Communications of the Association for Information Systems

Volume 49 Article 5

7-28-2021

Preregistration of Information Systems Research

Eric Bogert
Primary Author, etbogert@uga.edu

Aaron Schecter
Department of MIS University of Georgia

Richard T. Watson
Department of MIS University of Georgia

Recommended Citation

Bogert, E., Schecter, A., & Watson, R. T. (2021). Preregistration of Information Systems Research. Communications of the Association for Information Systems, 49, pp-pp. https://doi.org/10.17705/1CAIS.04905

Communications of the Association for Information Systems

Research Article | DOI: 10.17705/1CAIS.04905 | 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:

In this paper, we introduce the preregistration concept for experiments in the information systems (IS) discipline. Preregistration constitutes a way to commit to analytic steps before collecting or observing data and, thus, mitigate 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, and common arguments against preregistration. We also offer a call to action for authors to conduct more preregistered work in IS.

Keywords: Preregistration, Philosophy of Science, Experiments.

This manuscript underwent editorial review. It was received 09/04/2020 and was with the authors for two months for one revision. Mary Tate served as Associate Editor.

1 Introduction

More than 60 years ago, researchers posited that, if enough people study a relationship between two unrelated variables, then, eventually an author will find a statistically significant relationship and publish it (Sterling, 1959). Researchers refer to such errors as type I errors (i.e., false positives) and have suggested that they, along with other false findings, constitute a majority of published research (Ioannidis, 2005). One cannot easily invalidate a false positive finding even in replications because a replication might fail for many reasons (Simmons, Nelson, & Simonsohn, 2011). Furthermore, top-tier journals do not commonly publish null results, which means that, in the rare event that researchers replicate a study and find that it contains type I errors, the academic community has little chance to learn about it (Simmons et al., 2011).

False findings may constitute so much published research for many reasons, and we do not cover every reason. Rather, we focus on the issues that have a higher chance to affect information systems (IS) research. First, journals are more likely to false positive findings 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 journals are more likely to publish statistically significant results (Mahoney, 1977; Schwab, Abrahamson, Starbuck, & Fidler, 2011). Even after controlling for an experiment's quality, authors typically choose to not submit papers with null results (Franco et al., 2014). Authors likely perceive that various incentives encourage them to submit research with significant findings because the top journals in a discipline tend to publish far more papers with significant results than non-significant results. 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 involves finding significant results by running many underpowered 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 uses dubious research practices and even without a culture that encourages teams to publish or otherwise disseminate null results, when enough teams work on the same problem, they will invariably produce type I errors.

Third, authors are more likely to submit false results than true results because some authors might use dubious research practices. Researchers have documented the problematic techniques that authors use in NHST that invalidate its assumptions and make it more likely that they will publish null findings (Ioannidis, 2005). Hypothesizing after authors knows the results (HARKing) constitutes a process in which they analyze data until they find relationships that they believe they can publish and then find theories related to those findings and hypothesize accordingly (Starbuck, 2016). Authors may be particularly excited when a theory does not predict a relationship because journals are more likely to publish surprising findings (Brembs, Button, & Munafo, 2013). HARKing represents 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 from among multiple moderately correlated dependent variables (Simmons et al., 2011). Running multiple calculations with variable or data subsets can result in overly small p-values (Mertens & Recker, 2020; Starbuck, 2016). Researchers can usually fix running multiple calculations with a correction for multiple tests (Dunnett, 1955; Tukey, 1949), but, if they do not consistently record every test they run, then they cannot correct for multiple tests (Nosek, Ebersole, DeHaven, & Mellor, 2018; Starbuck, 2016). Unassailable evidence has shown p-hacking across disciplines. Political science contains approximately twice as many p-values immediately below 0.05 compared to immediately above 0.05 (Gerber & Malhotra, 2008). In economics, studies that examine how the minimum wage affects unemployment tend to have effect sizes twice the standard error and many different values for the standard error (Card & Krueger, 1995). Finally, one can easily "prove" that either Republican or Democratic politicians better benefit the economy depending on the variables one chooses (FiveThirtyEight, 2020).

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

In these replication projects, authors replicate original work and often follow a preregistered analysis plan. In preregistered plans, authors need 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. Many replication projects include papers based on how well they adhere to the original study's methodology and based on how frequently others have already replicated the hypothesis.

However, not all preregistrations are for replication projects. An increasing number of journals have begun to require or encourage researchers to preregister new experiments. The medical discipline has required preregistrations as the standard for decades, and United States' law requires preregistration in clinical trials since lawmakers passed 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 has roughly doubled every year in the most common preregistration repository (Kupferschmidt, 2018).

In this paper, we introduce the IS community to preregistering experiments (particularly in non-replication contexts). We focus on experimental research, although researchers can use much of the process in non-experimental research as well. We discuss preregistration's advantages and disadvantages, discuss the preregistration process, and provide a sample preregistration for a published experiment in *Management Science* in the appendix.

2 What is Preregistration?

Preregistration refers to "committing to analytic steps without advance knowledge of the research outcomes" (Nosek et al., 2018, p. 2601). One can preregister by posting an analysis plan in a scientific repository, such as the Open Science Foundation (OSF). While preregistrations commonly involve an experiment, researchers can preregister analysis plans for datasets not gathered through an experiment. For example, if researchers wanted to use government data that a government would soon release, they could preregister their analysis plan before the data appeared (Nosek et al., 2018). Similarly, if researchers wanted to examine a phenomenon that would soon unfold in social media (e.g., the months before an election or a predictable event such as the Olympics), then they could preregister their analysis plan for the upcoming event.

Researchers should not claim to have preregistered their study if they know too much about the data before they preregister. However, we lack a perfect rule to determine whether researchers know too much about the dataset to honestly claim that they made their analysis plan before research outcomes. For example, if researchers gave a dataset to their colleagues and told them about some interesting correlations they have already found, that would usually constitute too much information to honestly preregister future analysis plans (Nosek et al., 2018). Similarly, if researchers are already familiar with a relationship between a dependent variable and independent variable in their data, they should not preregister a study about two similar, likely highly correlated dependent variables and independent variables in the same dataset (Nosek et al., 2018).

Preregistration generally comes in two types. The first type, registered reports, requires 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 and, thereby, reduces publication bias by not providing a strong incentive for authors to find significant results. The second type (which we focus on in this paper) requires authors

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.

to submit 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 after a journal has published their paper. In both preregistration types, 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 they did not use questionable research practices in producing the focal paper due to the transparency that preregistration affords.

3 Why Preregister?

In this section, we explore four reasons that prior scholarship has introduced to justify preregistration's importance (Nosek et al., 2018; Roloff & Zyphur, 2019).

3.1 Reducing Publication Bias

Preregistration strongly predicts whether a study has true positive findings rather than false positive findings (Swaen, Teggeler, & van Amelsvoort, 2001). In their meta-analysis of medical trials, Kaplan and Irvin (2015) found preregistration to be strongly correlated with null findings. The International Committee of Medical Journal Editors now mandates that authors preregister experiments that adhere to the committee's policies. In short, preregistration constitutes the best way to reduce type I errors in a discipline.

3.2 Clarifying Prediction and Postdiction

Good science begins with openly and ethically explaining how one conducted the scientific process and the type of theorizing that one used. Preregistration makes it clear to readers whether theory informed a finding a priori or when post hoc analyses determined it (Nosek et al., 2015, 2018). Preregistration does not preclude authors from using inductive theorizing, but it necessitates that they clearly state they will use post hoc analyses.

This additional clarity makes it easier for one to assess what is theory generating versus theory testing (Nosek et al., 2018). While analyzing data, authors might find it easy 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 they originally envisioned their final model with significant results (Nosek et al., 2018). First, humans are more likely to conclude what they want to see due to motivated reasoning (Kunda, 1990). One's ability to justify one's conclusions constrains motivated reasoning. In ambiguous contexts such as scientific research, one can often justify actions in multiple ways such that one can justify almost any decision (Nosek et al., 2018). Second, humans are prone to hindsight bias; that is, when reflecting on past events to which they assigned probabilities, they tend to inflate those probabilities that they assigned to events that actually happened and deflate those probabilities that they assigned to 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 constitutes one partial antidote to p-hacking. An all-too-common syllogism in social science involves observing a phenomenon in one's data, writing a hypothesis about that data, and finding the hypothesis to be true. In doing so, authors conceal circular logic as theorizing (Nosek et al., 2018).

Of course, preregistered experiments can also be prone to p-hacking if 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 the authors will measure each variable. We openly admit that preregistration does not stop dubious research practices in all forms; for example, it cannot prevent authors from making up data entirely.

3.4 Multiple Testing

One can use many methodologies to adjust for multiple hypotheses tests on the same dataset. For example, a Bonferroni correction requires authors to divide the p-value required (usually 0.05) by the number of tests conducted (Bonferroni, 1936). Without preregistration, one would find it exceedingly difficult to identify how many tests authors conducted, which makes any adjustment to required p-values meaningless (Nosek et al., 2018). With preregistration, identifying how many tests an author conducted 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, which we argue would represent 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 theory informed their hypotheses. They can label any post hoc findings as such and, if possible, investigate them 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, authors can conduct follow-up experiments 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 envisioned when beginning a study. Due to hindsight bias and motivated reasoning, authors can often convince themselves that they envisioned their final model (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 analyses. 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). Preregistration can improve research design as one cannot “fix” many problems later. As such, preregistration can help strengthen studies’ validity and reduce the likelihood that authors need to rerun a faulty experiment.

4.4 Enabling Research Streams

Preregistration also helps enable research streams, which Gable (202) has recently highlighted as critical to the IS discipline’s success (Gable, 2020). When a study has surprising findings (whether a study finds significant and theoretically interesting but non-hypothesized interactions or lacks an expected effect), future research can investigate these effects. In non-preregistered studies, one may find it 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 authors answer in the affirmative, then they should preregister it (Simmons et al., 2011). Thus, authors do not need to preregister an 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 plays an important role in the scientific process, many preregistration templates do not require authors to explicate the theory behind the propositions and their operationalization as hypotheses because one can expect that the theorizing will take place in the published paper tied to the preregistration.

5.2 Proposition and Hypothesizing

We advocate a two-step procedure. First, authors should outline the propositions they will investigate. Second, they should specify how they will operationalize each proposition as they can often measure a concept in many ways and they should not adjust how they measure concepts after collecting data. Some preregistration templates allow authors to differentiate between exploratory hypotheses and confirmatory hypotheses.

5.3 Committing to Analytic Steps

Choosing an analysis plan generally constitutes the most important part of preregistration. The more specific the analytic steps, the fewer "researcher degrees of freedom" (Simmons et al., 2011). Among the most important factors to clarify include 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 dependent variables, flexibility in choosing a dependent variable almost doubles the likelihood that researchers will produce a false positive finding (Simmons et al., 2011). Similarly, if researchers willingly stop a study during data collection after finding a significant finding, they increase the chance of a false positive finding by about 50 percent (Simmons et al., 2011). Combining common researcher degrees of freedom raises the total probability that one produces a false positive to 61 percent; thus, researchers with high flexibility in analysis are more likely to find a false positive (Simmons et al., 2011).

Researchers commonly use these researcher degrees of freedom to falsely demonstrate significant results. After surveying academics, John, Loewenstein, and Prelec (2012) found that 70 percent reported stopping data collection early because they had found statistical significance. This finding likely does not assess the problem to its fullest extent either given that the study relied on self-reported behavior.

Choosing theoretically justifiable models represents strong science. If that model does not amount to the best-fitting one, researchers should acknowledge that and explain post hoc why they observed it. Researchers have the option to preregister that they will choose the best-fitting model (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. Authors can explain how they computed the number of subjects for an experiment (e.g., whether 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 cross-validating the data. IS papers that study large datasets commonly cross-validate data, and researchers can use cross-validation to test how well a model predicts a dependent variable on a held-out dataset (Lin, Lucas, & Shmueli, 2013).

5.3.2 Data Exclusions

Authors should list any process for excluding observations prior to gathering data. When conducting online experiments, such processes can include whether to include subjects who fail attention checks, manipulation checks, or subjects with duplicative IP addresses. Data exclusions in advance have particularly importance. For example, authors do not usually treat outliers in a consistent manner (Simmons et al., 2011). Consider time as a basis to exclude subjects; authors can exclude the fastest or slowest one percent or 2.5 percent of subjects, subjects 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 before their application.

Although most preregistration templates do not specifically ask for it, we advocate identifying possible robustness checks related to data subsets in advance.

5.4 Preregister on Repository

Authors need to choose what preregistration template and repository works best for them. AsPredicted.org provides perhaps the simplest and least burdensome template. Specifically, it 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 authors will measure, manipulate, define, and (if applicable) combine variables into an index. Some repositories require authors to indicate a date of public release when submitting a preregistration after which the author loses the ability to prevent the repositories from publicly releasing their research plan. In this way, they ensure that authors cannot bury preregistrations if they become inconvenient to authors.

We summarize important differences between three prominent repositories in the social sciences. One such repository, ResearchBox, launched in late 2020 and builds on AsPredicted. Specifically, it adds new functionality to AsPredicted such as the ability for authors to upload data and materials. We consider AsPredicted and ResearchBox a single consolidated platform: they share a server, have APIs that connect AsPredicted's preregistration to ResearchBox's platform, and both reside in the Wharton Credibility Lab. We compare their differences to the OSF and the American Economic Association (AEA) Randomized Control Trial (RCT) Registry in Table 1. All three platforms allow authors to upload preregistrations, data, and materials such as a survey from Qualtrics. One important distinction between the platforms concerns public preregistrations in that one cannot search them on ResearchBox, although it plans to release this functionality in late 2021. Many authors (ourselves included) 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 concerns 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 others have already done and provides an antidote to the "file drawer problem" in which authors do not publish null results (Franco et al., 2014). AsPredicted argues that forcing authors to publish all preregistrations may deter some preregistrations—a significant cost (AsPredicted, n.d.-a). Making all preregistrations public has two important benefits (AsPredicted, n.d.-a). First, authors cannot simply upload multiple preregistrations that each predict opposite effects for the same study. This practice would be unethical and against preregistration principles. AsPredicted's algorithmic sweeping has not detected this behavior as yet (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 before authors gather and analyze data such as in experiments or meta-analyses. 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 them as post hoc.

5.7 Report Results

Authors should report all preregistered hypotheses. If a study has multiple hypotheses, authors should not focus on supported hypotheses in order to avoid accidentally highlighting type I error results. If a preregistration has 20 hypotheses and authors find support for only one, the supported hypothesis has a relatively high chance to constitute a type 1 error. Similarly, completely dismissing hypotheses that lack support constitutes biased reporting. If a researcher conducts 20 highly similar experiments per year and finds one supported result per 20 experiments, the researcher should report the results from the other 19 experiments to provide the context necessary to evaluate whether the finding likely arose due to a false positive (Nosek et al., 2018).

5.8 Follow-up Studies on Postdictions

Preregistration lends itself to creating research programs. Conducting follow-up studies on results that authors found post hoc constitutes good practice and helps to build an edifice of IS research (Tiwana & Kim, 2019) and ensures that they can quickly find type I errors. We outline these processes below in Figure 1.

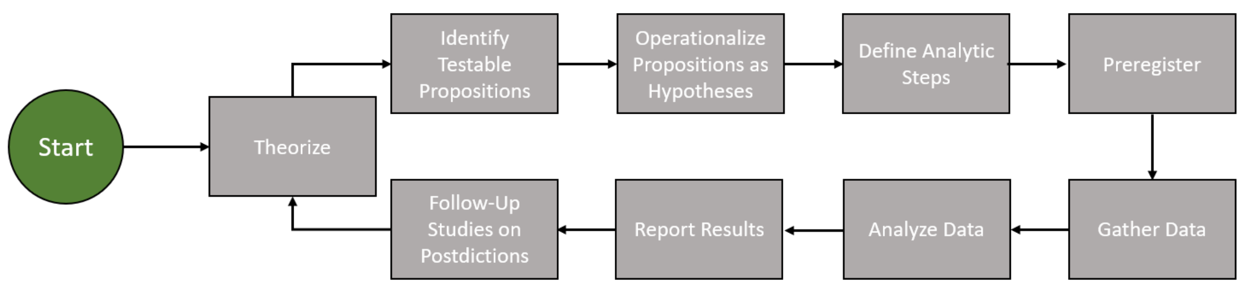

Figure 1. Preregistration Process

6 Arguments against Preregistration

In this section, we list some arguments against preregistration, though others may also exist. For additional work that analyzes 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, and so on, they 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, one can use certain sections from a preregistration in a paper. For example, most papers explain the variables, data-collection procedures, and sample, and authors will have mostly written these sections in advance if they preregister. Furthermore, the preregistration process can clarify thinking that saves valuable time throughout an investigation. Good planning, such as preregistration, usually saves time over a project's duration. For instance, the experiments we have preregistered usually take around 10 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 does not conform to their research philosophy. We do not mean to imply 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 want to conduct more exploratory research in areas that lack guiding theories to frame propositions, they can still preregister their studies. If authors plan multiple studies, they can preregister an initial study and investigate interesting post hoc findings in future preregistered, theory-testing studies. If they plan only one large study, we recommend that they preregister that they will cross-validate data (Fafchamps & Labonne, 2017; Nosek et al., 2018). After analyzing the first data set and drawing initial conclusions, they can test whether they can find support for those conclusions in the held-out data.

6.4 Replication Makes Preregistration Unnecessary

Although the IS discipline is not currently amid a replication crisis (Dennis et al., 2020), preregistration aligns with many recommendations that scholars have made to advance the discipline. 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). Authors should contain include much of this information in a preregistration, which makes almost no additional work for scholars who intend to make their work replicable while forcing them to carefully document these matters in advance rather than delaying until they report their research. The Open Science Foundation, which enables authors to upload materials in addition to preregistration, shows 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 benefit replication endeavors. The scholars who authored the example preregistration in Appendix 1 have included code, materials, data, and the preregistration in their submission to the OSF, which *Management Science* later published (https://osf.io/b5m3n/). We assert that preregistration and replication form 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.

Preregistration also adds value above and beyond replication for two other reasons. First, preregistration makes it clear what hypotheses a study did *not* find support for and, thereby, helps a discipline focus on more fruitful areas of inquiry. Second, preregistration helps prevent authors and journals from disseminating type I errors *prior* to publication (Nosek et al., 2018)—an important consideration because it becomes exceedingly difficult to correct the scientific record once a journal has published a paper. Thinking about replication in juxtaposition with a formal retraction represents 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—the most extreme, documentable, and public case of correcting the scientific record—many authors still positively cite retracted papers for their original findings even if the retracting journal retracted the paper for data fabrication, ethical misconduct, or false reports (Bar-Ilan & Halevi, 2017; Borenemann-Cimenti, Szilagyi, & Sandner-Kiesling, 2016). We expect that replications that 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 about the retraction, databases make no such indication next to papers that authors have failed to replicate. Thus, we should incentivize the IS academic community to stop errors *prior* to publishing through preregistration.

7 Extending to Non-experiments

Authors most commonly use preregistration in experiments and clinical trials, but preregistration has broader applicability to all empirical research. For example, qualitative researchers can preregister who they intend to interview, their interview questions, interview location, 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 the IS discipline already use preregistration in studies that involve archival data such as natural experiments. One could even preregister literature reviews or meta-analyses. In the first preregistered paper published in one of the Senior Scholars' basket of eight journals, Mertens and Recker (2020) conducted a meta-analysis of null-hypothesis significance testing in IS. The authors preregistered how they sampled papers from top journals, their hypotheses about how those papers used the NHST, and their analysis plan. The IS discipline 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 that established frameworks offer. This process will also help authors respond to recent calls for increased transparency in literature reviews (Templier & Paré, 2018). In analyzing 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 variable measurements. Meta-analyses frequently list the databases from which authors will take papers, inclusion criteria, exclusion criteria, and analysis plans.

8 Call to Action

Authors, reviewers, and journals have responsibility for preregistration's future in the IS discipline. 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 represents a low-cost, impactful commitment to academic practice and quality, and it already commonly occurs in many journals and disciplines. Authors can choose to be at the forefront of best practices in research while increasing their work's replicability. Authors who embrace preregistration may also find that it helps reduce errors before they happen by forcing them 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 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 result from analyzing a preregistered experiment post hoc, then reviewers, readers, and editors should be willing to consider them if (and only if) the authors report that they did not preregister the models.

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

8.3 Journals

The Transparency and Openness Promotion (TOP) guidelines give journals a score between zero and three based on the degree to which they accept preregistration (Mellor, 2020). They define each score as follows:

0) Journal says nothing about preregistration.
1) Papers state if authors preregistered work.
2) Papers state if authors preregistered work. Journal verifies adherence to preregistration plan.
3) Journals require authors to preregister confirmatory or inferential research.

As of this writing, the TOP guidelines have scored *Nature*, *Science*, and *The Proceedings of the National Academy of Sciences* one and *Nature Human Behavior* two. Relatively few journals have policies that would grant them a score of 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" (Association for Psychological Science, n.d.). Lastly, the editor-in-chief of *MIS Quarterly* has stated that preregistration exemplifies an editorial innovation he has considered exploring for the journal[3].

Preregistration advocates offer a fourth option that journals can use to embrace preregistration: offering registered reports (Graf, 2017). Registered reports allow journals to commit to publishing work prior to data collection without knowing the outcomes. In this way, they reduce incentives to engage in dubious research practices.

The data indicate that preregistration has begun to gain acceptance as a scientific convention, and IS journals will possibly adopt the practice at some stage. The Association for Information Systems (AIS) constitutes the appropriate body, through its journals, to lead the IS discipline in adopting a norm that promotes research excellence since the organization "serves society through the advancement of knowledge and the promotion of excellence in the practice and study of information systems" (AIS, n.d.).

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 discuss and overview preregistration, which, while not a cure all for accepting biased significant findings, represents a step in the right direction. Preregistration aligns with recent calls to action for more programmatic research in which studies cumulatively build on each other using a thoughtful, long-term strategic plan to thoroughly investigate a phenomenon (Gable, 2020).

We should note that it is not always dubious or unethical to allow data to develop new theories. Carefully analyzing data can result in robust, important findings and has for centuries. For example, scientists discovered the electron 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 resulting from theory generation rather than 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 represents the best step to ensure that this distinction remains salient in disseminating academic research.

[3] He stated such comments in a presentation to MIS department at the University of Georgia on 16 October, 2020

References

Achinstein, P. (2001). *The book of evidence*. Oxford, UK: Oxford University Press.

AIS. (n.d.). *About AIS*. Retrieved from https://aisnet.org/page/AboutAIS

AsPredicted. (n.d.-a). *Why does AsPredicted allow pre-registrations to remain private forever?* Retrieved from https://aspredicted.org/messages/private_forever.php

AsPredicted. (n.d.-b). *Terms of use*. Retrieved 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*(1), 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*(4), 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*, *8*, 3-62.

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., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, 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., Johannesson, M., Kirchler, M., Nave, G., Nose, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N. Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J., & 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*(2), 238-243.

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

FiveThirtyEight. (2020). *Hack your way to scientific glory.* 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. *MIS 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), 1072-1102.

Nosek, B., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B. A, Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., & 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*, *66*(11), 4944-4957.

Association for Psychological Science. (n.d.). *Preregistration of research plans*. 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*(3), 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, J. W. (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

Readers can find the following preregistration at https://osf.io/tvbj5/. This preregistration belongs 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 papers than they would if no papers 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 3,000 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 is a PhD candidate at the University of Georgia in Management Information Systems. He has been published in the proceedings of 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 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. His 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. He also works to develop and test new research methods leveraging event-based data.

Rick 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. He is a currently a senior visiting professor at the Southern University of Science and Technology (SUSTech) in Shenzhen, China.