



Journal of Behavioral Public Administration Vol 3(2), 2020, doi: 10.30636/jbpa.32.182

Is self-reported social distancing susceptible to social desirability bias? Using the crosswise model to elicit sensitive behaviors

Ulrich Thy Jensen

School of Public Affairs, Arizona State University

&

Crown Prince Frederik Center for Public Leadership, Aarhus University

Address correspondence to Ulrich Jensen at ujensen@asu.edu.

Is self-reported social distancing susceptible to social desirability bias? Using the crosswise model to elicit sensitive behaviors

Abstract:

Sensitive behaviors such as self-reported performance or (un)ethical behaviors often carry strong social connotations of appropriate or inappropriate conduct. In return, social norms can artificially inflate or deflate individuals' responses and bias scientific results on their prevalence and effects. As a core part of governments' mitigation strategy against the outbreak of COVID-19, social distancing might represent one of these behaviors. Can researchers expect honest responses when surveying citizens about their social distancing behaviors? This question is examined using the sensitive survey technique, "the crosswise model", to elicit aggregate-level prevalence estimates of (1) self-reported social distancing, and (2) honest reporting in a prediction dice game. Since the number of wins in the dice game follows a known probability distribution, it offers an excellent setting for illustrating the utility of the crosswise model before applying it to self-reported social distancing. In a survey of 1,059 adults living in the US, the crosswise model outperforms direct questioning in revealing respondents' dishonest behavior in the dice game. While the crosswise model also indicates some social desirability bias when asking respondents directly about their social distancing behaviors, the extent of this bias seems small and does not appear to overtly inflate individuals' self-reported measures of social distancing.

Keywords: COVID-19, Social distancing, Cheating, Survey sensitivity, Crosswise model

Social scientists have long taken interest in phenomena that are inherently embedded in social norms about appropriate and inappropriate action. Issues such as tax evasion, corruption, unethical behaviors, performance, and many others are prime examples (Alm 2012; Boyne et al. 2005; Bozeman et al. 2018; Menzel 2015). While researchers often turn to surveys to capture the manifestation and prevalence of such phenomena, their social context remains a critical concern for the validity of scientific findings (Zerbe and Paulhus 1987).

A brief illustration exemplifies the issue at hand: A research team is keen on understanding citizens' behavioral responses to the current outbreak of the novel coronavirus and related COVID-19 disease (e.g., Fetzner et al. 2020; Pedersen and Favero 2020). The researchers are particularly interested in the extent to which citizens engage in social distancing and what predicts their actions. This is informative because social distancing has been touted as a quintessential part of governments' mitigation strategies to the pandemic. The team fields a survey asking a sample of citizens about their willingness to stay at home, cancel upcoming birthday parties etc. They also ask about citizens' retrospective behaviors like: Did you leave your house within the last couple of days for non-essential purposes? Should the research team expect to receive honest and truthful responses to such questions? We don't know. The challenge is that we might expect social distancing to carry strong socially desirable connotations exactly because it represents a core part of governments' mitigation strategies. "Stay at home!" was the central message across press briefings and news coverage in the Spring of 2020, and remains so in several states as confirmed cases continue to surge across parts of the US. Since social distancing is touted as the appropriate behavior and the responsible way to act to help slow the spread of COVID-19, the social values placed on

such behaviors might dictate individuals' responses to the research team's survey rather than reflect people's true preferences or actions. For this reason, we might be concerned that citizens depict their behaviors in ways that reconcile with social norms to promote social distancing; especially when asked explicitly to reveal their status on the sensitive behaviors making up social distancing, such as staying at home and refraining from social interaction in close physical spaces.

This article demonstrates how an established sensitive survey technique, the "crosswise model" (Yu et al. 2008), can be useful to examine the extent to which sensitive measures, like questions about social distancing, are likely to suffer from social desirability bias. The purpose here is not to formally introduce, review, or validate the crosswise model as several existing studies offer excellent reviews and validations (Jann et al. 2008; Höglinger and Jann 2008), but rather to showcase its utility to public administration research where it has gained little traction. In doing so, I use a two-tier approach. First, I implement a similar design to that of study 1 in Olsen et al. (2019) using a prediction dice game that incentivizes respondents to cheat for personal gain. Since respondents are not asked to reveal their true prediction, scholars are unable to verify individual wins or losses. Given equal probability of outcomes (1-6) of each die, however, I can calculate the aggregate "cheat rate" and expected number of wins based on this known distribution, to compare it to the sensitive survey technique's ability to elicit truthful responses when respondents are asked to report their (dis)honest behavior in the game. After illustrating the usefulness of the crosswise model in the prediction game, I adopt the same technique to examine the degree to which questions

about citizens' retrospective social distancing behaviors are likely inflated by social desirability bias.

Based on a sample of 1,059 adults living in the United States recruited via Amazon's Mechanical Turk, I show that the crosswise model significantly outperforms directly asking respondents whether they reported their predictions truthfully in the dice game. This result showcases the utility of the crosswise model on sensitive behaviors. While the article cannot rule out that self-reported measures of citizens' retrospective social distancing suffers from some social desirability bias, the crosswise model only shows modest efficiency gains compared to direct questioning. If corroborated by studies with larger and more diverse samples, this result can help mitigate concerns that survey measures of social distancing are highly susceptible to social desirability bias and offer support for the validity of surveying individual citizens directly about such behaviors.

The Crosswise Model: What is it Good for and How Does it Work?

The crosswise model builds on a simple idea: Protecting individual respondent's status on a sensitive question will elicit a more truthful answer than asking the respondent directly to reveal their status on the sensitive question. The model is a variant of the randomized response technique framework (Yu et al. 2008), and guarantees respondents' privacy by bundling the response to two questions together. More specifically, the crosswise model presents respondents with the sensitive question alongside an unrelated non-sensitive question with two response options: (A) Yes (or no) to both questions, or (B) Yes to one question, but no to the other. By asking respondents to provide a joint response to the two

questions, researchers are unable to disentangle the status on the sensitive question from the status on the non-sensitive item. The trick of this model is to make sure the non-sensitive question is unrelated to the sensitive items and has a known probability distribution different from 0.5. Under these conditions, an aggregate-level prevalence estimate for a “yes” to the sensitive item can be calculated using the following formula as shown by Höglinger and Jann (2018, 9):

$$\hat{\pi} = Pr(Y^* = 1) = \frac{Pr(Y=1)+p_z-1}{2 \cdot p_z - 1}$$

, where Y^* is the unobserved answer to the sensitive question, Y is the observed joint answer to the sensitive and non-sensitive question, and p_z the known probability of a “yes” to the non-sensitive question. As shown by Jann et al. (2012), the sampling variance of the prevalence estimate can be calculated using the following formula:

$$SE(\hat{\pi}) = \sqrt{\frac{Pr(Y=1) \cdot (1 - Pr(Y=1))}{n \cdot (2 \cdot p_z - 1)^2}}$$

The prevalence estimate obtained for the sensitive behavior can then be compared to the prevalence estimate obtained when respondents are asked the sensitive question directly without privacy. Working under the “more-is-better” (or “less-is-better”) assumption, we expect to reveal higher (or lower) prevalence estimates for the sensitive question in case of the crosswise model compared to direct questioning. For instance, in the case of social distancing we would expect potential social desirability bias to artificially inflate affirmative

responses resulting in an overreporting. If social desirability bias is a concern, we would therefore expect the crosswise model to produce lower estimates for social distancing behaviors compared to direct questioning. Whether we should expect affirmative responses to be inflated or deflated as a function of social desirability bias naturally depends on the wording of the specific sensitive behavior question as we will see below.

A number of comparative validation studies have evaluated the ability of the crosswise model to elicit more truthful answers compared to direct questioning for sensitive behaviors. For instance, in a study of student plagiarism, Jann, Jerke, and Krumpal (2012) found that 22.3 % of students admitted to partial plagiarism behaviors compare to only 7.3 % when asked directly. In a study of tax evasion, Korndörfer and colleagues (2014) report that 16.7 % admitted to evading taxes sometime in the last 10 years when asked directly. This number was 27.8 % when using the crosswise model, with the difference in prevalence estimates across the two approaches being statistically significant. Building on these studies, I show how the crosswise model can be applied to elicit the prevalence of (dis)honest behavior in an online dice prediction game. After benchmarking the crosswise model against the direct questioning approach, the same technique is used to test for social desirability bias in self-reported social distancing behaviors.

Study Overview and Design

The study is designed as a between-subjects experiment embedded in an electronic survey to 1,059 adults living in the United States. In the experiment, respondents were randomly assigned to one of two approaches for asking sensitive questions about (1) cheating in a dice

game and (2) past social distancing behaviors. One group of respondents was asked the sensitive questions directly; as is commonly done in survey research. The second group was asked the same questions using the crosswise model where each question was bundled with an unrelated non-sensitive question. Descriptive statistics and test for differences in respondent characteristics across the two groups is presented in Appendix A. Before I outline each of these survey approaches in more detail, I first discuss the data collection and sample.

Mechanical Turk: Sample and data quality checks

The survey data was collected between April 6–8, 2020 via Amazon’s Mechanical Turk (Mturk). Mturk provides an extensive online labor market platform that has rapidly become a central part of the methodological toolkit among social scientists (Buhrmester et al. 2011) including PA scholars (Stritch, Pedersen and Taggart 2017). Mturk offers many advantages including quick turnaround times, affordable pricing, and opportunities to construct panels.

However, multiple concerns have also been raised, most notably about the quality of the data generated from these convenience samples. To limit access for people outside the US masking their location with VPN/VPS services, I followed the protocol by Burleigh and colleagues (2018). A JavaScript was implemented to strip the respondents’ IP address at the beginning of the survey and run it against known IP addresses using a third-party service (IPHub). 140 individuals were screened out using this approach. I also used a recaptcha verification mechanism to detect and prevent automated non-human/bot respondents from taking the survey. Finally, 98 individuals did not complete the survey’s last section on honest

reporting and social distancing. All 98 individuals exited the survey prior to being randomized to the sensitive survey questions, making it very unlikely that attrition is a function of the experiment reported here. The final sample is made up of 1,059 individuals.

A second concern levelled against Mturk is that convenience samples do not adequately reflect the general population, raising questions about the external validity of findings generated using this platform. As shown in Table A1, the sample shows great variation along several key demographic characteristics. 42 % are women, the mean age is 38.3 years old, and 60 % identify as liberal. Recent studies have replicated identical experiments on Mturk and national samples with largely similar results to follow (e.g., Coppock 2019; Mullinix et al. 2015), indicating that “Turkers” might not display attitudes or behaviors that – on average – differ fundamentally from the those of the broader population. While my sample therefore is not representative of the broader US population (e.g., younger and more liberal), it does afford us the opportunity to explore the extent to which survey measures of social distancing are susceptible to social desirability bias among a diverse group of adult members of the US public.

Prediction dice game

As part of the survey and prior to the sensitive questions, all respondents were asked to play 40 rounds of an online prediction dice game. The design of the game is largely similar to the one reported in Olsen et al. 2019 and has been validated against real-world behaviors (Barfort et al. 2019; Cohn and Maréchal 2017; Hanna and Wang 2017). In the dice game, respondents were asked ahead of each round to make a prediction with regard to the

outcome of a die roll (1-6). They were then instructed to roll the virtual die, observe its outcome, and report whether their prediction matched the outcome or not. If yes, the participant won \$0.05. If no, the participant received no additional compensation. While this incentive might seem negligible, prior research indicate that prevalence of dishonest behavior in dice games is quite insensitive to the size of the incentive (Olsen et al. 2019; Abeler et al. 2019). The important feature of the design is thus to create an opportunity for people to reveal their (dis)honest behavior or cheating for personal gain. Respondents had the opportunity to win an additional total \$2.00 on top of the base compensation of \$1.50 for completing the survey. The compensation was calculated to meet the federal minimum wage of \$7.50 in the absence of a win rate of 1/6. All respondents were therefore compensated a minimum of \$1.80.

In contrast to other versions of the dice game, cheating cannot be verified or detected at the individual level. However, having individuals play multiple rounds provides us a win-rate to compare against the “true” win-rate of 1/6, assuming an equal probability function for the outcome of the die. We can thus calculate an observed estimated cheat rate for the sample using the formula provided by Barfort et al. (2019, 105) and compare it to self-reported honesty (or cheating). Data yields a cheat rate of ~ 0.43 which is fairly similar to that reported in Olsen et al. (2019) for Danish students (0.38) and that of others studies (e.g., Hanna and Wang 2017), bolstering our confidence in the implementation of the game and the credibility of this design. The observed dishonest behavior gives us a benchmark for illustrating the usefulness of the crosswise model, not only in comparison to a direct questioning approach, but in reflecting individuals’ actual dishonest behavior.

After the dice game, respondents were randomly exposed to a sensitive question on whether they reported the outcomes honestly in the dice game. Respondents in the first experimental condition were asked the question directly: “In the dice game, did you always honestly report whether your prediction of the dice roll was right?”. Respondents had the response options “yes” or “no” yielding a binary outcome variable affirming honest reporting or not. The response distribution to this question provides us the extent to which people report cheating when asked directly while not provided any privacy in relation to the sensitive behavior.

Respondents in the second experimental condition were asked the same sensitive question using the crosswise model. The sensitive question was therefore bundled with a second unrelated, non-sensitive question for which we have a known distribution: “Is your mother's birthday in January or February? (if you don't know please use the birthday of another family member or a good friend)”. Respondents were given two response options: (1) Yes (or no) to both, or (2) yes to one question, but no to the other. This procedure guarantees complete privacy to the individual respondent as researchers cannot disentangle the status on the sensitive question from the status on the non-sensitive question. We can, however, calculate the aggregate-level prevalence estimate for the sensitive question using the formula listed above.

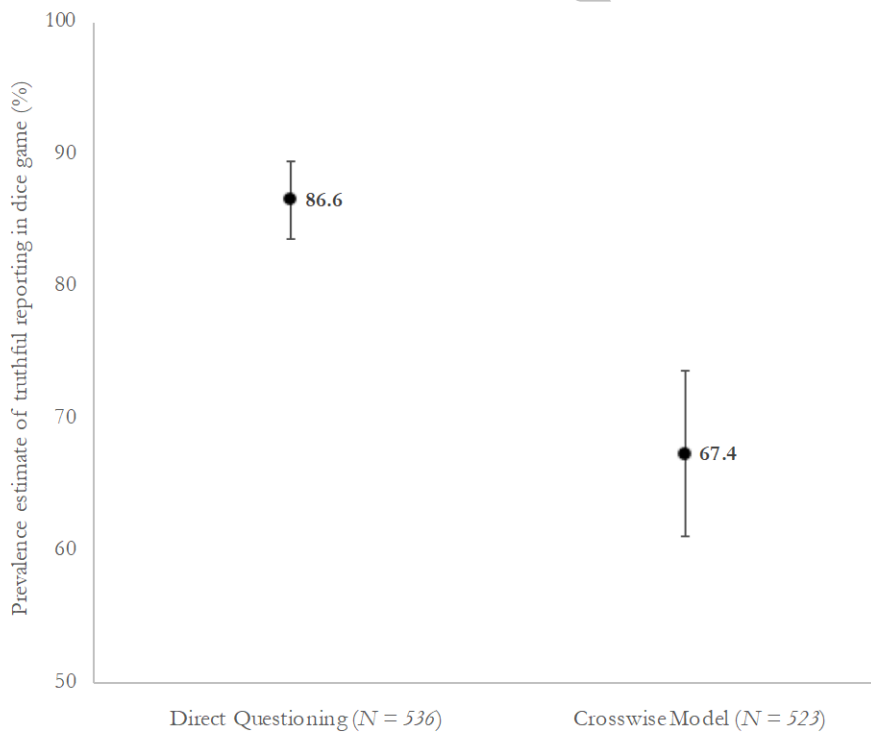
Using the Crosswise Model to Estimate Dishonest Behavior

Figure 1 reports the estimated prevalence of truthful reporting in the dice game. For respondents that were asked directly whether they always reported honestly in the dice game,

an estimated 86.6 % (SE = 1.5) reported that they always provided a true account of whether the roll of the die matched their prediction. Only 13.4 % of the respondents thus admitted to cheating. As expected, the crosswise model produces a much smaller prevalence estimate with an estimated 67.4 % (SE = 3.2) always reporting truthfully. A two-sample z test for equality of means shows that the crosswise model significantly outperforms the direct questioning approach with 19.2 % (SE = 3.5, $z = 5.4$) more respondents admitting to cheating.

Figure 1

Prevalence estimates of truthful reporting of wins in prediction dice game

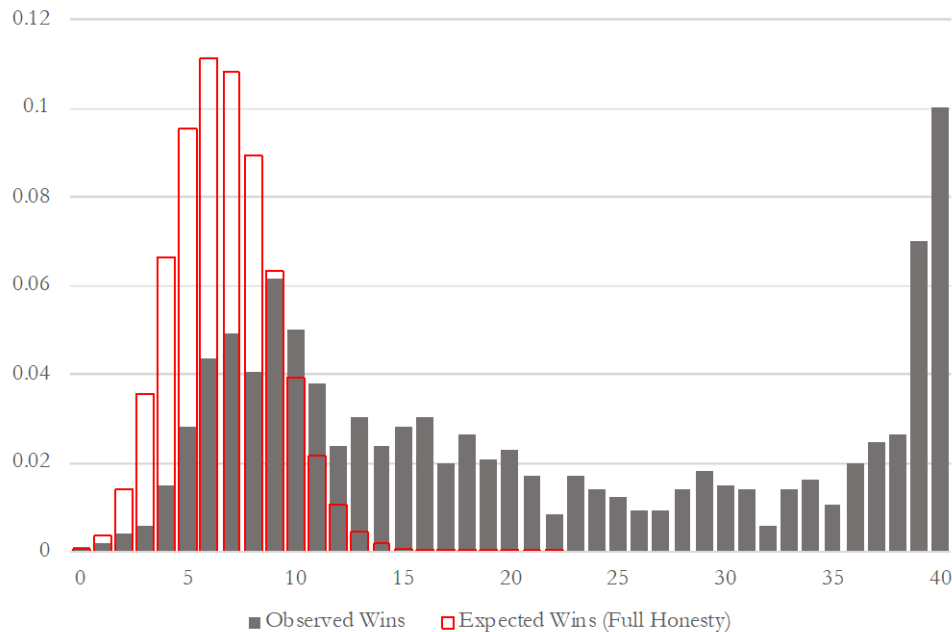


Notes: Point estimates depict the prevalence of always reporting wins in dice game truthfully for the direct questioning approach and the crosswise model. Bars show 95 percent confidence intervals.

Overlaying the expected number of wins with the observed number of wins yields an estimate of the true honest reporting of 35.9 % (or 64.1 % showing dishonest behavior by cheating). Figure 2 plots the two distributions. If we assume that people never underreport actual wins, the overlapping areas of the shaded bars (observed wins) and the red hollow bars (expected wins) indicate that just over one-third of all respondents report truthfully on the match between the outcome of their die and their prediction. While the crosswise model thus shows substantial efficiency gains over the direct questioning of almost 20 percentage points, its prevalence estimate is still double that of the “true” prevalence of people reporting truthfully in the dice game.

Figure 2

Observed wins vs expected wins under full honesty in prediction dice game



Notes: Solid bars depict observed wins as self-reported by respondents. Hollow bars depict the expected probability of wins under assumption of full honesty for fair 6-sided die over 40 independent rolls.

Is Self-reported Social Distancing Behavior Susceptible to Social Desirability Bias?

Figure 3 reports the estimated prevalence of social distancing in percentage. Similar to the structure of the questions on cheating in the dice game, one group of respondents was randomly assigned to provide a direct response to the question: “Have you at one or more times during the past four days left your house/apartment for a non-essential purpose?”. The question was accompanied by a short statement clarifying that “non-essential purposes commonly include going to work or engage in schooling outside the home if you don't have

to or can work remotely, gathering in large social groups outside the home or inviting family and friends over who don't reside in your home, leaving the house for shopping (other than for groceries), dining out at restaurants, bars etc.”. Respondents could either respond (1) no, (2) yes – once, (3) yes – a couple of times, or (4) yes – at least 5 times. For the purpose of this analysis, the three affirmative categories are combined to make a binary, “yes”/”no”, response variable. The response distribution to this question reveals the extent to which people report leaving their home for a non-essential purpose – and thus failing to comply with a core feature of social distancing – when asked directly without any guaranteed privacy on the sensitive behavior.

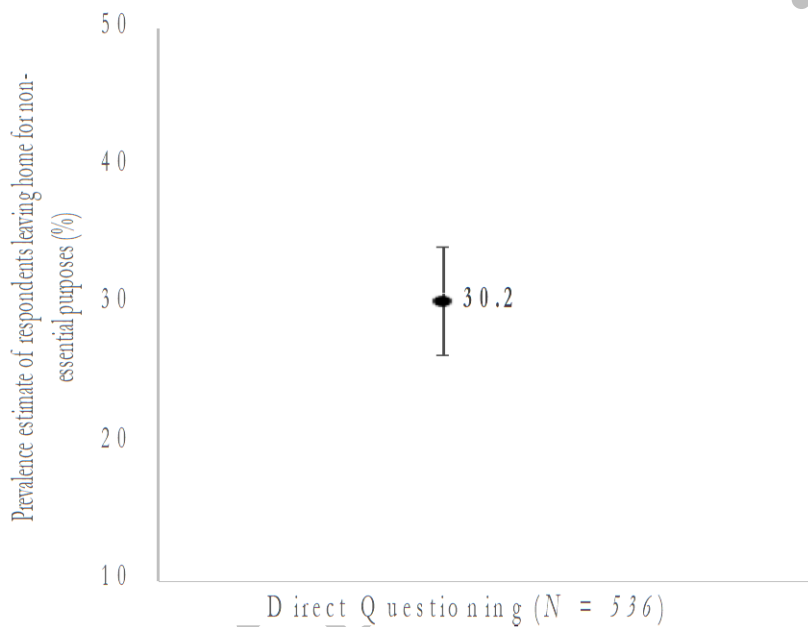
Respondents in the second experimental condition were asked the same question but bundled with a second unrelated, non-sensitive question with a known distribution: “Is your father's birthday in January or February? (if you don't know please use the birthday of another family member or a good friend)”. Again, respondents could either respond “yes (or no)” to both, or “yes to one question, but no to the other”. While the crosswise model ensures privacy for the individual respondent, we can calculate the aggregate-level prevalence estimate for the presumed sensitive question – self-reported social distancing.

If self-reported social distancing behaviors are susceptible to social desirability bias, asking respondents directly should deflate or underestimate the number of people admitting to leaving their home for non-essential purposes. Figure 3 reveals a fairly small efficiency gain from the crosswise model. When asked directly, 30.2 % (SE = 2.0) of respondents admit to leaving their home for non-essential purposes while the prevalence estimate increases to

37.0 % (SE = 3.2) when the crosswise model is used. The difference of 6.7 % (SE = 3.8, $z = 1.7$) is statistically significant as a one-sided two-sample z test of equality of means ($p = 0.038$) but not for a two-sided test ($p = 0.076$).¹

Figure 3

Prevalence estimates of leaving one's home for non-essential purposes



Notes: Point estimates depict the prevalence of leaving one's home for non-essential purposes for the direct questioning approach and the crosswise model. Bars show 95 percent confidence intervals.

Discussion and Conclusion

This article offers a double-sided contribution. The first contribution is generic in nature and illustrates how research on a host of behaviors embedded in strong social norms about appropriate conduct can be advanced using the “crosswise model” – an established sensitive survey technique. Sensitive behaviors span longstanding issues of corruption, unethical behaviors, and performance (Alm 2012; Boyne et al. 2005; Bozeman et al. 2018; Menzel 2015), as well as more recent phenomena like people’s actions to physically distance themselves from others in times of a pandemic (Van Bavel et al. 2020). The second contribution is more specific to the current COVID-19 public healthcare crisis and illustrates that while self-reported measures of social distancing do display some social desirability bias, the extent of this bias seems small; and possibly smaller than researchers might have initially expected.

Sensitive behaviors share a common feature. People who engage in corrupt behaviors are likely to suppress the report of such activities when asked directly by researchers. When asked to report how well their organization is doing, managers are likely to paint a rosier picture of the current state of affairs. Whether respondents suppress or exaggerate the behavior in question, the key question for researchers remains: How do we elicit more accurate reflections of people’s true actions when it comes to behaviors that are sensitive to social norms?

In this article, I illustrate the usefulness of a well-known technique – the crosswise model – that has gained little traction in public administration research. The “trick” of the model is to

guarantee individual respondents' privacy on the sensitive behaviors by bundling the respondent's status on this question with the response to a second, unrelated non-sensitive question with a known distribution (Jann, Jerke and Krumpal 2012). This enables us to assess the degree of social desirability bias for the direct questioning approach at the sample level and guide our intuition about the appropriateness of measures to capture individual self-reported behaviors.

Using the crosswise model to elicit the prevalence of honest reporting in an online dice game similar to that of Olsen and colleagues (2019), I find about 67 % self-report honest reporting of wins, or 33 % admit to cheating. However, only ~13 % admit to cheating when asked directly. The crosswise model thus represents a quite powerful tool for eliciting sensitive behaviors compared to direct questioning, albeit it falls short of the estimated ~64 % of respondents who cheat when we compare respondents' self-reported number of wins to the expected number of wins using the probability of 1/6 chance of winning in each round over 40 rounds.

While the crosswise model reveals substantial social desirability bias in reporting (dis)honest behavior, social desirability bias in respondents' self-reported social distancing behavior seems less prevalent. When asked directly, about 30 % of respondents admit to leaving their home for non-essential purposes, while this number increases to about 37 % when I use the crosswise model. While the estimated effect of 7 % bias is not negligible, this difference is only statistically significant by a one-sided z test, and could thus indicate that social desirability bias is less of a threat to self-reported measures of social distancing behaviors than

researchers might have initially expected. However, affirmation of this conclusion is pending replications with larger and more diverse samples than was feasible to obtain as part of this study.

While I am unable to verify respondents' true social distancing behaviors – and thus evaluate the absolute effectiveness of direct questioning and crosswise model approaches – some recent findings bolster our confidence in the results presented above. In a study of American adults, Gollwitzer and colleagues (2020) report correlations between self-reported social distancing and actual movement data at the individual and state-level. These findings are important because they utilize smartphone tracking data to observe individuals' real-world actions, and thus move us beyond mere intentions or recall of self-reported behaviors. The results also corroborate the findings presented here that self-reported social distancing behaviors – despite a well-founded concern for susceptibility to social desirability bias – appear to be a viable approach for survey research. This is critical as many research projects aimed at estimating, understanding, and impacting citizen's social distancing behaviors around the world heavily rely on surveys (e.g., Fetzer et al. 2020; <https://hope-project.dk>).

Another limitation of this study is its rather crude measure of social distancing. I rely on a one-item compound measure to capture “staying at home”, arguably one of the most central facets of social distancing. However, many other behaviors are critical for governments' mitigation strategies to work. People, for example, might have friends or family come over to visit. Here, a respondent would comply with social distancing as captured by my broad measure, but not by more fine-grained measures. An interesting observation here is that the

broad self-reported social distancing measure correlates positively with respondents' *intention* to engage in other social distancing behaviors such as wearing a mask whenever in public ($r = 0.23$) and cancelling private get-togethers with close friends ($r = 0.26$). Other findings from ongoing research offer more reassurance. In a study of a representative sample of adults in Denmark, Larsen and colleagues (2020) find no measurable difference in self-reported social distancing like visiting or getting a visit from a friend among respondents asked directly versus using a list experiment – another sensitive survey technique. Despite these encouraging observations, researchers are strongly encouraged to replicate and extend this article's findings using a broader range of more fine-grained measures of social distancing as well as recruiting larger samples to offer more conclusive evidence on the extent to which social desirability bias influences citizen's self-reported social distancing behaviors.

Notwithstanding the limitations of this study, this article offers two important contributions with implications for both scholarship and practice. First, it illustrates the application and usefulness of the crosswise model as a tool in public administration and management researchers' toolbox. Its relatively simple application and quite powerful ability to detect and reduce social desirability bias makes it apt for estimating and understanding the prevalence of sensitive behaviors like corruption, cheating, or other unethical behaviors. Second, the article demonstrates that self-reported social distancing – in contrast to cheating in a dice game – likely suffers from limited social desirability bias; and possibly less than researchers might have initially expected. This is encouraging news for researchers and policy makers

who conduct and rely on survey measures of social distancing as central parts of designing and revising mitigation strategies in the fight against COVID-19.

Acknowledgment

I appreciate the valuable comments to earlier versions of this article from Asmus Olsen and Justin Stritch as well as the insightful and helpful suggestions from two anonymous reviewers and the editorial team at JBPA.

Notes

1. A sensitivity analysis shows a minimum detectable effect size of 0.15 using the study parameters of sample group sizes of 536 and 523, an error rate of 0.05 and power of 0.80. The estimated observed effect size is only half of that and falls shy of the lower common threshold of a small effect of 0.1.

References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115-1153.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. *International Tax and Public Finance*, 19(1), 54-77.

Barfort, S., Harmon, N. A., Hjorth, F., & Olsen, A. L. (2019). Sustaining honesty in public service: The role of selection. *American Economic Journal: Economic Policy*, 11(4), 96-123.

Boyne, G. A., Meier, K. J., O'Toole Jr, L. J., & Walker, R. M. (2005). Where next? Research directions on performance in public organizations. *Journal of Public Administration Research and Theory*, 15(4), 633-639.

Bozeman, B., Molina Jr, A. L., & Kaufmann, W. (2018). Angling for sharks, not pilot fish: Deep corruption, venal corruption, and public values failure. *Perspectives on Public Management and Governance*, 1(1), 5-27.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.

Burleigh, T., Kennedy, R., & Clifford, S. (2018, October 18). How to screen out vps and international respondents using Qualtrics: A protocol. <https://ssrn.com/abstract=3265459>

Cohn, A., & Maréchal, M. A. (2017). Laboratory measure of cheating predicts misbehavior at school. (November 2017). CESifo Working Paper Series No. 5613. Available at SSRN: <https://ssrn.com/abstract=2706473>

Coppock, A. (2019). Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. *Political Science Research and Methods*, 7(3), 613-628.

Fetzer, T., Witte, M., Hensel, L., Jachimowicz, J., Haushofer, J., Ivchenko, A., ... Yoeli, E. (2020, May 20). Perceptions of an insufficient government response at the onset of the COVID-19 pandemic are associated with lower mental well-being. PsyArXiv Preprint (Not Peer-Reviewed). <https://doi.org/10.31234/osf.io/3kfmh>

Gollwitzer, A., Martel, C., Marshall, J., Höhs, J. M., & Bargh, J. A. (2020, June 4). Connecting self-reported social distancing to real world behaviors at the individual and U.S. state level. PsyArXiv Preprint (Not Peer-Reviewed). <https://doi.org/10.31234/osf.io/kvnwp>

Hanna, R., & Wang, S. Y. (2017). Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy*, 9(3), 262-90.

Höglinger, M., & Jann, B. (2018). More is not always better: An experimental individual-level validation of the randomized response technique and the crosswise model. *PloS one*, 13(8).

Jann, B., Jerke, J., & Krumpal, I. (2012). Asking sensitive questions using the crosswise model: an experimental survey measuring plagiarism. *Public Opinion Quarterly*, 76(1), 32-49.

Korndörfer, M., Krumpal, I., & Schmukle, S. C. (2014). Measuring and explaining tax evasion: Improving self-reports using the crosswise model. *Journal of Economic Psychology*, 45, 18-32.

Larsen, M. V., Petersen, M., & Nyrup, J. (2020, April 9). Do survey estimates of the public's compliance with COVID-19 regulations suffer from social desirability bias? PsyArXiv Preprint (Not Peer-Reviewed). <https://doi.org/10.31234/osf.io/cy4hk>

Menzel, D. C. (2015). Research on ethics and integrity in public administration: Moving forward, looking back. *Public Integrity*, 17(4), 343-370.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, 2(2), 109-138.

Olsen, A. L., Hjorth, F., Harmon, N., & Barfort, S. (2019). Behavioral dishonesty in the public sector. *Journal of Public Administration Research and Theory*, 29(4), 572-590.

Stritch, J. M., Pedersen, M. J., & Taggart, G. (2017). The opportunities and limitations of using Mechanical Turk (Mturk) in public administration and management scholarship. *International Public Management Journal*, 20(3), 489-511.

Yu, J. W., Tian, G. L., & Tang, M. L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika*, 67(3), 251-263.

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... & Drury, J. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4, 460-471.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review*, 12(2), 250-264.

Author Version

Appendix A: Table A1. Summary Statistic and Test for Balance Across Sensitive Survey Technique Groups

Variable	All	(1)	(2)	(1-2)
		Direct Questioning (DQ)	Crosswise Model (CM)	F-Test ^a
<i>Sensitive Behaviors</i>				
DQ: Honesty in Dice Game (0 No, 1 Yes)		.87 (.34)		
CM: Honesty in Dice Game & Birthday of Parent (0 Different, 1 Same)			.38 (.49)	
DQ: Social Distancing Behavior (0 No, 1 Yes)		.30 (.46)		
CM: Social Distancing Behavior & Birthday of Parent (0 Different, 1 Same)			.59 (.49)	
Gender (0 Male, 1 Female)	.42 (.49)	.42 (.49)	.41 (.49)	F(1,1051)=.05, ns
Age (19-75)	38.29 (11.83)	38.58 (11.84)	37.99 (11.83)	F(1,1057)=.65, ns
Household Income (1-13)	6.20 (3.38)	6.09 (3.37)	6.32 (3.39)	F(12,1046)=.65, ns
White (0 Non-White, 1 White)	.70 (.46)	.72 (.45)	.67 (.47)	F(1,1057)=3.02, ns
<i>Education</i>				
Less than High School (0 No, 1 Yes)	.00 (.04)	.00 (.06)	.	
High School Graduate (0 No, 1 Yes)	.11 (.31)	.11 (.31)	.10 (.30)	
Some College (0 No, 1 Yes)	.20 (.40)	.19 (.39)	.20 (.40)	
2-Year Degree (0 No, 1 Yes)	.10 (.30)	.12 (.32)	.08 (.27)	
4-Year Degree (0 No, 1 Yes)	.44 (.50)	.46 (.50)	.41 (.49)	
Professional Degree (0 No, 1 Yes)	.14 (.35)	.11 (.31)	.18 (.38)	
Doctorate (0 No, 1 Yes)	.02 (.13)	.02 (.13)	.02 (.13)	
Political Ideological Orientation (0 Conservative, 1 Liberal)	.60 (.49)	.58 (.49)	.63 (.48)	F(1,878)=1.48, ns
Cognitive Ability (# Correct Reponses, 0-3)	1.59 (1.23)	1.56 (1.21)	1.63 (1.25)	F(1,1057)=.75, ns
Extraversion (34.38-100)	65.98 (11.39)	66.25 (11.42)	65.70 (11.36)	F(1,1035)=.58, ns

Agreeableness (33.33-100)	67.59 (9.80)	67.60 (9.67)	67.58 (9.94)	F(1,1044)=.00, ns
Conscientiousness (41.67-100)	77.68 (11.76)	77.45 (11.75)	77.92 (11.77)	F(1,1041)=.43, ns
Neuroticism (28.13-100)	58.53 (16.96)	58.88 (16.67)	58.18 (17.27)	F(1,1038)=.44, ns
Openness (20-87.50)	59.28 (10.86)	59.51 (10.69)	59.05 (11.05)	F(1,1039)=.47, ns
Observations	1,059 ^b	536	523	1,059 ^b

Notes: Columns "All", "1" and "2" report variable means with standard errors in parentheses ^a F-tests based on analysis of covariance. Test of group mean differences for individual covariates across treatment conditions. ^b Number of observations vary between 1,037 to 1,059 depending on missing values on items for individual Big 5 personality factors with exception for dichotomized political ideological orientation variable that contains 880 observations

Author Version