responses and we imbedded a series of questions designed to detect shirking amongst our respondents to ensure attentiveness. All told, 3,256 respondents were referred to our survey, 385 of these did not complete the instrument, and 253 failed one or more of our shirking "trap" devices. We acquired demographic information on all referrals from SSI's respondent bank — specifically their gender, age, race, education, household income, and personal employment status, which we will compare to their responses behavior. These separately acquired, validated demographic variables are important. If we were forced to rely on demographic information from our own instrument, they may be incomplete in the case of breakoff responders or inaccurate in the case of shirkers.

Our instrument contained three shirking-detection items, based on four questions that were distributed about the instrument at roughly the beginning, 1/3 mark, 2/3 mark, and end. The bookending questions asked the respondent to report their age in number of years and then their year of birth. Those who reported a year of birth that could not produce their reported age in years were marked as shirkers; 172 respondents.[2] The middle two questions were more classic attention devices, embedded in instructional preamble text, where a response was required in order to advance the instrument. The first question, read: "We are interested in how much information people have about politics. Please answer each of the following questions. If you don't know the answer, make your best guess. To show us you understand, please select 'triangle' below." The respondents were then given "triangle," "square," and "circle" responses to choose from; 92 did not choose triangle. The second question read: "For each of the following statements, please indicate how much you agree or disagree. To show us you understand, please select 'yellow' below." The respondents were then given "red," "green," "blue," and "yellow" responses to choose from; 25 did

[1] The firm has since rebranded as Dyanata.

[2] Note that these are *not* the age values used in the analysis below. Those are acquired directly from the survey firm.

not choose yellow. Both questions randomized response orderings.

Along with these questions that explicitly identify shirking via inattentiveness are questions on the survey that may allow us to identify worrisome behaviors that may also be plausibly sincere responses or manifestations of uncertainty. Specifically, our survey had multiple matrix format questions, where respondents were asked for their support of various proposals, or asked to evaluate the liberal-conservativeness of candidates participating in the presidential primary, etc. We identify respondents who gave the same response to all questions in a matrix as "straight-lining," a behavior associated with shirking that may potentially be a series of sincere responses, however improbable — we identified 142 straight-liners. Our survey also asked respondents to identify the party of the president, at that time Barack Obama, and gave only two alternatives, Democratic and Republican. Roughly 7% of our respondents (239), reported that a Republican controlled the presidency. We believe that this is a reasonable companion item to examine. Given the historic nature of the Obama presidency and the enduring association between Black Americans and the Democratic Party, we believe that this is a piece of information that nearly all adult Americans should possess, and so wrong answers here may be indicative of inattentiveness. That said, there are undoubtedly a fair number of American who simply do not have this information and it is certainly possible that our sample reflects this, so we are restrained in our interpretation here.

All told we have three distinct types of non-response, or shirking behaviors to analyze and compare: breakoff, shirking, and plausibly sincere, but unlikely responses. Below, we regress indicators for these response types on the demographic covariates measured above. Specifically, an indicator for female, age (from the survey bank) measured in years, an indicator for white (as compared to non-white), education measured on a 7-point scale (1 = no high school diploma; 7 = graduate degree), household income measured on a 9-point scale (1 = less than $\$10,000$; 9 = $\$250,000$ or more), and indicators for part time employed and unemployed, where full time employment is the baseline. Descriptive statistics are given in Table 1. Though our respondents were drawn from a representative bank, we retained all shirkers, calibrating our final sample of "quality responses" to the benchmark, rather than the total sample. As such, some characteristics of the total sample are demonstrably off-balance (e.g., it is 64% women). Nonetheless, we are not interpreting population effects for a particular treatment here, merely non-response probability across demographic subgroups groups (e.g., employed and unemployed), while holding other demographic factors constant,

Table 1: Descriptive statistics

Variable	Mean	SD	Min	Max
Female	0.644	0.4800	0	1
Age	45.098	16.429	18	96
White	0.753	0.430	0	1
Education	3.560	1.301	0	7
Employed Part Time	0.135	0.341	0	1
Unemployed	0.426	0.495	0	1
Income	4.763	2.533	1	9

statistically.

Before we could analyze the data, we first had to impute missing observations on the demographic characteristics. Across our whole sample, we possess complete information on gender and age, but some respondents elected to keep other information private: 73 observations on race, 87 on employment, 100 on education, and 223 on household income. These are much lower levels of missingness than on most large scale, trusted electoral surveys, like the ANES or CSES.[3] These missing values are imputed 10 times and the parameter estimates and substantive effects that we report below are the aggregation of 10 model estimates, one for each imputation following King et al. (2001) and Rubin (1987). This means that the standard errors and substantive effects convey our uncertainty about the estimated relationships and also the underlying measurement error. However, the reader need not interpret these estimates any differently than they would typical logistic regression estimates.

Table 2 presents our breakoff results, where we have calculated predicted effects for all of the parameters with 95% (two-tailed) confidence intervals in parentheses. We calculated these effects by holding all binary variables at 0 and all non-binary variables at their mean and making first-difference changes to the focal variable. The model results suggest that women are substantially more likely (almost 5%) to break off and that no other demographic characteristic even comes close to being as substantively significant. Prior work has shown that women are more likely to answer "don't know" in response to political knowledge questions (Mondak and Anderson 2004). Possible psychological underpinning of this might be due to an unwillingness to guess or not provide genuine responses. By extension, we think that these same motivations would make a female survey

[3]For example, this rate of missingness presents 15% better coverage on income than the ANES trend file.

Table 2: Breakoff

Variable	Estimate	Effect
Female	0.519	**0.046**
	(0.126)	**(0.024, 0.07)**
Age	0.004	0.005
	(0.004)	(-0.003, 0.015)
White	0.093	0.007
	(0.136)	(-0.014, 0.026)
Education	-0.144	**-0.012**
	(0.050)	**(-0.021, -0.004)**
Income	-0.013	-0.002
	(0.028)	(-0.012, 0.007)
Employed Part Time	0.047	0.004
	(0.175)	(-0.02 0, 0.031)
Unemployed	0.063	0.005
	(0.134)	(-0.014, 0.024)
Intercept	-2.097	
	(0.263)	
N	3256	
ln(likelihood)	-1163.461	

respondent more likely to break off during a survey than other forms of survey shirking that we consider like straight-lining which does not provide sincere answers to questions. Indeed, education is the only other significant predictor of breakoff, where increasing education reduces the probability of incomplete response, and the effect is roughly 1/4 the gender effect (the education effect comports with previous research).[4]

Table 3 analyzes the relationship between respondent demographics and responses that *may* indicate inattentiveness. There is a robust and negative correlation between age and both straight-lining and mistaking Obama's partisanship, but the effect is roughly twice as large for latter than the former. There is also a large and robust effect for education in straight-lining behavior. The effect of education on presidential partisanship just misses traditional levels of significance ($p = 0.08$). Taking the age and education results together the data suggest that the best predictor of these plausibly sincere, though potentially alarming responses, are the actual predictors of political knowledge, not necessarily the predictors of shirking, which we will see below.

Table 4 analyzes inattentiveness and, with one exception, the results are remarkably stable over each of the "traps." The data suggest that young, non-white, males are most likely to be

[4]See Peytchev (2009) for discussion of the mechanisms and explanation of the empirical variation in effects of education on survey breakoff.

Table 3: Plausibly sincere responses that may indicate inattentiveness

	Straightlining		Mistake Obama's party	
Variable	Estimate	Effect	Estimate	Effect
Female	0.193	0.008	-0.193	-0.011
	(0.192)	(-0.007, 0.023)	(0.144)	(-0.028, 0.005)
Age	-0.024	**-0.012**	-0.035	**-0.027**
	(0.006)	**(-0.018, -0.006)**	(0.005)	**(-0.037, -0.019)**
White	0.082	0.003	0.279	0.019
	(0.199)	(-0.013, 0.017)	(0.161)	(-0.003, 0.039)
Education	-0.234	**-0.010**	-0.113	-0.008
	(0.082)	**(-0.017, -0.003)**	(0.062)	(-0.018, 0.001)
Income	-0.039	-0.003	-0.045	-0.007
	(0.042)	(-0.011, 0.004)	(0.032)	(-0.016, 0.003)
Employed Part Time	-0.581	-0.015	-0.425	-0.021
	(0.315)	(-0.032, 0.001)	(0.236)	(-0.043, 0.001)
Unemployed	-0.150	-0.005	-0.097	-0.006
	(0.202)	(-0.019, 0.008)	(0.160)	(-0.024, 0.013)
Intercept	-1.163		-0.491	
	(0.407)		(0.318)	
N	3256		3256	
ln(likelihood)	-562.771		-816.8192	

careless. The effect of being male on failing at least one trap is an increase of over 5%. Non-white respondents are also more likely to issue a careless response, with the overall effect being nearly 6%. Similarly consistent, the models reveal that younger respondents are significantly more likely to issue inattentive responses, where a one standard deviation increase in age correlates to a 3% decrease in inattentive response, but the effect is not linear. Looking at the raw data, the rate of trap failure amongst respondents over 45 is about 5%, but for those under 45, the rate is 10%. That rate is 12% for respondents aged 25 years or less and for men aged 25 years or less that rate is 17%. Careless responses is pervasive amongst these young male respondents, which makes understanding their political behavior very difficult.

One interesting difference is the effect of income. While higher earners are less likely to fail the typical attentiveness traps, they are more likely to fail the age trap. The data reveal that this is partially a function of higher income respondents giving flip responses to the first age question, reporting that they are, for example, 4 years old (only respondents 18 and older were invited into our survey). This type of response is significantly less common amongst lower income respondents. There is also the regularity that the unemployed are much less likely to fail at least one trap question. One may speculate that this could partially be a function of them being more dependent

Table 4: Shirking

Variable	Yellow Estimate	Yellow Effect	Triangle Estimate	Triangle Effect	Age Estimate	Age Effect	Any Estimate	Any Effect
Female	-0.487	**-0.021**	-0.759	**-0.010**	-0.303	-0.024	-0.475	**-0.053**
	-0.219	**(-0.043, -0.002)**	-0.413	**(-0.029, 0.000)**	-0.163	(-0.052, 0.002)	-0.136	**(-0.087, -0.022)**
Age	-0.029	**-0.02**	-0.037	**-0.008**	-0.012	**-0.015**	-0.016	**-0.031**
	-0.008	**(-0.030, -0.011)**	-0.016	**(-0.017, -0.002)**	-0.006	**(-0.028, -0.001)**	-0.005	**(-0.047, -0.015)**
White	-0.561	**-0.023**	-1.288	**-0.014**	-0.485	**-0.036**	-0.533	**-0.059**
	-0.229	**(-0.049, -0.004)**	-0.453	**(-0.035, -0.003)**	-0.175	**(-0.066, -0.010)**	-0.145	**(-0.096, -0.026)**
Education	0.099	0.008	-0.265	-0.005	0.072	0.009	0.063	0.011
	-0.086	(-0.005, 0.022)	-0.226	(-0.017, 0.004)	-0.069	(-0.007, 0.027)	-0.058	(-0.008, 0.033)
Income	-0.092	-0.011	-0.295	**-0.010**	0.079	**0.019**	0.005	0.002
	-0.051	(-0.024, 0.001)	-0.123	**(-0.022, -0.002)**	-0.039	**(0.001, 0.040)**	-0.033	(-0.019, 0.024)
Employed Part Time	0.121	0.008	-1.736	-0.014	0.184	0.019	0.173	0.025
	-0.309	(-0.024, 0.049)	-1.062	(-0.037, 0.006)	-0.22	(-0.021, 0.066)	-0.193	(-0.026, 0.084)
Unemployed	-0.482	-0.02	-0.417	-0.006	-0.713	**-0.048**	-0.599	**-0.064**
	-0.271	(-0.045, 0.002)	-0.457	(-0.025, 0.009)	-0.212	**(-0.079, -0.020)**	-0.17	**(-0.102, -0.029)**
Intercept	-1.47		-0.003		-2.323		-1.204	
	-0.472		-0.909		-0.372		-0.303	
N	3256		3256		3256		3256	
ln(likelihood)	-124.1527		-397.5824		-643.4759		-850.2164	

on whatever compensation they are receiving for taking the survey; it may also be a function of the unemployed having more time. Amongst our respondents of retirement age (65 years or greater), 78% were not working, which is well over twice the rate of those under 65. We suspect that differentiating retirement from unemployment status could shed more light on what is driving this correlation, but, unfortunately, we do not have that information on hand.

What are the potential effects of failing to account for breakoff and shirking? As discussed, breakoff leads to the unbalancing of samples or underrepresentation of respondents with certain characteristics. As such, it makes estimating population traits or subgroup behaviors onerous, but does not necessarily bias the estimation of correlations in controlled analyses. Shirking, on the other hand, has more pernicious effects, particularly given its concentration among certain population subgroups. As an example, let us examine the correlation between partisanship and economic retrospections — one of the strongest predictors of vote choice (e.g., Duch and Stevenson 2008).

A well-established finding in election and opinion research is that evaluations of economic outcomes under an executive tend to be higher for in-partisans than out-partisans (e.g., Evans and Pickup 2010). We should therefore expect that Democrats are more likely to report that unemployment has improved under Obama's leadership than Republicans. Let us say that we had some theory of economic retrospections that led us to hypothesize that women were more inclined toward motivated reasoning than men (we do not believe this to be the case, it is merely illustrative). This would manifest in interaction terms between gender and partisanship that showed moderating effects of being male on partisan rationalization — that the constituent term on being a Democrat (Republican) was positive (negative), but the interaction of being male on being a Democrat (Republican) was negative (positive). This is what column 2 of Table 5 shows.

Estimating the substantive effect: the net effect of being male on Democratic responses is null (note the large, positive coefficient on the male constituent term), however, the probability of male Republicans reporting the economy has improved under the Obama administration is 0.12 (0.06, 0.18) greater than female Republicans. However, we know that men are more likely to shirk than women, and shirking should have a moderating effect on this relationship — where a strong correlation is truly present, groups with higher levels of shirking will display a weaker correlation than groups with lower shirking levels as more random or invalid responses will drive the estimated relationship toward 0. This is what we see in column 3 of Table 5. Estimating the substantive effect: the probability of non-shirking Democrats reporting that unemployment has improved under Obama is 0.15 (0.04, 0.26) greater than shirking Democrats, while the net effect of shirking is null for Republicans (shirkers are significantly less likely to report that unemployment improved — likely due to the static response category ordering, where decline preceded improvement).

We see two central solutions to this problem, *given* that the researcher has built tools into their instrument to discover shirking. The first is to discard the offending respondents. This is the most simple and direct, but also the most costly, particularly in light of the reality that there are bound to be both false positives and false negatives in any detection device — mistakes happen. The second is to account for shirking behaviors statistically. There are at least two paths here. One would be to include a shirking covariate in the model directly. This is the most simple and direct, but also the most onerous for interpretation and explanation, particularly in non-linear modeling frameworks. The second is to allow shirking estimates to inform the weights put on observations. This is potentially less simple and direct in application — the researcher must chose a weighting scheme — but also more economical and substantially more simple for interpretation and explanation. Such results are given in column 4 of Table 5.[5] The recovered results are substantially closer to the partisan constituent terms from the shirking interacted (column 3) model than the unconditioned estimates (column 1). Though the results also show that the standard errors are slightly less efficient, we see this as a worthwhile tradeoff in reducing the bias induced by inattentive

[5] The weights here are $1 - \frac{1}{1+n(\text{trap fails})}$.

Table 5: Potential bias in recovering motivated reasoning effects. Dependent variable indicates respondent believes unemployment improved under Obama administration.

Variable	Unconditioned	Gender	Shirking	Weighted
Democratic	0.665	0.789	0.703	0.696
	(0.089)	(0.112)	(0.093)	(0.094)
Republican	−0.811	−0.995	−0.838	−0.862
	(0.105)	(0.143)	(0.109)	(0.111)
Male		0.335		
		(0.131)		
Male × Democratic		−0.344		
		(0.185)		
Male × Republican		0.339		
		(0.214)		
Shirker			−0.216	
			(0.232)	
Shirker × Democratic			−0.409	
			(0.321)	
Shirker × Republican			0.366	
			(0.415)	
Constant	−0.428	−0.549	−0.411	−0.332
	(0.063)	(0.079)	(0.066)	(0.067)
N	2,908	2,908	2,908	2,908
ln(likelihood)	-1,853.920	-1,842.740	-1,849.379	-1,563.099

responding.

Ideally, surveyors and scholars running institutional electoral research projects, like the ANES or CSES, would include an array of reasonable attention checks and make a joint decision on an appropriate weighting scheme and build the weights into the data archives. This would allow a (transparent) standard for all consumers of those data products, while still allowing individual researchers to pursue alternative strategies tailored toward their particular application.

Discussion

What can these results tell about electoral research, or survey research more generally, and how can they help us contextualize previous findings? It is clear that, although women may be more expensive to survey, as a function of their higher breakoff rates, that their responses, on average, are of a substantially higher quality than their male counterparts. And this distinction between incomplete responses and inattentive responses is very important for the applied researcher. Given that survey firms typically deliver only completed data, the firms are doing a large part of the work in sorting out corrupted responses, but only one type, where remaining troubled responses are inordinately concentrated among young men and non-whites (in the case of US sampling). In other words, there are real, robust, quantitative and qualitative differences between respondents who are likely to drop out of surveys and the respondents that are likely to give inattentive responses. These differences may bear implications for our substantive conclusions from survey research if the inattentive are not successfully sorted out or otherwise accounted for. How can come to understand the political behavior of young men, if they are systematically more likely to thwart our efforts through inattention?

One such example comes from Prior and Lupia (2008) who offer financial incentives to survey respondents in an attempt to incentivize them to perform better on a political knowledge battery. Comparing those who were paid for correct responses to those who were unpaid, the authors find a large effect for men, but no effect for women. Given what we have learned here, part of this difference is likely to have come from purchasing attentiveness away from men's natural state of inattentiveness, rather than, say, incentivizing extra effort.

Ideally, researchers would have adequate resources to use financial incentives to get higher quality responses, however, this is often not the case. With this in mind, we offer two suggestions for survey practitioners to identify issues all types of nonresponse and alleviate issues associated with biased inferences. First, when designing surveys researchers should employ the use of attention checks throughout their survey in order to identify in what part of the survey and which respondents are using different satisficing behaviors. Second, researchers should report results that keep and drop inattentive respondents to make appropriate adjustments and careful interpretations of the substantive results in their work.

More broadly, the different patterns of results for breakoff and our various measures of survey inattention suggest that the psychological roots of these behaviors are quite different. While straight-lining and shirking have typically been understood in terms of satisficing (Krosnick 1991), past work has found that respondents who breakoff early are not inattentive (Peytchev 2009) and our findings reinforce this conclusion. Women are dramatically more likely to leave our surveys early than men even as they are much less likely to fall prey to inattention. This discrepancy is worthy of further scrutiny, especially as the motivational basis of breakoff is poorly understood. Leveraging known differences between men and women (e.g. in terms of personality traits) could help shed light on why people choose to break off early from surveys.

References

Conrad, Frederick G, Michael F Schober, and Wil Dijkstra. 2008. "Cues of Communication Difficulty in Telephone Interviews." *Advances in telephone survey methodology*: 212-30.

Duch, Raymond M., and Randolph T. Stevenson. 2008. *The economic vote: How political and economic institutions condition election results.* New York: Cambridge University Press.

Evans, Geoffrey, and Mark Pickup. 2010. "Reversing the causal arrow: The political conditioning of economic perceptions in the 2000-2004 US presidential election cycle." *Journal of Politics* 72: 1236-1251.

Galesic, Mirta. 2006. "Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey." *Journal of Official Statistics* 22: 313.

Galesic, Mirta, and Michael Bosnjak. 2009. "Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey." *Public Opinion Quarterly* 73: 349-60.

Groves, Robert M, Stanley Presser, and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68: 2-31.

Groves, Robert M, Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration." The *Public Opinion Quarterly* 64: 299-308.

Heerwegh, Dirk. "Web Surveys. Explaining and Reducing Unit Nonresponse, Item Nonresponse and Partial Nonresponse." University of Michigan, USA, 2005.

Knauper, Barbel. 1999. "The Impact of Age and Education on Response Order Effects in Attitude Measurement." *Public Opinion Quarterly*: 347-70.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied cognitive psychology* 5: 213-36.

Krosnick, Jon A, Allyson L Holbrook, Matthew K Berent, Richard T Carson, W Michael Hanemann, Raymond J Kopp, Robert Cameron Mitchell, Stanley Presser, Paul A Ruud, and V Kerry Smith. 2002. "The Impact of" No Opinion" Response Options on Data Quality: Non-Attitude Reduction or an Invitation to Satisfice?" *Public Opinion Quarterly* 66: 371-403.

Mondak, J.J. and Anderson, M.R., 2004. "The knowledge gap: A reexamination of gender-based differences in political knowledge." *The Journal of Politics* 66: 492-512.

Mooney, Geraldine M, Barbara Rogers, and Deborah Trunzo. 2003. "Examining the Effect of

Error Prompting on Item Nonresponse and Survey Nonresponse in Web Surveys." Paper presented at the Proceedings of American Association for Public Opinion Research, 58th Annual Conference, Nashville, Tennessee.

Peytchev, Andy. 2009. "Survey Breakoff." *Public Opinion Quarterly* 73: 74-97.

Peytchev, Andy. 2011. "Breakoff and Unit Nonresponse across Web Surveys." *Journal of Official Statistics* 27: 33.

Prior, Markus and Arthur Lupia. 2008. "Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills.? *American Journal of Political Science* 52: 169-183.

Steinbrecher, Markus, Joss Roßmann, and Jan Eric Blumenstiel. 2014. "Why Do Respondents Break Off Web Surveys and Does It Matter? Results from Four Follow-up Surveys." *International Journal of Public Opinion Research* 27: 289-302.