Communications of the Association for Information Systems

Accepted Manuscript

Preregistration of Information Systems Research

Eric Bogert
Department of MIS
University of Georgia
etbogert@uga.edu

Aaron Schecter
Department of MIS
University of Georgia

Richard T. Watson
Department of MIS
University of Georgia

Please cite this article as: Bogert, Eric; Schecter, Aaron; Watson, Richard T.: Preregistration of Information Systems Research, *Communications of the Association for Information Systems* (forthcoming), In Press.

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

Preregistration of Information Systems Research

Eric Bogert
Department of MIS
University of Georgia
etbogert@uga.edu

Aaron Schecter
Department of MIS
University of Georgia

Richard T. Watson
Department of MIS
University of Georgia

Abstract:

We introduce the concept of preregistration for experiments in information systems. Preregistration is a way to commit to analytic steps before collecting or observing data, thus mitigating any biases authors may have (consciously or not) towards reporting significant findings. We explain why preregistration matters, how to preregister a study, the benefits of preregistration, address common arguments against preregistration, and offer a call to action for authors to conduct more preregistered work in IS.

Keywords: Preregistration, Philosophy of Science, Experiments.

[Department statements, if appropriate, will be added by the editors. Teaching cases and panel reports will have a statement, which is also added by the editors.]

[Note: this page has no footnotes.]

This manuscript underwent [editorial/peer] review. It was received xx/xx/20xx and was with the authors for XX months for XX revisions. [firstname lastname] served as Associate Editor.] **or** The Associate Editor chose to remain anonymous.]

1 Introduction

More than 60 years ago, it was posited that if enough people study a phenomenon for which there is no true relationship, eventually an author will find a statistically significant relationship and publish it (Sterling, 1959). This is the *error of the first kind*, or a *Type 1* error. It has been suggested that these errors, along with other false findings, constitute a majority of published research (Ioannidis, 2005). False positive findings are pernicious – it is difficult to invalidate a false-positive finding even in replications, because there are many reasons a replication might fail (Simmons, Nelson, & Simonsohn, 2011). In the rare event of a replication study, it is uncommon for top-tier journals to publish null results, making it even less likely for the field to learn that original study was a false positive (Simmons et al., 2011).

There are many reasons why false findings may constitute so much published research, and we do not cover every reason. Rather, we focus on the issues that are more likely to affect IS research. First, false positive findings are more likely to be published than true (null) findings because authors are more likely to submit statistically significant results (Franco, Malhotra, & Simonovits, 2014). Authors demonstrate this propensity to submit statistically significant results because statistically significant results are more likely to be published (Mahoney, 1977; Schwab, Abrahamson, Starbuck, & Fidler, 2011). Even after controlling for the quality of an experiment, authors typically choose to not submit papers with null results (Franco et al., 2014). Authors likely perceive that they are incentivized to submit research with significant findings because the top journals in a field tend to publish far more papers with significant results than non-significant. In psychology, for example, 96 percent of papers that use null hypothesis significance testing (NHST) report significant outcomes for their main hypotheses (Bakker, van Dijk, & Wicherts, 2012). If publishing papers is a game, an effective strategy is to find significant results by running many under-powered experiments with few participants (Bakker et al., 2012).

Second, false findings can constitute a majority of published research when multiple teams work on the same research stream (Sterling, 1959). Even if no team is using dubious research practices, when there are enough teams working on the same problem, there will be occasional Type 1 errors. If there is not a culture of publishing or otherwise disseminating null results, then if enough authors pursue a research topic they will invariably find and publish spurious correlations.

Third, false results are more likely to be submitted than true results because some authors might use dubious research practices. There are well-documented, problematic techniques used during NHST that invalidate the assumptions of NHST and make it more likely to publish null findings (Ioannidis, 2005). Hypothesizing after results are known (HARKing) is a process in which authors analyze data until they find relationships that they believe are publishable, then find theories related to those findings and hypothesize accordingly (Starbuck, 2016). Authors may be particularly excited when the theory would not predict the relationship, because journals are more likely to publish surprising research (Brembs, Button, & Munafo, 2013). HARKing is a serious problem – 92 percent of management professors claim to know someone who has HARKed (Bedeian, Taylor, & Miller, 2010).

Authors can also use p-Hacking to find statistically significant, but not necessarily replicable, results. P-Hacking “involves subjecting data to many calculations or manipulations in search of an equation or classification system that captures strong patterns” (Starbuck, 2016, p. 171). P-Hacking techniques include: selectively excluding participants from analyses, selectively including control variables and interactions, or choosing one of multiple moderately correlated dependent variables (Simmons et al., 2011). Running multiple calculations with subsets of variables or subsets of data can result in p-values that are too small (Mertens & Recker, 2020; Starbuck, 2016). Running multiple calculations usually can be fixed statistically with a correction for multiple tests (Dunnett, 1955; Tukey, 1949), but if researchers do not consistently record every test they run then correcting for multiple tests is impossible (Nosek, Ebersole, DeHaven, & Mellor, 2018; Starbuck, 2016). There is unassailable evidence of p-Hacking across disciplines. In political science, there are approximately twice as many p-values immediately below 0.05 as there are immediately above 0.05 (Gerber & Malhotra, 2008). In economics, studies of how the minimum wage affects unemployment tend to have effect sizes that are frequently twice the standard error, with many differing values of the standard error (Card & Krueger, 1995). Finally, it is exceedingly simple to ‘prove’ that either Republican or Democratic politicians are better for the economy, depending on how one chooses variables (“Hack Your Way To Scientific Glory,” 2020).

Worries about p-Hacking have led to high-profile, high-powered replication projects. In the Reproducibility Project: Psychology (RPP), 100 experiments from top journals were reproduced. One third to one-half found the original results, and effect sizes were generally half the size (Open Science Collaboration, 2015). In the Social Science Replication Project, (SSRP), 21 studies from Science and Nature were reproduced, with 13 finding significant effects in the same direction (Camerer et al., 2018). In the Experimental Economics Replication Project (EERP), 18 experiments from the *Quarterly Journal of Economics* and *The American Economic Review* were replicated – 11 had significant effects in the same direction and the average effect size was two thirds of the original (Camerer et al., 2016). The Information Systems Replication Project (ISRP) (Dennis, Brown, Wells, & Rai, 2018) is the first large-scale replication project in information systems. The ISRP found that many of the replications found similar results to their predecessors (Dennis et al., 2018; Dennis, Brown, Wells, & Rai, 2020). IS may have had more success in replication because IS researchers often replicate another field's theories in an IS context.

One aspect of these replication projects is that authors replicate the work, oftentimes following a preregistered analysis plan. In preregistered plans, authors are expected to list their exact model, their proposed control variables, how they will measure each variable, their hypotheses, and how they will exclude and include any data. For many replication projects, papers are included based on how well they adhere to the original study's methodology and based on how frequently the hypothesis has already been replicated.

However, not all preregistrations are for replication projects. An increasing number of journals are requiring or encouraging preregistration of new experiments. This has been the standard in medicine for decades, and preregistration is required by United States' law in clinical trials since the passage of the Food and Drug Administration Modernization Act of 1997 (FDAMA). Premier journals such as *Science, Nature, Proceedings of the National Academy of Science,* and *Management Science* now publish many preregistered studies.[1] Furthermore, the number of preregistrations is roughly doubling every year on the most common preregistration repository (Kupferschmidt, 2018).

The purpose of this article is to introduce the IS community to preregistration of experiments, particularly in non-replication contexts. We focus on experimental research, although much of the process can be used in non-experimental research as well. We discuss the advantages and disadvantages of preregistration, the process of preregistration, and provide a sample preregistration for a published experiment in Management Science in the appendix.

2 What is Preregistration?

Preregistration is defined as "committing to analytic steps without advance knowledge of the research outcomes" (Nosek et al., 2018, p. 2601). Preregistration occurs by posting an analysis plan on a scientific repository, such as the Open Science Foundation (OSF). It is commonly associated with an experiment, although it is possible to preregister analysis plans for datasets not gathered through an experiment. For example, if a researcher were interested in government data that was about to be released, they could preregister their analysis plan before the data were released (Nosek et al., 2018). Similarly, if a researcher were interested in a phenomenon that was about to unfold in social media (e.g., the months preceding an election, or social media around a predictable event such as the Olympics), then they could preregister their analysis plan of the upcoming event.

Researchers should not claim to have preregistered their study if they know too much about the data in advance of preregistration. There is no perfect rule to determine whether a researcher knows too much about the dataset to honestly claim that their analysis plan came in advance of the research outcomes. For example, if someone gives a dataset to their colleague and tells them of a few interesting correlations they have already found, that would usually be too much information to honestly preregister future analysis plans (Nosek et al., 2018). Similarly, if a researcher is already familiar with a relationship between a DV and IV in their data, they should not preregister a study about two similar, likely highly correlated DVs and IVs in the same dataset (Nosek et al., 2018).

There are generally two types of preregistration. One type, registered reports, require authors to submit a preregistration for review at a journal prior to gathering data. The journal can agree to publish the findings regardless of the results – thereby reducing publication bias by not providing a strong incentive for authors to find significant results. In the second type of preregistration, and the focus of this paper, authors submit

1We perused the most recent 21 papers from the Psychology and Cognitive Science section of the Proceedings of the National Academy of Science and found that four of eleven experiments in the sample were preregistered. One of ten non-experiments was preregistered.

an analysis plan to a repository. Authors usually allow reviewers to review the preregistration during the review process, and subsequently release it fully to the public upon publication of their paper. In both forms of preregistration, the preregistration occurs prior to data collection. Authors use time-stamped repositories as proof that they conducted the experiment after theorizing, with the (often implicit) argument that questionable research practices were thus less likely to accompany the focal paper due to the transparency afforded by preregistration.

3 Why Preregister?

Preregistration is important for many reasons. We will explore four of the rationales introduced in prior scholarship(Nosek et al., 2018; Roloff & Zyphur, 2019).

3.1 Reducing Publication Bias

Preregistration is the strongest predictor of whether a study has true positive findings instead of false positive findings (Swaen, Teggeler, & van Amelsvoort, 2001). In a meta-analysis of medical trials, preregistration was strongly correlated with null findings (Kaplan & Irvin, 2015). Preregistration is now mandated for experiments that adhere to the policies of the International Committee of Medical Journal Editors. In short, preregistration is the best way to reduce Type 1 errors in a field.

3.2 Clarifying Prediction and Postdiction

Good science begins with an ethical, open explanation of the scientific process conducted and the type of theorizing occurring. Preregistration makes it clear to readers when a finding was informed *a priori* by theory and when a finding was determined through post-hoc analysis (Nosek et al., 2015, 2018). Preregistration does not preclude the use of inductive theorizing, but it necessitates that the authors are clear about using post-hoc analyses.

This additional clarity of preregistration makes it easier to assess what is theory generating versus theory testing (Nosek et al., 2018). While analyzing data, it might be easy for authors to convince themselves that the model specification that gives significant findings matches their original intentions. Two psychological biases work in tandem to make scientists believe that their final model with significant results was the model envisioned originally (Nosek et al., 2018). First, humans are more likely to conclude what they want to see because of motivated reasoning (Kunda, 1990). Motivated reasoning is constrained by one's ability to justify one's conclusions. In ambiguous contexts such as scientific research, there can be multiple ways to justify actions such that almost any decision is justifiable (Nosek et al., 2018). Secondly, humans are prone to hindsight bias; when reflecting on past events to which they assigned probabilities, humans tend to inflate those they assigned to events that actually happened, and deflate those of events that did not happen (Fischhoff & Beyth, 1975). Preregistration makes it easier to combat these biases and ensures that authors do not accidentally conflate postdiction with prediction.

3.3 P-Hacking

Preregistration is one partial antidote to p-Hacking. An all-too-common syllogism in social science is to observe a phenomenon in one's data, write a hypothesis about that data, and find the hypothesis to be true. This is circular logic concealed as theorizing (Nosek et al., 2018).

Of course, preregistered experiments can also be prone to p-Hacking if the authors leave themselves too much leeway in the analysis plan. A properly preregistered study should explicitly state the data exclusion criteria (e.g., acceptable answers for manipulation checks and attention checks), controls, and ways of measuring each variable. We openly admit that preregistration does not stop all forms of dubious research practices – it cannot, for example, prevent authors from making up data whole cloth.

3.4 Multiple Testing

There are many methodologies to adjust for multiple hypotheses testing on the same dataset. For example, a Bonferroni correction requires the author to divide the p-value required (usually 0.05) by the number of tests conducted (Bonferroni, 1936). Without preregistration, it is exceedingly difficult to identify how many tests were conducted. This makes any adjustment to required p-values meaningless (Nosek et al., 2018). With preregistration this becomes simple – authors can outline in advance how many models they intend to run, and proactively state what, if any, adjustments for multiple tests they intend to make.

4 How Preregistration Helps Authors

Some authors may consider preregistration an onerous burden without significant benefits. We argue that is a misguided view. We review four ways in which preregistration directly helps authors (Nosek et al., 2018):

4.1 Reducing Doubt

When authors preregister a study, they make it clear to reviewers that their hypotheses were informed by theory. Any post-hoc findings can be labeled as such, and if possible, can be investigated in future research. If authors want to theorize based on post-hoc findings, they can do so – editors and readers can decide whether they find the post-hoc analysis appropriate and believable. For relatively low-cost research, such as experiments using subjects recruited from Amazon Mechanical Turk, follow-up experiments can be conducted to investigate these post-hoc findings. Due to the additional constraints on authors, some researchers have advocated that reviewers be less exacting on findings that adhere to these more transparent practices (Nosek et al., 2018; Simmons et al., 2011).

4.2 Reminding Authors What Was Prediction and Postdiction

Authors will be more likely to have replicable, believable work when they preregister, because they have a concrete guide to the process they were envisioning at the onset of the paper. Due to hindsight bias and motivated reasoning, authors are likely to convince themselves that their final model was the model they envisioned (Nosek et al., 2018). Preregistration makes it simple for authors to check that and ensures that they distinguish between prediction and postdiction.

4.3 Clarifying Experimental Designs

Preregistration can help authors identify and correct mistakes prior to running their experiments or data analysis. Putting the design on paper with the expected analysis plan can help researchers see faults in their design that they may have otherwise not seen. A founder of AsPredicted.org reported that he has twice realized his study did not make sense as he wrote the preregistration (Kupferschmidt, 2018). Pre-registration can improve research design as many problems cannot be 'fixed' later. This would both strengthen the validity of studies and reduce the likelihood that authors need to rerun a faulty experiment.

4.4 Enabling Streams of Research

Preregistration also helps enable streams of research, which has recently been advocated as critical to the success of the IS field (Gable, 2020). When a study has surprising findings – whether they are significant and theoretically interesting but un-hypothesized interactions, or the lack of an expected effect – future research can investigate these effects. In non-preregistered studies, it is tempting to simply add a new hypothesis that posits the effect observed post-hoc. In preregistered studies, interesting, but non-hypothesized effects, practically necessitate follow-up studies and additional research.

5 The Preregistration Process

The guiding principle behind what to preregister and what not to preregister should be "Does preregistering reduce my options during data analysis?" If the answer is yes, then preregister it (Simmons et al., 2011). Thus, there is no need to preregister the introduction, literature review, discussion, or conclusion.

5.1 Theorizing

The preregistration process begins with theorizing. After reviewing the literature and identifying the causal models and relevant theories, authors theorize about the effects they expect to see. While this step is important to the scientific process, many preregistration templates do not require an explication of the theory behind the propositions and their operationalization as hypotheses. This is because it is expected that the theorizing will take place in the published paper to which the preregistration is tied.

5.2 Proposition and Hypothesizing

We advocate a two-step procedure. First, outline the propositions for investigation. Second, specify how each proposition is operationalized, as there are often multiple ways of measuring a concept and post data collection adjustments of concept measurement should be discouraged. Some preregistration templates allow authors to differentiate between exploratory hypotheses and confirmatory hypotheses.

5.3 Committing to Analytic Steps

Choosing an analysis plan is generally considered the most important part of preregistration. The more specific the analytic steps, the fewer “researcher degrees of freedom” (Simmons et al., 2011). Four of the most important factors to clarify are: which control variables to include, how to measure the dependent variable, the proposed sample size, and choosing subsets of experiment conditions (Simmons et al., 2011).

Imagine an IS study with two potential dependent variables, behavioral intention to use a system and affect towards a system. Assuming a correlation at r = 0.50 between the DVs, flexibility in choosing a DV almost doubles the likelihood of a false positive finding (Simmons et al., 2011). Similarly, if a researcher is willing to stop a study during data collection after finding a significant finding, the researcher increases the chance of a false positive finding by about 50 percent (Simmons et al., 2011). Combining common researcher degrees of freedom makes the total probability of a false positive 61 percent; researchers with high flexibility in analysis are thus more likely to find a false positive (Simmons et al., 2011).

These researcher degrees of freedom are commonly used to falsely demonstrate significant results. In a survey of academics, 70 percent reported stopping data collection early because they had found statistical significance (John, Loewenstein, & Prelec, 2012). This is likely an under-assessment of the extent of the problem given that the study relied on self-reported behavior.

It is strong science to choose models that are theoretically justifiable. If that model is not the best-fitting, so acknowledge and explain post hoc why that was observed. Researchers have the option to preregister that they will choose the model of best fit (by ROC curve, AIC, or some other criteria). Authors should not preregister that they will determine fit by choosing the model in which the key variable or interaction has a significant p-value (Nosek et al., 2018).

5.3.1 Power Analyses

Many preregistration templates ask for a power analysis explanation. The author can explain how they computed the number of subjects for an experiment, whether that is due to cost, time, or statistical power. Papers that use historical, non-experimental data can preregister an alternative plan, such as specifying the number of observations to evaluate or the use of cross-validation of the data. Cross-validation of the data is a common technique used in IS papers that study large datasets, and can be used to test how well a model predicts a dependent variable on a held-out dataset (Lin, Lucas, & Shmueli, 2013).

5.3.2 Data Exclusions

It is important to list any process for excluding observations prior to gathering data. In the context of online experiments, this can include whether to include subjects who fail attention checks, manipulation checks, or subjects with duplicative IP addresses. Data exclusions in advance are particularly important. For example, researchers are usually not consistent in how they treat outliers (Simmons et al., 2011). Consider time as a basis to exclude subjects; authors can exclude the fastest or slowest 1 percent, or 2.5 percent of subjects, or subjects who were one or two standard deviations above the mean, or some other cutoff (Simmons et al., 2011). Authors should absolutely retain the ability to choose the cutoffs for excluding outliers, but in advance of their application.

Although not specifically asked on most preregistration templates, we advocate identifying possible robustness checks related to subsets of the data in advance.

5.4 Preregister on Repository

Authors need to choose what preregistration template and repository works best for them. Perhaps the simplest and least burdensome template is from AsPredicted.org, which asks the following questions:

1) Have any data been collected for this study already?

2) What's the main question being asked or hypothesis being tested in this study?
3) Describe the key dependent variable(s) specifying how they will be measured
4) How many and which conditions will participants be assigned to?
5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations
7) How many observations will be collected or what will determine sample size?
8) Anything else you would like to preregister?

Other templates may ask slightly different questions. The standard OSF template asks whether a study includes blinding, how authors will use randomization in selecting subjects, and how variables will be measured, manipulated, defined, and (if applicable) combined into an index. When submitting a preregistration, some repositories require authors to indicate a date of public release, after which the author loses the ability to prevent the public release of their research plan. This ensures that preregistrations cannot be buried if inconvenient to authors.

We summarize important differences between three prominent repositories in the social sciences. One of these repositories, ResearchBox, was launched in late 2020 and builds on AsPredicted.org, allowing AsPredicted new functionality such as uploading data and materials. We consider AsPredicted and ResearchBox as a single consolidated platform: they share a server, have APIs that connect AsPredicted's preregistration to ResearchBox's platform, and are both housed in the Wharton Credibility Lab. We also review the OSF and the American Economic Association (AEA) Randomized Control Trial (RCT) Registry in Table 1. All three platforms allow for uploading preregistrations, data, and materials such as a survey from Qualtrics. One important distinction between the platforms is that public preregistrations on ResearchBox are currently not searchable, although this functionality is intended to be released in late 2021. Many authors, including the authors of this manuscript, use the AsPredicted template but post their responses to that template, along with their code, data, and other materials to OSF or ResearchBox.

Perhaps the most important distinction between repositories is whether they eventually make all preregistrations public. AsPredicted and ResearchBox do not make preregistrations public unless the authors make the explicit decision to do so. The argument for making all preregistrations public after a certain duration is that it allows scholars to investigate what has already been done, providing an antidote to the "file drawer problem" in which authors do not publish null results (Franco et al., 2014). AsPredicted's stance is that forcing publication of all preregistrations may deter some preregistrations, which is a significant cost (AsPredicted, n.d.-a). There are two important benefits to making all preregistrations public (AsPredicted, n.d.-a). First, authors cannot simply upload multiple preregistrations, each predicting opposite effects, for the same study. This practice would be unethical and against the principles of preregistration. AsPredicted's algorithmic sweeping has detected no instances of this behavior (AsPredicted, n.d.-b). Second, if preregistrations become public, then authors could theoretically comb them to see what results were not published and adjust their research accordingly.

Table 1. Differences Between Preregistration Repositories

	AsPredicted.org / ResearchBox	Open Science Foundation	American Economic Association RCT Registry
Can store data, materials, and code	Yes, through ResearchBox	Yes	Yes
All preregistrations eventually become public	No	Yes, the maximum embargo period is four years	Some data is made public initially, other data is kept private unless released publicly.
Public content can be searched	Searching materials on ResearchBox will be possible eventually.	Yes	Yes

5.5 Gather data (if necessary)

Preregistration is most common and appropriate prior to data gathering or analysis, such as in an experiments or meta-analysis. After preregistering,

authors collect data or start analysis, as appropriate.

5.6 Analyze Data

After gathering data, authors test their preregistered models. Authors can choose to run additional models that interest them and label any results from those as post-hoc.

5.7 Report Results

All preregistered hypotheses should be reported. If there were multiple hypotheses in a study, focusing on supported hypotheses is discouraged, in order to avoid accidentally highlighting Type 1 error results. If a preregistration has twenty hypotheses and only one is supported, the likelihood that the one supported hypothesis is a Type 1 error is relatively high. Similarly, completely dismissing hypotheses that are not supported is biased reporting. If a researcher conducts twenty highly similar experiments per year and finds one supported result per twenty experiments, the researcher should report the results from the other nineteen to provide the context necessary to evaluate whether the finding is likely due to a false positive (Nosek et al., 2018).

5.8 Follow-Up Studies on Postdictions

Preregistration lends itself to creating programs of research. It is a good practice to conduct follow-up studies on results that were found post-hoc. This helps to build an edifice of IS research (Tiwana & Kim, 2019) and ensures that Type 1 errors are quickly found. We outline these processes below in Figure 1.

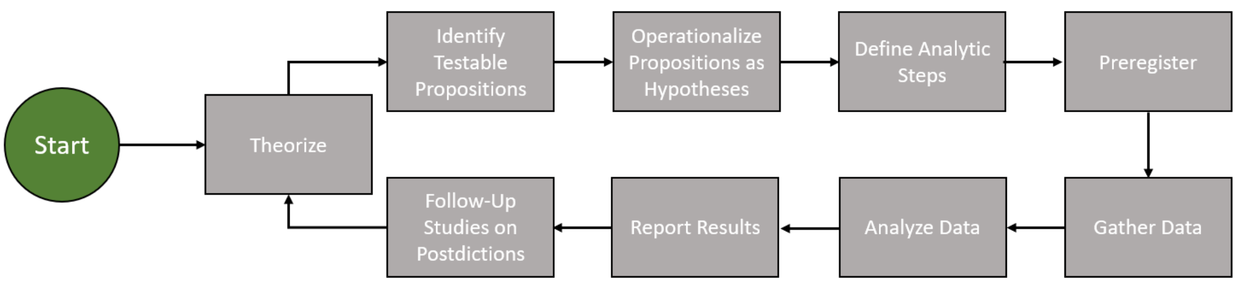

Figure 1. Process of Preregistration

6 Arguments against Preregistration

This is not a comprehensive list of the arguments against preregistration. For additional analyses of the arguments against preregistration and their rebuttals, see Nosek et al., (2018).

6.1 Preregistration Does Not Completely Solve Publication Bias

Authors who want to engage in research practices, such as p-Hacking and HARKing, could manage to do so even if they preregister a study. Even if authors specify every control variable, interaction, sample size, model specification, etc., authors might still be able to find dubious ways to produce significant results that support their hypotheses. However, pre-registration should make it easier for reviewers to identify such behavior, and should provide an incentive for authors to not engage in dubious research practices.

6.2 Preregistration Is Too Burdensome

Although it might seem more work to write a preregistration, sections of it can be leveraged in the associated paper. For example, most papers include an explanation of the variables, data collection procedures, and sample – these sections will be mostly written in advance if the authors preregister. Furthermore, the process of preregistration can clarify thinking that saves valuable time throughout an investigation. Good planning, such as preregistration, is usually time-saving over the span of a project. For instance, the experiments we have preregistered usually take around ten minutes.

6.3 Preregistration Uses a Flawed Epistemology

Some IS scholars might be reluctant to preregister because they believe it requires accepting an epistemology that is not in accordance with their research philosophy. We are not implying that they should change their beliefs, but if preregistration becomes more common, or even required, authors who do not preregister may find themselves at a disadvantage.

If scholars are interested in more exploratory research where there are few guiding theories to frame propositions, preregistration can still be used. If multiple studies are planned, preregister an initial study and investigate interesting post-hoc findings in future preregistered, theory-testing studies. If only one large study is planned, we recommend preregistering cross-validation of the data (Fafchamps & Labonne, 2017; Nosek et al., 2018). After analyzing the first set of data and drawing initial conclusions, test whether those conclusions are supported in the held-out data.

6.4 Replication Makes Preregistration Unnecessary

Although the IS discipline is not in the midst of a replication crisis (Dennis et al., 2020), preregistration aligns with many of the recommendations offered to advance the field. To maximize replicability in IS, researchers should provide details around context (i.e., temporal, cultural, and task-specific factors), methodology (i.e. research, design, analysis, and interpretation of results), and techniques (e.g. code) (Dennis et al., 2020). Much of these materials should be present in a preregistration, making it almost no additional work for scholars who intend to make their work replicable, while forcing careful documentation of these matters in advance rather than delaying until research reporting. The Open Science Foundation, which enables authors to upload materials in addition to preregistration, is an example of how preregistration goes hand-in-hand with replication. Authors can upload not only their preregistration, but also their code, materials, and data – all of which are beneficial to replication endeavors. The authors of the example preregistration in Appendix 1 have included code, materials, data, and the preregistration in their submission to the OSF, later published in *Management Science* (https://osf.io/b5m3n/). We assert that preregistration and replication are part of a three-phase approach to raising research quality: (1) preregister, (2) research, and (3) replicate. They go hand-in-hand-in-hand and have been doing so for decades in rigorous disciplines such as medicine.

There are two other reasons preregistration adds value above and beyond replication. First, preregistration makes it clear what hypotheses were *not* supported in a paper, thereby helping the field focus on more fruitful areas of inquiry. Second, preregistration helps to prevent Type 1 errors from being disseminated *prior* to publication (Nosek et al., 2018). This aspect is important, because once papers are published it becomes exceedingly difficult to correct the scientific record. Thinking about replication in juxtaposition with a formal retraction is a useful exercise to understand why a process that prevents erroneous research from being published is far more effective than attempting to correct the record after publication. With retraction, which is the most extreme, documentable, and public case of correcting the scientific record, many papers are still cited positively for their original findings, even if the paper was retracted for data fabrication, ethical misconduct, or false reports (Bar-Ilan & Halevi, 2017; Borenemann-Cimenti, Szilagyi, & Sandner-Kiesling, 2016). We expect that replications which fail to replicate an original experiment will have substantially less impact than a retraction on correcting the scientific record. Unlike retractions, which usually result in databases making a note that a paper has been retracted, there is no such indication next to papers that failed to be replicated. Thus, our field should be incentivized to stop errors *prior* to publishing, through preregistration.

7 Extending to Non-Experiments

Preregistration is most commonly used in experiments and clinical trials, but it has broader applicability to empirical research of all sorts. For example, qualitative researchers can preregister who they intend to interview, their interview questions, locations of interviews, the number of interviews they expect to need to reach data saturation, any variables they intend to capture, and their methods (e.g. Glaser). Researchers outside of IS are already using preregistration in studies using archival data such as natural experiments. Preregistration is possible even for literature reviews or meta-analyses. The first preregistered paper published in one of the senior scholar basket of eight journals was a meta-analysis of null-hypothesis significance testing in IS (Mertens & Recker, 2020). The authors preregistered how they sampled papers from top journals, their hypotheses about the use of NHST in those papers, and their analysis plan. IS also has established frameworks for literature reviews (Templier & Paré, 2015; Webster & Watson, 2002). Preregistration complements structured literature reviews because authors can preregister how they will follow the guidelines offered by the established frameworks. This process will also help authors respond to recent calls for increased transparency in literature reviews (Templier & Paré, 2018). In an analysis of the 20 most recent public preregistrations on the Open Science Foundation registry, we found six experiments, eight meta-analyses, five observational studies (e.g. RDD, natural experiments), and one qualitative study.[2] Observational studies in this sample frequently included any statistical transformations, proposed models, hypotheses, data exclusion plans, and measurements of variables. Meta-analyses frequently list the databases from which articles will be taken, inclusion criteria, exclusion criteria, and analysis plans.

8 Call to Action

Authors, reviewers, and journals are responsible for the future of preregistration of IS studies. They have the power to make preregistration an IS research norm, and without their support preregistered studies will remain voluntary and on the periphery of IS research.

8.1 Authors

Preregistration is a low-cost, impactful commitment to academic practice and quality that is already commonplace for many journals and disciplines. Authors can choose to be at the forefront of best practices in research while increasing the replicability of their work. Authors who embrace preregistration may also find that it helps reduce errors before they happen by forcing authors to fully think through experimental manipulations, variables, and data exclusion practices.

8.2 Reviewers

A similar call for increased transparency in experimental research argues that reviewers should appreciate the transparency of studies in which authors report rules for data collection, all experimental conditions (including failed), and the effect of removing any eliminated observations (Simmons et al., 2011). Reviewers should remember that low power, non-preregistered studies with opaque methods and perfect results are the studies that deserve the most scrutiny (Simmons et al., 2011). If recommending secondary or tertiary studies to enhance a focal study, reviewers can recommend that authors preregister any additional analyses. If authors report interesting findings that are the result of post-hoc analysis of a preregistered experiment, then reviewers, readers, and editors should be willing to consider those, if (and only if) the authors report that the models were not preregistered.

8.3 Journals

The Transparency and Openness Promotion (TOP) Guidelines give journals a score between zero and three based on their acceptance of preregistration (Mellor, 2020). Their definitions for each score are:

0. Journal says nothing about preregistration
1. Articles will state if work is preregistered
2. Articles state if work was preregistered. Journal verifies adherence to preregistration plan
3. Journals require that confirmatory or inferential research must be preregistered

As of this writing, *Nature*, *Science*, and *The Proceedings of the National Academy of Sciences* are all at level one of the TOP Guidelines, and *Nature Human Behavior* is at level two. Relatively few journals have policies that reach level three, although some have intermediate policies. *Psychological Science*, for example, states that “Manuscripts reporting preregistered research will have an advantage over otherwise comparable manuscripts reporting studies that were not pre-registered” (Preregistration of Research

[2] We conducted this analysis on August 3rd, 2020. We excluded one non-English preregistration, and sampled only papers using the OSF preregistration template.

Plans, n.d.). Lastly, the Editor in Chief of MISQ has stated that preregistration is an example of an editorial innovation he is considering exploring at MISQ.[3]

Advocates of preregistration offer a fourth option that journals can use to embrace preregistration – by offering registered reports (Graf, 2017). Registered reports allow journals to commit to publishing work prior to data collection, without knowledge of the outcomes, thereby reducing the incentives to engage in dubious research practices.

The data indicate that preregistration is gaining acceptance as a scientific convention, and it is probably a practice IS journals will adopt at some stage. AIS is the appropriate body, through its journals, to lead the IS field in adopting a norm that promotes research excellence, as AIS "serves society through the advancement of knowledge and the promotion of excellence in the practice and study of information systems."

9 Conclusion

In the same vein as prior publications designed to further IS research (Jarvenpaa, Dickson, & DeSanctis, 1985; Lin et al., 2013; Mertens & Recker, 2020; Straub, Boudreau, & Gefen, 2004), we provide an overview of preregistration, which while not a cure-all for biased acceptance of significant findings, is a step in the right direction. Preregistration aligns with recent calls to action for more programmatic research, in which studies cumulatively build upon each other using a thoughtful, long-term, strategic plan to thoroughly investigate a phenomena (Gable, 2020).

We should note that it is not always dubious or unethical to allow data to develop new theories. Careful analysis of data can result in robust, important findings, and has for centuries. For example, the electron was discovered in an experiment without postulating in advance that it existed (Achinstein, 2001). Rather, we echo previous calls for researchers to openly present such unanticipated findings as part of the theory generating rather than the theory-testing (Nosek et al., 2018). This call to transparency and iteratively approaching data echoes previous IS papers that approach theory generation with quantitative data (Evermann & Tate, 2011). We believe that preregistration is the best step to ensure that this distinction remains salient throughout the dissemination of academic research.

[3] Presentation to Department of MIS, October 16, 2020

References

Achinstein, P. (2001). *The Book of Evidence*. Oxford: Oxford University Press.

AsPredicted. (n.d.-a). AsPredicted. Retrieved July 10, 2020, from https://aspredicted.org/messages/private_forever.php

AsPredicted. (n.d.-b). Terms of Use. Retrieved October 11, 2020, from https://aspredicted.org/messages/terms.php

Bakker, M., van Dijk, A., & Wicherts, J. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, *7*(6), 543–554.

Bar-Ilan, J., & Halevi, G. (2017). Post retraction citations in context: a case study. *Scientometrics*, *113*, 547–565.

Bedeian, A. G., Taylor, S. G., & Miller, A. N. (2010). Management Science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning and Education*, *9*, 715–725.

Bonferroni, E. (1936). Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni Del R Istituto Superiore Di Scienze Economiche e Commerciali Di Firenze*.

Borenemann-Cimenti, H., Szilagyi, I. S., & Sandner-Kiesling, A. (2016). Perpetuation of Retracted Publications Using the Example of the Scott S. Reuben Case: Incidences, Reasons, and Possible Improvements. *Science and Engineering Ethics*, *22*, 1063–1072.

Brembs, B., Button, K., & Munafo, M. (2013). Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, *7*(291).

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644.

Card, D., & Krueger, A. (1995). Time-series minimum wage studies: A meta-analysis. *American Economic Review*, *85*, 238–243.

Dennis, A., Brown, S., Wells, T., & Rai, A. (2018). Information Systems Replication Project. Retrieved from https://aisel.aisnet.org/trr/aimsandscope.html

Dennis, A., Brown, S., Wells, T., & Rai, A. (2020). Replication Crisis or Replication Reassurance: Results of the IS Replication Project. *Management Information Systems Quarterly*, *44*(3), 3–10.

Dunnett. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Society for Information Science and Technology*, *50*(272), 1096–1121.

Evermann, J., & Tate, M. (2011). Fitting Covariance Models for Theory Generation. *Journal of the Association for Information Systems.*, *12*(9), 631–661.

Fafchamps, M., & Labonne, J. (2017). Using split samples to improve inference on causal effects. *Political Analysis*, *25*(4), 465–482.

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, *13*(1), 1=16.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505.

Gable, G. (2020). Viewpoint: Information Systems Research Strategy. *Journal of Strategic Information Systems*, *29*(2).

Gerber, A., & Malhotra, N. (2008). Do Statistical Reporting Standards Affect What is Published? Publication Bias in Two Leading Political Science Journals. *Quarterly Journal of Political Science*, *3*(3), 313–326.

Graf, C. (2017). 8 Answers About Registered Reports and Research Preregistration. Retrieved from https://www.wiley.com/network/researchers/being-a-peer-reviewer/8-answers-about-registered-reports-research-preregistration-and-why-both-are-important

Hack Your Way To Scientific Glory. (2020). Retrieved from https://projects.fivethirtyeight.com/p-hacking/

Ioannidis, J. (2005). Why Most Published Research Findings Are False. *PLoS One Medicine*, *2*(8).

Jarvenpaa, S., Dickson, G., & DeSanctis, G. (1985). Methodological Issues in Experimental IS Research: Experiences and Recommendations. *Management Information Systems Quarterly*, *9*(2), 141–156.

John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532.

Kaplan, R., & Irvin, V. (2015). Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased Over Time. *PLoS One*, *5*(10).

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480--498.

Kupferschmidt, K. (2018). More and more scientists are preregistering their studies. Should you? *Science*. Retrieved from https://www.sciencemag.org/news/2018/09/more-and-more-scientists-are-preregistering-their-studies-should-you

Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research Commentary - too big to fail: Large samples and the p-value problem. *Information Systems Research*, *34*(4), 906–917.

Mahoney, M. (1977). Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System. *Cognitive Therapy and Research*, *1*, 161–175.

Mellor, D. (2020). TOP Factor Rubric. Retrieved from https://osf.io/t2yu5/

Mertens, W., & Recker, J. (2020). New Guidelines for Null Hypothesis Significance Testing in Hypothetico-Deductive IS Research. *Journal of the Association for Information Systems*, *21*(4).

Nosek, B., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Nosek, B., Ebersole, C., DeHaven, A., & Mellor, D. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600–2606.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

Pennycook, G., Bear, A., Collins, E., & Rand, D. (2020). The Implied Truth Effect: Attached Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. *Management Science*.

Preregistration of Research Plans. (n.d.). Retrieved from https://www.psychologicalscience.org/publications/psychological_science/preregistration

Roloff, J., & Zyphur, M. (2019). Null Findings, Replications and Preregistered Studies in Business Ethics Research. *Journal of Business Ethics*, *160*, 609–619.

Schwab, A., Abrahamson, E., Starbuck, W., & Fidler, F. (2011). Researchers Should Make Thoughtful Assessments Instead of Null-Hypothesis Significance Tests. *Organization Science*, *22*(4), 1105–1120.

Simmons, J., Nelson, L., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366.

Starbuck, W. (2016). 60th Anniversary Essay: How Journals Could Improve Research Practices in Social Science. *Administrative Science Quarterly*, *61*(2), 165–183.

Sterling, T. D. (1959). Publications decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34.

Straub, D., Boudreau, M.-C., & Gefen, D. (2004). Validation Guidelines for IS Positivist Research. *Communications of the Association for Information Systems*, *13*, 379–427.

Swaen, G., Teggeler, O., & van Amelsvoort, L. G. (2001). False positive outcomes and design characteristics in occupational cancer epidemiology studies. *International Journal of Epidemiology*, *30*(5), 948–954.

Templier, M., & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, *37*, 112–137.

Templier, M., & Paré, G. (2018). Transparency in literature reviews: an assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems*, *27*(5), 503–550.

Tiwana, A., & Kim, S. (2019). From Bricks to an Edifice: Cultivating Strong Inference in Information Systems Research. *Information Systems Research*, *30*(3), 1029–1036.

Tukey. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, *5*(2), 99–114.

Webster, J., & Watson, R. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *Management Information Systems Quarterly*, *26*(2), 13–23.

Appendix A: Sample Preregistration

The following preregistration can be found at https://osf.io/tvbj5/. This preregistration is tied to a recent Management Science paper (Pennycook, Bear, Collins, & Rand, 2020) that found that the presence of fake news flags makes people believe more in unflagged articles than they would if no articles had fake news flags. We received explicit written permission to reproduce the preregistration in our paper.

1. **Have any data been collected for this study already?**
   No, no data have been collected for this study yet.
2. **What's the main question being asked or hypothesis being tested in this study?**
   A previous study showed that tagging a subset of false news articles with a “disputed by third-party fact-checkers” warning increases perceived accuracy of untagged news stories. Here we ask whether this effect extends to intention to share stories, and whether including “verified” tags on a subset of true news articles lower perceived accuracy of untagged stories.
3. **Describe the key dependent variable(s) specifying how they will be measured.**
   Participants will be presented with a series of false and true news headlines and asked for each: "If you were to see the above article on social media, would you consider sharing it?". They will respond “no” or “yes”.
4. **How many and which conditions will participants be assigned to?**
   Participants will be randomly assigned to one of three conditions: 1) Control condition – headlines are not flagged with any labels, 2) False Flag condition – 75% of the false headlines will be flagged with a “false” stamp, 3) False+True Flags condition – 75% of the false headlines will be flagged with a “false” stamp, and 75% of the true headlines will be flagged with a “true” stamp.
5. **Specify exactly which analyses you will conduct to examine the main question/hypothesis.**
   The main analysis will be conducted at the level of sharing decision, using a logistic regression predicting sharing (0=don’t share, 1=share) with robust standard errors clustered on subject and item. It will include the following independent variable dummies: labeled false in False Flag condition, labeled false in False+True Flags condition, labeled true in False+True Flags condition, unlabeled in False Flag condition, unlabeled in False+True Flags condition. Furthermore political concordance of the headline (-0.5 = discordant, 0.5 = concordant) and political leanings of the subject (Democrat vs Republican binary variable, z-scored) will be interacted with each of these dummies, and headline type (-0.5=false, 0.5=true) will be added to the interaction terms for the two untagged dummies. Our main tests will be for the existence of a warning effect in each treatment condition (indicated by the coefficients on the two “labeled false” dummies), a verified effect in the False+True Flags condition (indicated by the coefficient on the “labeled true” dummy), and the existence of an illusory truth effect (indicated by the coefficients on the two “untagged” dummies). We will also test whether the Warning Effect and Implied Truth effects differ significant between the two treatment conditions by testing whether the two coefficients significantly differ from each other.
6. **Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**
   Participants will be asked at the outset of the study if they have a social media account and if they ever share political content on social media (as well several other distractor questions).
   Participants who answer “no” to either of these questions will not be allowed to participate in the study.
   Also, we will ask respondents about whether they answered randomly or googled the sources, but we do not intend to remove individuals based on these questions.
7. **How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**
   We will aim to recruit 3000 participants on Mechanical Turk but retain all individuals who complete the study.
8. **Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**
   Nothing else to pre-register.

About the Authors

Eric Bogert. Eric is a Ph.D. Candidate at the University of Georgia in Management Information Systems. He has been published in the International Conference on Information Systems and the American Conference on Information Systems. Prior to graduate school he was a technology consultant at Deloitte Consulting.

Aaron Schecter. Aaron Schecter is an Assistant Professor of Management Information Systems at the University of Georgia, Terry College of Business. He received his PhD in Industrial Engineering and Management Science from Northwestern University. Professor Schecter's research interests include new forms of teamwork, digital innovation, and the role of artificial intelligence in the workplace. His work draws on computational social science techniques such as social network analysis, econometrics, and operations research modeling. Professor Schecter also works to develop and test new research methods leveraging event-based data.

Rick Watson. Richard Watson is a Regents Professor and the J. Rex Fuqua Distinguished Chair for Internet Strategy in the Terry College of Business at the University of Georgia. He is a former President of the Association for Information Systems. In 2011, he received the Association for Information Systems' LEO award, which is given for exceptional lifetime achievement in Information Systems. The University of Liechtenstein has established with government support a Consortium for Digital Capital Creation based on the ideas in his recent book, Capital, Systems, and Objects. For about a decade, he was Research Director for the Advanced Practices Council of the Society of Information Management and a visiting researcher at the Research Institute of Sweden (RISE) Viktoria in Gothenburg. Dr. Watson is a currently a senior visiting professor at the Southern University of Science and Technology (SUSTech) in Shenzhen, China.