***Title (required)***

**Provide the working title of your study. It is helpful if this is the same title that you submit for publication of your final manuscript, but it is not a requirement.**

Survey Automation Detection Methods and its Implications on Psychological Research

***Authors (required)***

**The author who submits the preregistration is the recipient of the award money and must also be an author of the published manuscript. Additional authors may be added or removed at any time.**

John E. Scofield and Erin M. Buchanan

***Research Questions (required)***

**Please list each research question included in this study.**

Do automated survey responses differ in characteristics (i.e. skewness, kurtosis, response time, number of options chosen from a range of given scale options) compared to participants taking a survey without the use of an automated form filler, or when participants randomly provide answers to a survey?

What is the prevalence of participants utilizing survey bots or automated form fillers on Amazon Mechanical Turk?

***Hypotheses (required)***

**For each of the research questions listed in the previous section, provide one or multiple specific and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here.**

Automated survey responses will have shorter response times than participants taking the survey. (directional)

Automated survey responses will be less skewed than participants taking the survey. (non directional)

Automated survey responses will be more kurtotic than participants taking the survey. (non directional)

Automated survey responses will use more survey options in a given range of scale options than participants taking the survey (i.e. if a scale question has seven options, we predict that automated survey responses will use the majority of those options, whereas participants taking the same survey will use less). (non directional)

We predict that we will find a small subset of Amazon Mechanical Turk participants using survey bots or automated form fillers to take the survey. (non directional)

We predict that, from a sensitivity analysis, the inclusion of these detected automated survey responses will change the magnitude of effects found from the survey. (non directional)

A function will be created to screen and detect potential automated survey responses.

***In this section we’ll ask you to describe how you plan to collect samples, as well as the number of samples you plan to collect and your rationale for this decision. Please keep in mind that the data described in this section should be the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.***

***Existing Data (required)***

**Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).**

**Registration prior to creation of data**



**Registration prior to any human observation of the data**



**Registration prior to accessing the data**



**Registration prior to analysis of the data**



**Registration following analysis of the data**



***Explanation of existing data (required)***

**If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study. The purpose of this question is to assure that the line between confirmatory and exploratory analysis is clear.**

NA

***Data collection procedures (required)***

**Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that don’t include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.**

Data will be collected in two parts. In the first part, undergraduate students from a large Midwestern university will be recruited to participate in the study. Recruitment will be done via the online psychology participant pool or through manual classroom recruiting. These participants will not be compensated, but will either receive course or extra credit for their participation. Inclusion criteria include being an undergraduate student above the age of 18. Exclusion criteria include any student who is below the age of 18. There are no other exclusion criteria. The study timeline for this portion of the experiment will be one year or less, depending on the speed at which participants are recruited.

For the second part, participants will be recruited via Amazon Mechanical Turk. These participants will be compensated up to $0.50 for their participation. Inclusion criteria include being above the age of 18. Exclusion criteria include any participant who is below the age of 18. There are no other exclusion criteria. The study timeline for this portion of the experiment will be one year or less, depending on the speed at which participants are recruited.

***Sample size (required)***

**Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?**

For undergraduates recruited at a large Midwestern University, we aim to collect data from 100 participants. For participants recruited via Amazon Mechanical Turk, we aim to collect data from 1,000 participants.

***Sample size rationale (required)***

**This could include a power analysis or an arbitrary constraint such as time, money, or personnel.**

This sample size rationale is based on time and money. For undergraduate students, time to recruit and test participants is of concern if exceeding the sample size described. 100 participants also allows us to reach the central limit theorem to compare distributions.

For participants recruited via Amazon Mechanical Turk, budget restrictions allow us to afford to pay up to 1,000 participants.

***Stopping rule (required)***

**If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.**

The stopping rule for undergraduate participants will be 100 participants.

The stopping rule for Amazon Mechanical Turk participants will be 1,000 participants.

***In this section you can describe all variables (both manipulated and measured variables) that will later be used in your confirmatory analysis plan. In your analysis plan, you will have the opportunity to describe how each variable will be used. If you have variables which you are measuring for exploratory analyses, you are not required to list them, though you are permitted to do so.***

***Manipulated variables (required)***

**Describe all variables you plan to manipulate and the levels or treatment arms of each variable. For observational studies and meta-analyses, simply state that this is not applicable.**

We aim to manipulate how a survey is taken. The levels in this variable will a survey taken naturally (by reading each question and answering it based on that question), randomly taking a survey (not reading each question, and just plugging in random answers), and using an easily accessible Chrome extension automated form filler (which completes an entire survey in one mouse click) to take a survey.

For detecting the prevalence of automated survey responses via Mechanical Turk, basic manipulations from the survey will include age and sex (male, female). A sensitivity analysis will be performed by including and excluding any detected automated survey responses.

***Measured variables (required)***

**Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.**

The measured variables will be the scores from the survey taken, response time, skewness, kurtosis, and the number of scale options utilized.

A function will also be created, based on these measured variables, to help develop a screening method to identify potential automated survey responses for part two of the study to determine the rate of automated survey responses.

***Indices (required)***

**If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If your are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.**

Survey scores will be combined into one averaged score for sensitivity analysis, otherwise individual answers will be used to calculate skew/kurtosis and number of answers.

***In this section, you will be asked to describe the overall design of your study. Remember that this research plan is designed to register a single study, so if you have multiple experimental designs, please complete a separate preregistration.***

***Study type (required)***

**Please check one of the following statements**

**Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.**



**Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, “natural experiments,” and regression discontinuity designs.**



**Meta-Analysis - A systematic review of published studies.**



**Other**



***Blinding (required)***

**Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.**

**No blinding is involved in this study.**



**For studies that involve human subjects, they will not know the treatment group to which they have been assigned.**



**Research personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments.**



**Research personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.**



***Study design (required)***

**Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.**

For undergraduate participants, the study design will be a one-way repeated measures design, with the independent variable being the manner in which a survey was completed (naturally, randomly, using an automated form filler), and the dependent variables will be survey scores, response times, skewness, kurtosis, and number of scale options utilized.

For Mechanical Turk participants, all participants will take the same survey. Participants will then be screened to identify potential automated survey responses. Simple comparisons between age and sex will be placed in the context of a sensitivity analysis to note changes including or excluding those participants.

***Randomization (required)***

**If you are doing a randomized study, how will you randomize, and at what level?**

For undergraduate participants, surveys will be taken three times, in a randomized order for each participant. For Mechanical Turk participants, each participant will simply take the same survey once.

***You may describe one or more confirmatory analysis in this preregistration. Please remember that all analyses specified below must be reported in the final article, and any additional analyses must be noted as exploratory or hypothesis-generating.  
  
A confirmatory analysis plan must state up front which variables are predictors (independent) and which are the outcomes (dependent), otherwise it is an exploratory analysis. You are allowed to describe any exploratory work here, but a clear confirmatory analysis is required.***

***Statistical models (required)***

**What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions that will be tested and remember that any test not included here must be noted as an exploratory test in your final article.**

One-way repeated measures ANOVAs will be run between the different manners in which the survey was taken, with the dependent variables used being described above. An R function will then be created based on the survey characteristics of when surveys were taken via an automated form filler.

For Mechanical Turk participants, an ANOVA will be run to test differences in sex and a regression age on survey responses. Potential automated survey responses will be identified using the created R function. A sensitivity analysis will then be performed to note any changes in magnitude of effects, given if automated survey responses are included or excluded. To determine if effects are changed by inclusion/exclusion of potential automated responses, we will examine eta squared from the ANOVA with their non-central confidence intervals. Items that overlap indicate that automatic responses do not change results.

***Transformations (required)***

**If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.**

NA

***Follow-up analyses (required)***

**If not specified previously, will you be conducting any confirmatory analyses to follow up on effects in your statistical model, such as subgroup analyses, pairwise or complex contrasts, or follow-up tests from interactions? Remember that any analyses not specified in this research plan must be noted as exploratory.**

*t*-tests will be run on all significant effects from the ANOVA analyses, using a Bonferroni correction to control for multiple comparisons.

***Inference criteria (required)***

**What criteria will you use to make inferences? Please describe the information you’ll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?**

For inference criteria, both *p*-values and Bayes factors will be used. For *p*-values, .05 will be used as the significance criterion, using two tailed tests. Bonferroni corrections will be used to control for multiple comparisons. For Bayes factors, anything between 1-3 will be noted as ambiguous evidence. Anything from 4-20 will be considered positive evidence, and anything above 20 will be considered strong evidence. Any Bayes factor below one will be taken as evidence for the null.

***Data exclusion (required)***

**How will you determine which data points or samples (if any) to exclude from your analyses? How will outliers be handled?**

Any participant missing more than 5% of their data will be excluded. Any participant missing less than 5% of their data will be imputed through multiple imputation analyses (*mice* package). Outliers will be checked using z-scores. If any outliers are found, they will be excluded from further analyses.

***Missing data (required)***

**How will you deal with incomplete or missing data?**

Any participant missing more than 5% of their data will be excluded. Any participants missing less than 5% will be imputed through the *mice* package.

***Exploratory analysis (optional)***

**If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.**

Any exploratory analyses will be further labelled as exploratory.

***The purpose of a fully commented analysis script is to unambiguously provide the responses to all of the questions raised in the analysis section. This step is not common, but we encourage you to try creating an analysis script, refine it using a modeled dataset, and use it in place of your written analysis plan.***

**This optional step is helpful in order to create a process that is completely transparent and increase the likelihood that your analysis can be replicated. We recommend that you run the code on a simulated dataset in order to check that it will run without errors.**

All R scripts used for analyses will be made publicly available on OSF to help make the current project completely reproducible.

***Other (optional)***

**If there is any additional information that you feel needs to be included in your preregistration, please enter it here.**