Communications of the Association for Information Systems

Volume 52 Paper in press

April 2023

HARKing and P-Hacking: A Call for More Transparent Reporting of Studies in the Information Systems Field

Han-fen Hu
Department of Management, Entrepreneurship, and Technology University of Nevada, Las Vegas, han-fen.hu@unlv.edu

Gregory D. Moody
Department of Management, Entrepreneurship, and Technology University of Nevada, Las Vegas

Dennis F. Galletta
Katz Graduate School of Business University of Pittsburgh

Recommended Citation
Hu, H., Moody, G. D., & Galletta, D. F. (in press). HARKing and P-Hacking: A Call for More Transparent Reporting of Studies in the Information Systems Field. Communications of the Association for Information Systems, 52, pp-pp. Retrieved from https://aisel.aisnet.org/cais/vol52/iss1/37

Communications of the Association for Information Systems

Accepted Manuscript

HARKing and *P*-Hacking: A Call for More Transparent Reporting of Studies in the Information Systems Field

Han-fen Hu
Department of Management, Entrepreneurship, and Technology
University of Nevada, Las Vegas
han-fen.hu@unlv.edu

Gregory D. Moody
Department of Management, Entrepreneurship, and Technology
University of Nevada, Las Vegas

Dennis F. Galletta
Katz Graduate School of Business
University of Pittsburgh

Please cite this article as: Hu, H-f., Moody, G. D., & Galletta, D. F. (in press). HARKing and *P*-hacking: A call for more transparent reporting of studies in the information systems field. *Communications of the Association for Information Systems.*

Communications of the Association for Information Systems

Research Paper ISSN: 1529-3181

HARKing and *P*-Hacking: A Call for More Transparent Reporting of Studies in the Information Systems Field

Han-fen Hu
Department of Management, Entrepreneurship, and Technology
University of Nevada, Las Vegas
han-fen.hu@unlv.edu

Gregory D. Moody
Department of Management, Entrepreneurship, and Technology
University of Nevada, Las Vegas

Dennis F. Galletta
Katz Graduate School of Business
University of Pittsburgh

Abstract:

While researchers are expected to look for significant results to confirm their hypotheses, some engage in intentional or unintentional HARKing (Hypothesizing After Results are Known) and *p*-hacking (repeated tinkering with data and retesting). If these practices are widespread, one possible result is field-wide exaggerated (inflated) results reported in Information Systems (IS) publications. In this paper, we summarize the literature in HARKing and *p*-hacking across different disciplines. We offer an illustrative example of how an IS study could involve HARKing and *p*-hacking in various stages of the project to generate a more "publishable" result. We also report on a survey targeted at IS researchers to explore their experiences and awareness of this issue. Finally, we provide recommendations and suggestions based on the review of practices in other fields and advocate for more transparency in reporting research projects, so that study results can be interpreted properly, and reproducibility and replicability can be increased.

Keywords: *p*-Hacking, HARKing, Transparent Reporting, Questionable Research Practices, Selective Reporting.

[Note: All authors contributed equally.]

1 Introduction

Researchers in various fields have raised concerns about journals only publishing hypothesis-testing studies that show significant results supporting many of the hypotheses (e.g., Bosco, Aguinis, Field, Pierce, & Dalton, 2016; Bruns & Ioannidis, 2016; Simonsohn, Nelson, & Simmons, 2014). It has been found that papers without significant results have lower publication rates (Franco, Malhotra, & Simonovits, 2014), and many become filed away or placed at a low priority in researchers' to-do lists. Many of those papers fail to ever see the light of day. That "file drawer effect" leads to two types of biases in the literature, known as HARKing and *p*-hacking, described below. These practices can mislead researchers and make it difficult to know what results can be trusted.

HARKing is defined as "presenting a *post-hoc* hypothesis (i.e., one based on or informed by one's results) in one's research report as if it were, in fact, an a priori hypotheses" (Kerr, 1998, p. 196). HARKing can be observed when researchers derive hypotheses from the analysis results or suppress hypotheses that are not supported by the data (Kerr, 1998).

P-hacking "occurs when researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results" (Head, Holman, Lanfear, Kahn, & Jennions, 2015, p.1). *P*-hacking practices can occur at various stages of a research project, including study design, data collection, data analysis, and reporting the results. It often occurs as "researcher degrees of freedom," (Simmons, Nelson, & Simonsohn, 2011, p. 1359) when researchers make various decisions in the course of designing and conducting a research project; however, in a quantitative study, design and analysis decisions should be made in an *a-priori*, rather than *post-hoc*, manner.

Both HARKing and *p*-packing result in bias in the published results, with a higher proportion of supported hypotheses than is warranted. Omitting some subjects or making use of multiple constructs can sometimes nudge some marginal hypotheses into significance. Also, most or all hypotheses without support can be left out of a paper entirely. In an extreme case, at the widely-accepted .05 level of significance, if 100 hypotheses are tested with data including only random numbers, five would be expected to be marked "significant." If a researcher reports only on the five, the paper appears strong.

It is suggested that the root cause of the issue in a hypothesis-testing paper is a perceived or real need for finding results. Although the research outcomes are still reported following the format of rigorous scientific studies, bending or twisting the results through manipulation of hypotheses presented, study design, data collection, analysis methods, and result reporting increase chances of false positive results (Mertens & Recker, 2020; Simmons et al., 2011) and can inflate effect size estimates (Simonsohn et al., 2014; Wicherts et al., 2016).

Both HARKing and *p*-hacking are consequences of a common focus on confirmatory and statistically significant results for studies taking a hypothetico-deductive approach (Kerr, 1998). The confirmation approach is common in our field, as it is in other softer science fields such as Psychology, which share with us a lack of "hard sciences" universal or absolute properties (see Sanbonmatsu et al., 2015). Journals have an unspoken practice of favoring papers with significant results over those that do not find significance. Researchers have an implicit motivation to report study results that match predictions, in order to achieve acceptance and publication of their papers. Failure to find statistically significant results could be due to factors such as poor theorizing, poor instrumentation, or poor controls, possibly signaling a "failed study." Of course, another reason for the lack of statistically significant findings could be that the theorizing has flaws (such as boundary conditions) that should become known to other researchers; rejecting these papers will result in distortion of knowledge in a field. Taken to an extreme at the level of theory, if 100 researchers study a theory that has no merit, it would be expected that 5 papers would boast significant findings. The other 95 papers would most likely be rejected, and thus the field would perpetuate a flawed theory. Another more realistic example occurs when studies explore the boundary conditions for a theory. It is critical to show under what conditions the theory would not apply, evidenced by insignificant hypothesis testing results, so that the field can learn if the theory is not universal and needs to be applied only within the boundaries specified. If those boundary-testing papers are rejected, the field would also perpetuate a theory without knowing its limitations, which is also considered flawed. According to Popper's (1959) theory of falsification, disproving or disconfirming a theory is critical for scientific advancement; therefore, study results that contradict prior findings could provide new insights into certain phenomena.

Researchers in several fields have actually been found to engage in HARKing or *p*-hacking behaviors in conducting research and/or reporting the study results to maximize the number of supported hypotheses. The website RetractionWatch.com collects and displays instances where egregious manipulations are discovered. Unfortunately, more subtle infractions might never be discovered.

Studies have shown that HARKing and *p*-hacking are common in natural science and social science research alike. It would be naïve to assume that the Information Systems (IS) field is immune from such practices. Based on our analysis of the articles published in top journals between 2013-2017,[1] we found that more than 80% of the hypotheses reported in the quantitative articles were supported (see Appendix A for details). Although we are not aware of any retractions from any of the top journals in our field, we suspect that published studies in IS have been marred by HARKing and *p*-hacking biases, potentially to the same extent found to exist in other fields.

To highlight this potential concern in IS publications, we review the literature to reveal the findings from other fields and summarize recommendations for addressing these issues. We also provide an example from unpublished prior research of one of the authors and use a dataset from an experiment on the impacts of color on mood and website satisfaction to illustrate how HARKing and *p*-hacking can artificially generate significant results supporting the hypotheses. We also conducted a survey of IS researchers to understand their awareness of the issues and their own experiences with these practices. In the conclusion, we offer some recommendations for the field to address this issue.

2 Background and Literature Review

HARKing and *p*-hacking have been found to be alive and well in widely-varying fields, such as biology (Head et al., 2015), pharmacology (Motulsky, 2015), chemistry (Sweedler, 2015), medical imaging (Kwee, Almaghrabi, & Kwee, 2023), psychology and marketing (Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012; Shanks et al, 2015; Wicherts et al., 2016), management and organizational research (Bettis, 2012; Bosco et al., 2016; Honig, Lampel, Siegel, & Drnevich, 2014; Starbuck, 2016), and finance (Harvey, 2017). Scholars in different fields have suggested some methods to observe or detect HARKing and *p*-hacking. For example, Kerr (1998) has suggested some symptoms that can be observed from articles, such as convenient qualifiers, “perfect” theories (with all hypotheses supported), or omitted methodology details. Nevertheless, there are no certain diagnostic cues that can be used to determine whether a study is *p*-hacked (Kerr, 1998).

More systematically, meta-analysis has been used to investigate the publication biases issue (e.g., Dalton et al., 2012; Head et al., 2015). One popular approach taken is using *p*-curving, the distribution of significant *p*-values across a set of studies, to identify the percentage of possible inflation of the published articles in a certain field (Simonsohn et al., 2014). However, *p*-curving has been criticized for the potential sample selection issue that leads to an incomplete or non-representative set of studies and effects (Cuddy, Schultz, & Fosse, 2018), and *p*-curves in some cases cannot reliably distinguish true effects and null-effects with *p*-hacking (Bruns & Ioannidis, 2016).

While HARKing primarily relates to formulating hypotheses after knowing the results, common *p*-hacking practices can be observed at different stages of a study; it is also known as analytical elasticity (Harris, Pashler, & Mickes, 2014), data-dredging, snooping, fishing, or significance-chasing (Nuzzo, 2014). As summarized in Table 1, *p*-hacking practices can occur in any research project (Head et al., 2015; Motulsky, 2015; Nuzzo, 2014; Simmons et al., 2011; Sweedler, 2015; Wicherts et al., 2016).

[1] We analyzed the Association for Information Systems (AIS) basket of six journals: MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of Association for Information Systems, European Journal of Information Systems, and Information Systems Journal.

Table 1. Common HARKing and *p*-Hacking Practices

Category	Examples
Formulating intentionally vague hypotheses	• Having vague hypotheses without specifying the direction of the effects
Including redundancy in study design to create maneuverability in data analysis	• Creating multiple manipulated independent variables • Measuring additional variables • Measuring the same dependent variable in several alternative ways
Making *post-hoc* decisions regarding data collection	• Conducting analyses midway through studies to decide whether to continue collecting data • Choosing between different options of dealing with incomplete or missing data on *ad hoc* grounds
Making *post-hoc* decisions regarding data analysis	• Determining outlier treatment after analyzing the data • Excluding, combining, or splitting treatment groups post-analysis • Including or excluding covariates post-analysis • Stopping data exploration if an analysis yields a significant *p*-value • Choosing between different statistical models • Transforming data (i.e., logarithms, normalization) • Comparing the results of using different control variables and/or outcome variables
Reporting the study with some details omitted	• Selective reporting of variables included in study • Reporting without details of the study that prevents reproducibility or replicability • Misreporting results and *p*-values

Researchers of confirmatory quantitative studies view an ideal study as designed and conducted with the specific purpose of validating a set of propositions or hypotheses. There is significant incentive to design the studies or analyze the data in a way that would be likely to generate favorable results, and some cases have surfaced that involve unconscious *p*-hacking (Nuzzo, 2014; Sweedler, 2015). To illustrate, researchers may make some *post-hoc* decisions in data analysis when the distribution of data is not as expected, without any intention to *p*-hack the data. Quantitative IS studies that draw on null hypothesis significance testing (NHST) are susceptible to *p*-hacking (Mertens & Recker, 2020).[2] Based on our analysis of the articles published in top journals between 2013-2017, around 40% of the published research papers fall into this category. It is therefore critical for researchers to be aware of the issues and common practices of *p*-hacking so that it can be avoided.

To help clarify HARKing and *p*-hacking, we include an illustrative example. Thus, through "operationalizing" our conceptual discussion of *p*-hacking in the next section, the example provides a concrete illustration of various types of *p*-hacking, as well as the need for some potential solutions.

3 Illustrative Example

The data were collected as part of a master's degree program of one of the authors and has never been published. The initial research question of the project was whether the use of cool, neutral, or warm colors on a website would make it more satisfying for those using the website. The theory behind the research question is fairly weak but is based on previous research on sociological perceptions of color. It was hypothesized that warm colors (i.e., reds and yellows) would result in higher satisfaction with websites, while cool colors (i.e., blues and greens) would result in lower satisfaction. Likewise, it was predicted that this impact would be mediated by the mood-state of the individual visiting the site. Specifically, it was proposed that warm colors would evoke more aroused mood-states and cool colors would produce lower aroused mood-states, and that the impact of mood-state on satisfaction would be direct and linear. The original hypotheses were:

> **Original H1: When warm colors are used as the background of a website, it would evoke more aroused mood-states for the users, as compared to a website that cool colors are used.**

[2] The null hypothesis significance testing (NHST) paradigm is a term that also includes testing of directional hypotheses. NHST has been criticized for low replicability; some researchers even suggest using a *p*-value as just one among many pieces of evidence, rather than a threshold screening role (McShane et al., 2019). However, currently statistical significance is still widely used by many fields for NHST studies.

Original H2: When warm colors are used as the background of a website, the user would be more satisfied with the website, as compared to a website that cool colors are used.

Original H3: A website user's aroused mood-state would be associated with a higher level of satisfaction with the website, as compared to a lower aroused mood-state.

To test these few hypotheses, a study was designed with the following treatment conditions:

- Website (2 levels; between subjects): weather and music listening
- Color (2 levels; between subjects): warm and cool. Color was used as the background shade only, and all other elements of the website were constant across conditions
- Arousal (2 levels; within subjects): arousal and avoidance tasks on the consequent webpage (music listening). Arousal tasks were focused on creating a pleasant mood; subjects were asked to listen to whatever made them feel good. In the avoidance condition, subjects were told that they were going to be quizzed on the ranking of the music in the Top 100 Billboard Chart

The instrument to collect mood-state and satisfaction came from previous validated sources. The following control variables were gathered: gender, age, years in college, geographic/demographic background, and ethnicity. A total of 119 subjects were recruited from introductory information systems classes for this study. Subjects were randomly assigned to the between-subjects conditions. Thus, each cell had roughly 40 for the Website and Color treatments and 60 in each of the arousal vs. avoidance conditions.

After establishing the validity of the measurement, the first steps for analysis were to form the scores of the two reflective constructs. Mood consisted of the loaded factor scores of the 12 items from the telic-paratelic scale of mood-state. Satisfaction was formed from the factor loadings of its six items. Given the nature of the hypotheses, the effects of treatments on mood and satisfaction were tested using analysis of variance (ANOVA). The results are shown in Table 2, revealing the nonsignificant effects of the treatments. Figure 1 also shows that the mood-state and satisfaction are not statistically different in the three color conditions; the widely-used "box and whisker" diagrams, or simply boxplots (Dutoit, et al. 2012), show substantial overlap between the treatments within one standard deviation above and below the means.

Table 2. Summary of ANOVA Results

Dependent Variable	*F*-value	*p*-value	Significant?
Mood	1.18	0.309	No
Satisfaction	0.10	0.908	No

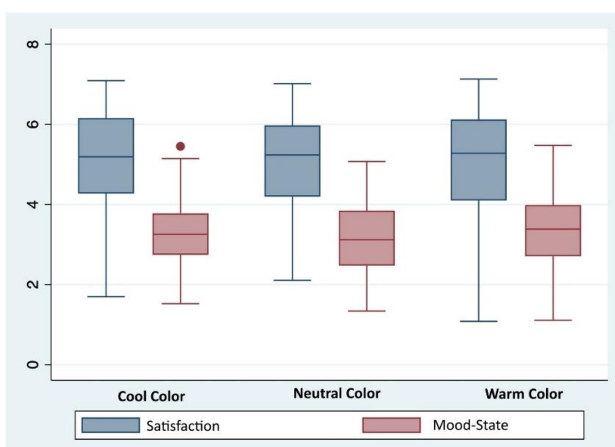


Figure 1. Distribution of Satisfaction and Mood-State across Conditions

Thus, the original H1 and H2 regarding the impacts of color on mood and satisfaction are not supported. The remaining hypothesis, impact of mood on satisfaction is supported, based on a regression model (b = 0.278; t = 5.71, p < .000, R2 = 0.098). Figure 2 also shows the positive relationship between satisfaction and mood-state. Thus, the effects of color did not seem significant, while the original research question.

H3 was supported. Unfortunately, the impact of mood on satisfaction is already well-known and the finding is not a novel contribution to research. Further, the correlation has little use in actual decision-making.

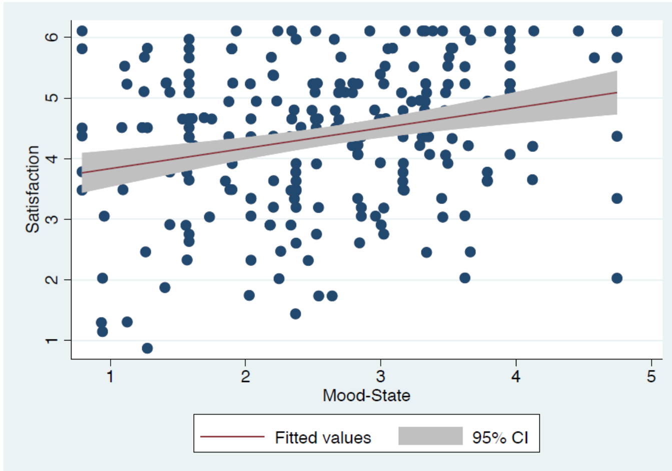


Figure 2. Distribution between Satisfaction and Mood-State

Given that the results are not as exciting as they had been expected to be upon completion of the analysis, some authors could engage in some steps to find more promising results that might be considered a contribution to research. The initial step that could be taken is often called a “fishing” or “dredging” expedition (Awati, 2022), where all of the variables are put into a model to find all relationships empirically. An analysis of the correlation of the control variables, treatments, and reflective constructs reveals that several pairs of variables are significantly correlated. Specifically, age negatively correlated with satisfaction; mood positively correlated with website and arousal; age negatively correlated with gender; and ethnicity correlated with some demographic items. With this knowledge it becomes possible to form more interesting hypotheses. Specifically, the impacts of arousal and website on mood can be examined, and the authors could create two new hypotheses that are known to be supported, namely:

> **New H4: An induced state of arousal will intensify one’s mood-state when compared to an avoidance induction.**

> **New H5: A more interactive and hedonic website will positively increase one’s mood-state when compared to a utilitarian website.**

These are in fact novel hypotheses and have a better potential to contribute to research given that it examines mood-states and how websites can alter mood-states, which in turn impact satisfaction, an important dependent variable in e-commerce research. To test these hypotheses, ANOVA was performed on the website and arousal treatments, which was collected as part of the research efforts for another potential study, as collection efforts were combined to maximize the use of subjects. Table 3 summarizes the analysis of H4 and H5. The result is also illustrated in Figure 3; the boxplots show that the data distributions are statistically different between the conditions.

Table 3. Summary of new ANOVA Results

Treatment	*F*-value	*p*-value	Significant?
Website (Music compared to weather)	40.28	0.000	Yes
Arousal (compared to avoidance)	20.35	0.000	Yes

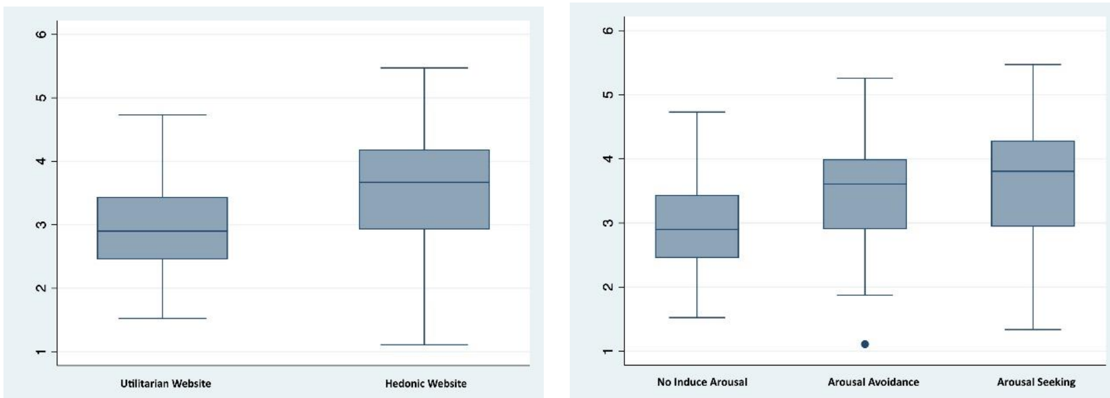


Figure 3. Distribution of Mood-State across Conditions

These results are strongly supportive of the treatments of website and arousal. Thus, if the hypotheses were altered to drop H1 and H2 and include H4 and H5 in their place, the paper would then have results that show that how subjects perform on a website, and the nature of the website both are strong predictors of the mood-state induced on the website, which significantly impacts website satisfaction. This is a more compelling story for eventual publication and could become the final model and analysis to be proposed in the paper.

These steps would not be revealed in the final paper, rather H3-5 would be the only proposed hypotheses and any mention of color and its impact would have been removed from the final, cleaned-up paper. Without disclosing that the model and analysis have changed, this study would have exhibited both HARKing and *p*-hacking.

To sum up, this illustrative example involved several HARKing and *p*-hacking behaviors. First, the study included a manipulated independent variable (i.e., website) that is not in the original hypotheses, just in case the original design would not work out as expected. Second, after seeing the original hypothesis testing result, the authors attempted different analysis methods and formulated new hypotheses to "generate" significant outcomes. Third, a post-analysis decision was made to use additional data in combination with the original study design. Fourth, the new hypotheses were proposed after the data were analyzed; the post-analysis decisions were not revealed in reporting the study.

4 Survey of IS Research Practices

As described above, we began this work without a sense of how much HARKing and *p*-hacking occur in our field. Evidence of these practices is completely absent from published quantitative papers. Therefore, we designed a survey focusing on those questionable research practices.

4.1 Data Collection

The items for questions about research behaviors were adapted from John, Loewenstein, and Prelec (2012). We also added a few questions focusing on the review process for journal publications based on our own experiences. To understand the possible drivers behind the questionable research practices, we adopted items from Boss, Galletta, Lowry, Moody, and Polak (2015) to examine the threat appraisal and coping appraisal of research practices from the perspective of Protection Motivation Theory. The threat appraisal includes fear and severity appraisal. Fear is defined as a feeling aroused in response to a dangerous situation. In the context of this study, it represents researchers' fear toward the condition that *p*-hacking is widespread in our discipline. Severity is the degree to which an individual believes the threat will cause consequential harm and captures whether a researcher considers *p*-hacking as a serious threat in our discipline. The coping appraisal focuses on response cost, which is the perceived personal cost of taking protective (response) actions (see Boss et al., 2015 for supporting this broader use of the word "cost"). In this study, "cost" refers to the high risk incurred in avoiding *p*-hacking, given that a study with no significant results would represent wasted effort and time in performing a non-publishable study. The items are listed in Appendix B.

The survey was distributed via the AISworld Listserv (http://listserv.aisnet.org/) in April 2020 and direct email invitations for authors who published in top IS journals between 2013-2020 (i.e., the original AIS basket of six journals) in February 2021. We received 118 complete responses (55 from the AIS listserv and 63 from the authors of top journal articles). We compared the results of the two waves of data collection and found no significant differences (all *p*-values are > 0.1) except one question[3].

4.2 Measure Assessment

An exploratory factor analysis revealed that four factors can be extracted from the items of questionable quantitative research practices. Each factor is defined as following:

- Review panel-driven HARKing: HARKing as requested by the review panel.
- Author-driven HARKing: HARKing before submission.
- Statistical hacking: Trying out several statistical analyses to obtain significant results.
- Selective reporting (reverse): Report the data analysis process not in a transparent manner.

For the fear and coping appraisal items, the factor analysis also showed that fear appraisal and severity appraisal loaded on the same factor while response cost loaded as its own factor. Thus, we combined the fear and severity appraisal and re-defined the factors as following:

- Fear and severity: The degree to which an individual believes *p*-hacking and HARKing will cause consequential harm.
- Response cost: The costs of engaging in HARKing and *p*-hacking will not exceed the perceived benefits.

The final items for each of the six factors are summarized in Appendix C, along with their descriptive statistics on a 7-point Likert scale. Each scale indicated each respondent's strength of agreement with several statements addressing each of the factors. We next examined the reliability and validity of the variables to ensure the quality of the measurement (see Appendix D for the details). The summarized descriptive statistics of the factors are shown in Table 4. The level of selective-reporting (mean = 4.81; standard deviation = 1.47) is the highest of the four types of questionable research practices, based on the results of pairwise comparisons using Bonferroni correction ($p < 0.0001$ for all comparisons). It indicates that selective reporting in the most widely-engaged questionable practice in the sample. In addition, the fear and severity appraisal is the higher of the two independent variables (mean = 4.74; standard deviation = 1.44), based on the result of a two-tailed t-test ($t = 4.99$; $p < 0.0001$); it shows that the respondents recognized and were concerned about the issue of *p*-hacking.

Table 4. Descriptive Statistics of Factors

	Factors	Mean	Std. Deviation
Dependent variable	Review panel-driven HARKing	2.38	1.80
	Author-driven HARKing	2.55	1.44
	Statistical hacking	1.32	0.85
	Selective reporting	4.81	1.47
Independent variable	Fear and severity	4.74	1.44
	Response cost	3.93	1.64

4.3 Results

As summarized in Table 5, the survey results showed that quantitative researchers in the IS field are aware of the issue. The average level of awareness of HARKing and *p*-hacking is 3.99 (standard deviation = 2.19) on a 7-point scale. In addition, as shown in Table 5 and Figure 4, on average the respondents estimated 40.38% (standard deviation = 30.54) of our discipline engaged in *p*-hacking, and they self-reported engaging in *p*-hacking in 9.92% (standard deviation = 19.77) of their own projects in the last five years.

[3] The question is "I think it is likely that *p*-hacking will continue in our discipline." The mean of the AIS listserv data is 2.47 (standard deviation = 1.76) while the other is 2.06 (standard deviation = 1.51).

Table 5. Descriptive Statistics of Awareness of HARKing and *p*-hacking in IS

Question	Scale	Mean	Standard Deviation
I have specific awareness of cases in which others in our discipline have engaged in *p*-hacking.	1-7	3.99	2.19
I estimate that *p*-hacking is engaged in by the following percentage of our discipline, on a regular basis.	1-100	40.38	30.54
Over my entire career, I have engaged in *p*-hacking for what percentage of my projects?	1-100	13.02	22.54
Over the last five years, I have engaged in *p*-hacking for what percentage of my projects?	1-100	9.92	19.77

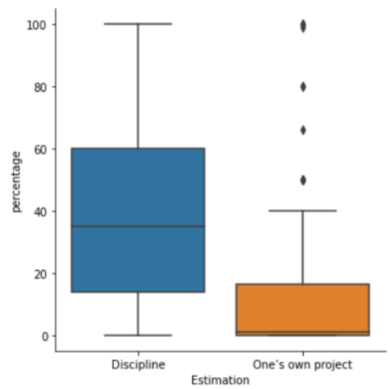
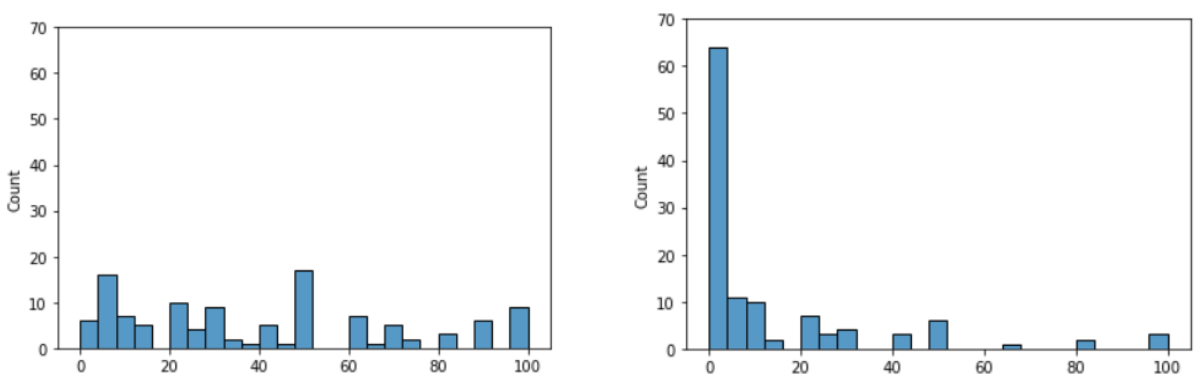


Figure 4. Respondents' Estimation of *p*-hacking Prevalence

Figure 5(A) and (B) further show the distribution of respondents' estimations of our discipline engaged in *p*-hacking, and they self-reported engaging in *p*-hacking. While more than 60 respondents indicated that they never engaged in *p*-hacking in their own projects, the majority of them estimated that a large number of researchers in our discipline engaged in *p*-hacking to some extent.



(A) Estimated percentage of *p*-hacking in our discipline (B) Estimated percentage of *p*-hacking in one's own project

Figure 5. Distribution of Respondents' Estimation of *p*-hacking Prevalence

Next, we used ordinary least squares (OLS) regressions to explore the associations between the appraisal factors and the four types of HARKing and *p*-hacking behaviors, in order to show how likely the threat appraisal (i.e., fear and severity appraisal) and coping appraisal (i.e., response cost appraisal) are associated with questionable research behaviors. We summarize the result of the OLS regressions in Table 6. The results suggest that the participants' fear and severity concerns of *p*-hacking is positively associated with their experience of review panel-driven HARKing. One of the possible explanations is that the respondents have been requested by the review panel to HARK, and therefore they were concerned about such practices in the IS field. In addition, their appraisals regarding response cost (i.e., the assessment that avoiding HARKing and *p*-hacking could take extra effort to explain and justify the procedures, and lead to a paper rejection due to the non-standard data analysis) is positively associated with all four types of questionable behavior.

Table 6. Summary of Regression Results

	Dependent Variables			
	Review panel-driven HARKing (R^2=0.11)	**Author-driven HARKing (R^2=0.07)**	**Statistical hacking (R^2=0.08)**	**Selective reporting (R^2=0.16)**
Fear and severity	0.20**	0.04	0.16	0.12
Response cost	0.25**	0.27**	0.23**	0.37***

Selective reporting is perceived to be the most common questionable research practice (please see Table 4), and the appraisal factors explain 16% of its variance, mainly by response cost. It implies that researchers engage in selective reporting to avoid problems in detailed reporting or reduce effort in reporting the statistical process. It is also not a common practice in our field to find published papers containing narratives about steps taken after original hypotheses failed to be supported.

5 Recommendations

To address the issues of HARKing and *p*-hacking, we would like to offer some recommendations for editors and reviewers of journals, researchers, and the IS field in general.

5.1 Recommendations for Editors and Reviewers

The root-cause of HARKing and *p*-hacking is the need for finding significant results in quantitative papers to increase the likelihood of acceptance, and ultimately to support authors' promotion and tenure cases. Such practices lead to inflated results. Reviewers and editors of peer-reviewed journals are already quite busy, and these results-exaggerating practices can be impossible to discover. Therefore, closer scrutiny of papers might not yield reduction or elimination of such misrepresentations. However, it could be more effective to simply require transparency, encouraging authors to report results that might have been contrary to their original expectations, showing revised hypotheses in those cases along with logic supporting those revisions. Also, being somewhat more tolerant of non-significant results in rigorously-conducted quantitative studies might actually provide value to other researchers.

Such non-significant results should be accompanied by detailed reports of study designs and procedures. Some of those reports can display how the original theorizing was flawed, and to provide advice about alternative theories or modified application of the theories that were used. Rather than exhibiting weakness, such reports can be quite instructive and interesting for future studies. We would also suggest that editors write editorials and communicate in public forums that they encourage well-done submissions by quantitative researchers that have insignificant hypothesis-testing results. Studies that disconfirm prior theories can form the basis of new theory generation and advance our knowledge in the field.

While "the research transparency movement is an institutional change in the field of science" (Burton-Jones et al., 2021, p. iii), it requires the efforts of both authors and review panels to promote transparent and truthful reporting. In the editorial of *MIS Quarterly* published in June 2021, Burton-Jones, Boh, Oborn, and Padmanabhan (2021) discuss the steps *MIS Quarterly* has taken to address research transparency. In the editorial, Burton-Jones et al. (2021) discusses the importance of transparency materials and the journal's new transparency requirements. The new transparency guidelines can help to reducing *p*-hacking to some extent because it requires many data analysis details (e.g., how missing values are managed, the main statistical tests in the paper, and key statistical output files) to enable reproducing the analysis in the paper. In this regard, the *MIS Quarterly* transparency guidelines encourage research replicability in the IS field. Nevertheless, the issue of HARKing and some *p*-hacking practices can still occur. To illustrate, authors might include redundancy in study design, try various analytical models *post-hoc*, or reformulate the hypotheses based on the analysis results (see Table 1 for common HARKing and *p*-hacking practices), in order to deliver more "perfect" results.

Stated simply, to further encourage research transparency in quantitative studies, reviewers and editors can take the initiative of changing their standard of acceptance to allow imperfect but theoretically and practically interesting results, as long as there is proper disclosure of procedures the researchers followed.

5.1.1 Detecting *p*-Hacking

During the review process, editors and reviewers can also look for signs of *p*-hacking, or require authors to conduct an exact replication (Simmons et al., 2011). Editors and reviewers can also request authors to analyze the data with different methods (or adding different variables) for both robustness checking and for potentially detecting *p*-hacking practices. From the readers' and reviewers' perspectives, it is almost impossible to be certain whether a quantitative paper is *p*-hacked or not, unless journals start requiring disclosure of all sample size rules, measures, and manipulations upon submission of articles (Simmons, Nelson, & Simonsohn, 2013). The *MIS Quarterly* guidelines on transparency material promote such disclosure to reduce *p*-hacking.

5.1.2 Replications

Encouraging replications of quantitative studies would be one subtle but powerful way to put authors on notice that their study could be refuted by future authors. However, a recent literature review by Brendel et al. (2023) reveals a paucity of replication papers in top journals in the IS field. While our top journals do not rule out replication studies, they do not encourage them. In fact, one of the authors of this paper submitted an attempted replication of a previous study using two separate, larger samples, and using structural equation modeling rather than simple t-tests, and four top journals rejected the paper because it was a replication. The previous study (from a top journal) found results, but ours did not, which could have provided important cautions to other authors embarking on the same topic. This was disappointing, as replication is essential for science (Lewandowsky & Oberauer, 2020) because the understanding or representation of a theory can be updated based on results found in replications (Brendel et al., 2023). Unsuccessful replications of studies could provide some evidence of *p*-hacking (Harris et al., 2014) or hidden boundary conditions in the original study.

5.2 Recommendations for Authors

Research transparency involves additional effort on the author's side and hypothesis-testing authors may consider avoiding HARKing and *p*-hacking as a burden or barrier for publication. However, transparency facilitates high-quality research (Burton-Jones et al., 2021). Authors who follow transparency guidelines can promote trustworthiness in research and increase the impact of their research (Burton-Jones et al., 2021), by reducing questionable research practices. Some suggestions for authors are discussed below.

5.2.1 Transparent Reporting

We agree that some of the common *p*-hacking practices we summarized in Table 1 are not unreasonable (such as measuring additional variables and *post-hoc* data treatment decisions), as long as they are reported clearly in the paper. For instance, if multiple techniques are applied to analyze the data, the results can be reported with a comparison of any divergent or unexpected results. Authors can also consider adding a section in manuscripts to discuss emergent findings that were not expected in the original study plan. We suggest researchers refer to the transparency guidelines by *MIS Quarterly* (Burton-Jones et al., 2021) to determine the amount of content to be reported and find the appropriate balance. In addition to the "transparency" principles in reporting the study design and result, researchers can also conduct blind data analysis (Nuzzo, 2015), so that they will be less likely to commit *p*-hacking since they do not know how close they are to the desired results.

5.2.2 Self-Check and Study Plan

We recommend that hypothesis-testing scholars use Table 1 in this paper to self-check whether they are involved in any problematic behaviors. Some researchers lacking proper training might not be aware of the problematic nature of HARKing and *p*-hacking. Authors should make explicit study plans, including hypotheses to test, data to be collected, and analysis methods and procedures to be used, and results to be reported (Simmons et al., 2011). Mertens and Recker (2020) also suggest a similar set of guidelines for hypothetico-deductive IS researchers. For making a study plan, authors can also identify ambiguous and unknown areas that need to be investigated or explored. By doing so, researchers can increase flexibility in their study plan for exploring new or less-known phenomena. In addition to preregistering hypotheses, protocols, and instruments, authors should specify hypotheses for competing theories or compare data to naive models, focus on effect size rather than statistical significance, and distinguish between *a-priori* expectations and *post-hoc* inferences (Mertens & Recker, 2020).

5.2.3 Highlighting Effect Size

Another suggestion we would offer is to focus more attention on effect sizes and somewhat less on significance in hypothesis-testing studies. In reporting analysis results, effect sizes and their corresponding confidence intervals have been considered a better measurement than significance. As defined by Kelly and Preacher (2012, p. 137), effect size is "a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest." When some statistical differences can be observed, it is critical to calculate and present effect sizes to allow the interpretation of the magnitude of the differences found. In particular, it is important to discuss whether the differences are meaningful or practically important. Authors should consider providing visualizations of the data and the results to illustrate the effect sizes and corresponding confidence intervals, to enhance the interpretability of their results.

Reporting only statistical significance can also obscure any issues of time-related factors. Researchers often encounter time-sensitivity of their results and should alert the reader if that is the case. For instance, it is possible to obscure results of a study by failing to provide screen shots, making it harder for the reader to understand fully the conditions and manipulations in a study. A researcher investigating stress under certain conditions using a 1980s text-based interface might have trouble replicating the results in a more modern graphical interface. Another example might be to try and publish data that was collected decades ago that focused on gender issues, which in today's world, could, in some circumstances, have become moot or otherwise completely obsolete.

5.2.4 Note for Not "*p*-Hacked" and Preregistered Study

To underscore the value of the quality of papers that are not *p*-hacked, Simmons et al. (2013) advocate hypothesis-testing researchers to note their papers as not "*p*-hacked" by including the following words in their methods sections: "We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study." (p. 775).

Another way to address the HARKing and *p*-hacking issue is to preregister a quantitative study (Harris et al., 2014), before data collection and analysis (Wicherts et al., 2016). The preregistration report should include a specific, precise, and exhaustive design of the study in advance, including the research hypotheses, data collection plans, specific analyses, and what will be reported in the paper. If the authors need to conduct additional analyses to address any issues (e.g., outliers or missing data) or to further explore intriguing patterns in the data, they should report these *post-hoc* analyses transparently (Aguinis, Cascio, & Ramani, 2017; Simmons et al., 2011). In the IS discipline, Bogert, Schecter, and Watson (2021) and Mertens and Recker (2020) also suggest preregistration as a way to mitigate authors' biases toward reporting significant findings. Preregistration has not yet been adopted as a common practice for the IS field; nevertheless, preregistered studies can still be prone to *p*-hacking if authors leave too much leeway in the analysis plan (Bogert et al., 2021).

Preregistration and reporting practices have been adopted in many fields. One of the examples is that the 2017 Annual Conference of Journal of Accounting Research implemented a two-stage editorial process: the authors first submit study proposals with design details, and then report the study results, as approved with the study proposal. Any modifications to the proposal are thus clearly noted and distinguishable through this approach. The proposals and the final paper then are presented side-by-side. The benefit of this two-stage, registered report approach is to encourage innovative research design. Without the concerns of quantitative papers being rejected due to "insignificant results," the authors with accepted proposals can undertake the work of data collection and analysis without being concerned about the outcome, since its design, methods and analysis have already been reviewed and approved. More and more journals in life sciences, biology, and psychology have joined the project of Registered Reports[4] to follow a two-stage review process.

5.3 Recommendations for the Information Systems Field

Research articles published in the IS field are positioned as either confirmatory (with hypothesis testing) or exploratory. However, even for confirmatory studies, some findings may emerge as unforeseen but have notable implications. To maximize the intellectual contributions to the field, the emergent findings, in addition to hypothesis-testing results, should be reported and elaborated in the manuscript. Although

[4] https://royalsocietypublishing.org/rsos/registered-reports

emergent findings may appear to be exploratory and do not “fit” in a confirmatory study, reporting such results can potentially offer new insights into an understudied phenomenon or uncover a unique context/situation that challenges known theories.

More recently, scholars also suggest that *post-hoc* analysis can still be used as scientific data in some cases (Hollenbeck & Wright, 2017) and discuss the distinction between *Sharking* (Secretly HARKing) and *Tharking* (Transparently HARKing). While Sharking is considered questionable, the practice of Tharking is not; it supports transparent research and reporting. Tharking is the process of presenting new hypotheses derived from *post-hoc* results; these *post-hoc* hypotheses and testing results should be reported in a separate section from *a-priori* hypotheses in a way that it is transparent to readers (Aguinis et al., 2017). Tharking can replace the practice of HARKing and enhance methodological transparency (Aguinis, Ramani, & Alabduljader, 2018). As argued by Hollenbeck and Wright (2017), “Tharking should be part of every published empirical study for any authors who do not have perfect ability to omnisciently predict the future” (p. 9). While researchers who are warned about HARKing might believe that they will need to follow only the original research plan for hypothesis testing for confirmatory studies, Tharking can reveal unexpected patterns or relationships that emerge from the data, suggesting new, intriguing directions for future research. The main purpose of Tharking should be to reconsider interesting questions given the analysis results available, aiding the process of discovery. Tharking should be used with caution and to focus attention on assessing alternative explanations for null results (Vancouver, 2018).

As compared to preregistration that focuses on documenting the history of a study, Tharking affords transparency that allows readers to identify differences between planned versus alternative hypotheses, methods, and analyses; it therefore also improves the credibility of research findings (Rubin, 2020). While preregistration requires effort before a study takes place and cannot be amended afterward, Tharking allows the flexibility to explore and report the possible alternative relationships between variables.

Tharking can also address the issue of review panel-driven HARKing. During the peer-review process, editors and reviewers may provide suggestions on how to present the research in a more cohesive and “interesting” way to attract the attention of readers, to underscore important research findings, or to frame the research in a novel way. While the published outcome might deviate from the researchers’ original plan, *post-hoc* hypotheses can still offer valuable insights derived from the study result. When the research results and the testing of *post-hoc* hypotheses are reported with transparency, it can facilitate reanalysis and replication of the study and promote trustworthiness in research, which is aligned with the purpose of research transparency (Burton-Jones et al., 2021). The IS field currently does not have a norm of reporting both *a-priori* and *post-hoc* hypotheses, and review panel-driven HARKing is not typically reported in published research articles, but these transparent measures are likely to result in findings that are more honest, complete, and dependable.

5.4 Future Research

To understand the extent and nature of HARKing and *p*-hakcing, the IS field would particularly benefit from meta-analyses that include unpublished studies and those from a variety of outlets rather than just those from top journals. Furthermore, theory building studies are not immune to practices that could distort findings. If informants provide “off the record” comments that provide substantial illumination,[5] for instance, the researcher is thrown into an ethical dilemma of whether and how to use that information. Also, researchers need to write a narrative that is interesting and cohesive, and could omit important details that, although providing a more complete story, might provide too much complexity or length to a paper, dampening its likelihood of acceptance. Finally, a researcher could pay closer attention to facts that they understand better and omit critical information that is outside of their realm. These are all important elements of distortion that are outside of this paper’s context. For brevity and sharper focus, as well as consistency with the research traditions into which we ourselves are embedded, we only consider the positivist, hypothesis-testing tradition herein. A future paper should concentrate on questionable practices in the interpretivist traditions (including case studies, action research, literature reviews, and design science, among others). The practices and implications are likely to be quite complex and nuanced, yet they are important for future researchers to pursue. We expect such a paper like this one, written by experts in a qualitative research tradition, to be very helpful to theory building researchers.

[5] One of the editors of this paper suggested we include this issue and disclosed that an interviewee concealed his true feelings about a system, and when off the record, admitted he lied during the formal interview.

6 Conclusion

Although the issues of HARKing and *p*-hacking are particularly prominent for confirmatory studies and studies that are exploratory in nature without pre-specified plans for data collection and analysis, the details of the studies should also be transparently reported, including any unexpected results obtained with models that were not originally considered. The results of such studies then can be interpreted with the appropriate level of caution (Bruns & Ioannidis, 2016), and could increase the reproducibility and replicability of the findings (Aguinis et al., 2017). Researchers need to make numerous choices during a research project. These choices need to be transparent in the final reporting of the studies (Baron, Zaltman, & Olson, 2017; Harvey, 2017; Hollenbeck & Wright, 2017).

We recommend that authors begin the practice of being more transparent about their model development and analytical procedures when submitting papers to conferences and journals submissions to reduce the inflated results that can accrue due to HARKing and *p*-hacking, and to provide more dependable, consistent, and meaningful accounts of their work. The entire field will benefit from this transparency and honesty.

References

Aguinis, H., Cascio, W. F., & Ramani, R. S. (2017). Science's reproducibility and replicability crisis: International business is not immune, *Journal of International Business Studies*, 48(6), 653–663.

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals*, 12(1), 83–110.

Awati, R. (2022). Data Dredging (data fishing), TechTarget Data Management, (https://www.techtarget.com/searchdatamanagement/definition/data-dredging).

Baron, A. S., Zaltman, G., & Olson, J. (2017). Barriers to advancing the science and practice of marketing, *Journal of Marketing Management*, 33(11-12), 893–908.

Bettis, R. A. 2012. The search for asterisks: Compromised statistical tests and flawed theories, *Strategic Management Journal*, 33(1), 108–113.

Bogert, E., Schecter, A. & Watson, R.T., (2021). Preregistration of information systems research, *Communications of the Association for Information Systems*, 49(1), 110–125.

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). HARKing's threat to organizational research: Evidence from primary and meta-analytic sources, *Personnel Psychology*, 69(3), 709–750.

Boss, S., Galletta, D., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors, *MIS Quarterly*, 39(4), 837–864.

Brendel, A. B., Diederich, S., & Niederman, F. (2021). An immodest proposal–going "All in" on replication research in information systems. *European Journal of Information Systems*, 32(2), 277-286.

Bruns, S. B., & Ioannidis, J. P. A. (2016). P-curve and p-Hacking in observational research, *PLoS ONE*, 11(2), e0149144.

Burton-Jones, A., Boh, W. F., Oborn, E., & Padmanabhan, B. (2021). Editor's Comments: Advancing research transparency at MIS Quarterly: A pluralistic approach, *Management Information Systems Quarterly*, 45(2), iii–xviii.

Cuddy, A. J. C., Schultz, S. J., & Fosse, N. E. (2018). P-curving a more comprehensive body of research on postural feedback reveals clear evidential value for power-posing effects: Reply to Simmons and Simonsohn (2017), *Psychological Science*, 29(4), 656–666.

Dalton, D. R., Aguinis, H., Dalton, C. M., Bosco, F. A., & Pierce, C. A. (2012). Revisiting the file drawer problem in meta-analysis: An empirical assessment of published and nonpublished correlation matrices, *Personnel Psychology*, 65(2), 221–249.

DuToit, S. H., Steyn, A. G. W., & Stumpf, R. H. (2012). Graphical exploratory data analysis. Springer Science & Business Media.

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

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Sage publications.

Harris, C. R., Pashler, H., & Mickes, L. (2014). Elastic analysis procedures: An incurable (but preventable) problem in the fertility effect literature: comment on Gildersleeve, Haselton, and Fales (2014), *Psychological Bulletin*, 140(5), 1260–1264.

Harvey, C. R. (2017). Presidential address: The scientific outlook in financial economics, *Journal of Finance*, 72(4), 1399–1440.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-Hacking in science, *PLoS Biology*, 13(3), e1002106.

Hollenbeck, J. R., & Wright, P. M. (2017). Harking, Sharking, and Tharking, *Journal of Management*, 43(1), 5–18.

Honig, B., Lampel, J., Siegel, D., & Drnevich, P. (2014). Ethics in the production and dissemination of management research: Institutional failure or individual fallibility? *Journal of Management Studies*, 51(1), 118–142.

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

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17(2), 137–152.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known, *Personality and Social Psychology Review*, 2(3), 196–217.

Kwee, R. M., Almaghrabi, M. T., & Kwee, T. C. (2023). Integrity in cardiovascular imaging research. Clinical Imaging, 96, 31–33.

Lewandowsky, S., & Oberauer, K. (2020). Low replicability can support robust and efficient science. *Nature Communications*, 11(1), 358.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, 73(sup1), 235–245.

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.

Meyer, R. (2013). Paul Green, Journal of Marketing Research, and the challenges facing marketing, *Journal of Marketing Research*, 50(1), 1–2.

Motulsky, H. J. (2014). Common misconceptions about data analysis and statistics, *British Journal of Pharmacology*, 172(8), 2126–2132.

Nuzzo, R. (2014). Scientific method: Statistical errors, *Nature*, 506, 150–152.

Nuzzo, R. (2015). How scientists fool themselves - and how they can stop, *Nature*, 526, 182–185.

Rubin, M. (2020). Does preregistration improve the credibility of research findings? *The Quantitative Methods for Psychology*, 16(4), 376–390.

Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., Tamman, A. J. F., & Puhlmann, L. M. C. (2015). Romance, risk, and Replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*, 144(6), e142–e158.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant, *Psychological Science*, 22(11), 1359–1366.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2013). Life after P-Hacking, *Meeting of the Society for Personality and Social Psychology*, New Orleans, LA, 17-19 January 2013.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer, *Journal of Experimental Psychology: General*, 143(2), 534–547.

Starbuck, W. H. (2016). 60th anniversary essay: How journals could improve research practices in social science, *Administrative Science Quarterly*, 61(2), 165–183.

Sweedler, J.V. (2015). Striving for reproducible science, *Analytical Chemistry*, 87(23), 11603–11604.

Vancouver, J. B. (2018). In defense of HARKing, *Industrial and Organizational Psychology* 11(1), 73–80.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., van Aert, R. C. M., & van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-Hacking, *Frontiers in Psychology*, 7(NOV), 1832.

Appendix A: Analysis of Articles in Top IS Journals

Table A1. Summaries of Supported Hypotheses in Top Information Systems Journals

Journal Name	Year	Percentage of Supported Hypotheses	Average
MIS Quarterly	2013	83.01%	85.84%
	2014	79.59%	
	2015	87.78%	
	2016	91.25%	
	2017	87.59%	
Information Systems Research	2013	89.07%	88.75%
	2014	89.92%	
	2015	90.14%	
	2016	91.96%	
	2017	82.64%	
Journal of Management Information Systems	2013	78.95%	79.90%
	2014	78.41%	
	2015	81.76%	
	2016	78.62%	
	2017	81.76%	
European Journal of Information Systems	2013	82.26%	87.15%
	2014	91.84%	
	2015	84.91%	
	2016	91.04%	
	2017	85.71%	
Information Systems Journal	2013	88.46%	80.42%
	2014	84.29%	
	2015	78.26%	
	2016	70.00%	
	2017	81.08%	
Journal of the Association for Information Systems	2013	93.10%	85.75%
	2014	78.95%	
	2015	94.94%	
	2016	88.75%	
	2017	73.03%	

Appendix B: Original Question Items in this Study

Questionable Research Practices (John et al., 2021)

Question Number	Item	Mean	Standard Deviation
QRP01	I failed to report all of a study's dependent measures that I collected.	3.00	1.67
QRP02	I decided whether to collect more data only after performing preliminary analysis and determining whether the results were significant.	2.30	1.53
QRP03	I did not include control variables that were collected.	2.59	1.59
QRP04	I stopped collecting data earlier than planned because I found the result that I had been looking for.	1.35	1.00
QRP05	I "rounded off" a *p*-value.	1.38	1.16
QRP06	In some papers, I inappropriately assumed a one-tailed test, in order to achieve significant results.	1.37	0.99
QRP07	I selectively reported constructs that "worked" in the overall model.	2.61	1.51
QRP08	I decided whether to exclude data after looking at the impact of doing so on the results.	1.95	1.44
QRP09	I reported an unexpected finding as having been predicted from the start.	2.28	1.56
QRP10	I claimed that results are unaffected by demographic variables (e.g., gender) when I am unsure (or knows that they do).	1.34	1.03
QRP11	I falsified data.	1.15	0.78
QRP12	I altered parts of the model to increase the overall results.	2.53	1.55
QRP13	I did entirely exploratory analysis on the data in order to arrive at a model, and then wrote up the study as if this model was finalized prior to collecting data.	2.25	1.64
QRP14	I clearly included all preliminary models that were tested for this study.	3.48	2.06
QRP15	I clearly identified all analytical procedures that I performed in this study in the paper.	5.11	1.91
QRP16	I included mediations or moderations in the study that I discovered during the data analysis as if they were planned prior to data collection.	2.80	1.67
QRP17	I clearly and completely reported how some data points are excluded from the analysis.	5.84	1.75
QRP18	I reported some *post-hoc* analysis results as if they were planned prior to data collection.	2.58	1.69
QRP19	I have been asked by a review panel to represent surprising results as hypothesized results by changing directionality of one or more hypotheses.	2.14	1.70
QRP20	In #19 above, I have followed through as they asked.	2.47	2.20
QRP21	I have been asked by a review panel to represent surprising results as hypothesized results by adding one or more hypotheses that were not in the original set.	2.29	1.73
QRP22	In #21 above, I have followed through as they asked.	2.62	2.21
QRP23	As a member of a review panel, I have asked authors to represent surprising results as hypothesized results by changing directionality of one or more hypotheses.	1.34	0.92
QRP24	As a member of a review panel, I have asked authors to represent surprising results as hypothesized results by adding one or more hypotheses that were not in the original set.	1.46	1.03

Appraisal (Boss et al., 2015)

Question Number	Item	Mean	Standard Deviation
Fear Appraisal			
FEAR01	I am worried about the extent of *p*-hacking in our discipline.	5.08	1.60
FEAR02	I am anxious about the extent of *p*-hacking in our discipline.	4.67	1.73
FEAR03	I am dismayed at the extent of *p*-hacking in our discipline.	4.37	1.83
Severity Appraisal			
SEV01	I believe *p*-hacking to be a severe problem in our discipline.	4.57	1.70
SEV02	I think *p*-hacking is a serious problem in our discipline.	4.61	1.77
SEV03	I think *p*-hacking is a significant issue in our discipline.	4.65	1.74
SEV04	I believe that our discipline is at risk due to *p*-hacking practices.	4.17	1.79
SEV05	I believe that our discipline is likely impacted by *p*-hacking.	4.95	1.62

SEV06	I think it is likely that *p*-hacking will continue in our discipline.	5.57	1.22
Response Cost Appraisal			
RCST01	I find that engaging in *p*-hacking is more beneficial to engage in than it is risky.	3.26	1.99
RCST02	I think that avoiding *p*-hacking practices simply takes too much effort to report everything in a paper.	3.60	1.95
RCST03	I feel that the effort to report all analyses and models would cause too many problems in a paper.	4.52	2.08
RCST04	I would feel silly for reporting all of my analyses and modeling efforts in a paper.	4.36	1.97

Awareness of HARKing and *p*-hacking in IS

Question Number	Item	Mean	Standard Deviation
AW01	I have specific awareness of cases in which others in our discipline have engaged in *p*-hacking.	3.99	2.19
AW02	I estimate that *p*-hacking is engaged in by the following percentage of our discipline, on a regular basis. (0-100)	40.38	30.54
AW03	Over my entire career, I have engaged in *p*-hacking for what percentage of my projects? (0-100)	13.02	22.54
AW04	Over the last five years, I have engaged in *p*-hacking for what percentage of my projects? (0-100)	9.92	19.77

Appendix C: Final Items Included in the Analysis

Review panel-driven HARKing

QRP19: I have been asked by a review panel to represent surprising results as hypothesized results by changing directionality of one or more hypotheses.

QRP20: In #19 above, I have followed through as they asked.

QRP21: I have been asked by a review panel to represent surprising results as hypothesized results by adding one or more hypotheses that were not in the original set.

QRP22: In #21 above, I have followed through as they asked.

Author-driven HARKing

QRP09: I reported an unexpected finding as having been predicted from the start.

QRP16: I included mediations or moderations in the study that I discovered during the data analysis as if they were planned prior to data collection.

QRP18: I reported some *post-hoc* analysis results as if they were planned prior to data collection.

Statistical hacking

QRP04: I stopped collecting data earlier than planned because I found the result that I had been looking for.

QRP05: I truncated a p value instead of rounding appropriately (for example p=.056 was truncated to p=.05).

QRP06: In some papers, I inappropriately assumed a one-tailed test, in order to achieve significant results.

QRP10: I claimed that results are unaffected by taking into account demographic variables when I am unsure or know that they really are affected by those variables.

QRP11: I falsified data.

Selective reporting (Reverse)

QRP14: I clearly included all preliminary models that were tested for this study

QRP15: I clearly identified all analytical procedures that I performed in this study in the paper

QRP17: I clearly and completely reported how some data points are excluded from the analysis.

Fear and severity

FEAR01: I am worried about the extent of *p*-hacking in our discipline.

FEAR02: I am anxious about the extent of *p*-hacking in our discipline.

FEAR03: I am dismayed at the extent of *p*-hacking in our discipline.

SEV01: I believe *p*-hacking to be a severe problem in our discipline.

SEV02: I think *p*-hacking is a serious problem in our discipline.

SEV03: I think *p*-hacking is a significant issue in our discipline.

SEV04: I believe that our discipline is at risk due to *p*-hacking practices.

SEV05: I believe that our discipline is likely impacted by *p*-hacking.

SEV06: I think it is likely that *p*-hacking will continue in our discipline.

Response cost

RCST01: I find that engaging in *p*-hacking is more beneficial to engage in than it is risky. (Reverse)

RCST02: I think that avoiding *p*-hacking practices simply takes too much effort to report everything in a paper.

RCST03: I feel that the effort to report all analyses and models would cause too many problems in a paper.

RCST04: I would feel silly for reporting all of my analyses and modeling efforts in a paper.

Appendix D: Measurement Assessment

As shown in Table D.1, Cronbach's α values of the constructs were all above 0.8 and demonstrated satisfactory internal consistency. Composite reliability values were also above the suggested level of 0.8. The average variance extracted (AVE) values ranged from 0.59 (Selective reporting) to 0.84 (Review panel-driven HARKing). We further examine whether the square root of AVE was larger than inter-construct correlations and whether measurement items load higher on the construct they intended to measure than on other constructs in the research model (Hair, Hult, Ringle, & Sarstedt, 2016). Both of these criteria were met, indicating appropriate discriminant validity.

Table D.1. Descriptive Statistics, AVE, Composite Reliability, and Alpha

		Mean	Std. Deviation	AVE	Composite Reliability	Cronbach's Alpha
Dependent variable	Review panel-driven HARKing	2.38	1.80	0.84	0.95	0.94
	Author-driven HARKing	2.55	1.44	0.77	0.91	0.85
	Statistical hacking	1.32	0.85	0.73	0.93	0.91
	Selective reporting	4.81	1.47	0.59	0.81	0.65
Independent variable	Fear and severity	4.74	1.44	0.74	0.96	0.96
	Response cost	3.93	1.64	0.68	0.89	0.84

Table D.2 shows the correlations between research constructs and Table D.3 shows the cross-loading of the items. Both these two criteria were met, indicating appropriate discriminant validity.

Table D.2. Latent Variable Correlations

	1	2	3	4	5	6
1. Review panel-driven HARKing	**0.92**					
2. Author-driven HARKing	0.46	**0.88**				
3. Statistical hacking	0.41	0.46	**0.86**			
4. Selective reporting	-0.20	-0.12	-0.03	**0.77**		
5. Fear and severity	0.22	0.05	0.17	-0.14	**0.86**	
6. Response cost	0.26	0.27	0.24	-0.38	0.05	**0.82**

Note: The bold items on the diagonal items are the square root values of AVE. All correlation coefficients between the first-order constructs are significant at the $p < 0.01$ level.

Table D.3. Cross-Loading

	Review process-driven harking	Autonomous harking	Statistical hacking	Selective reporting	Fear and severity	Response cost
QRP19	**0.92**	0.42	0.41	-0.21	0.22	0.23
QRP20	**0.92**	0.41	0.33	-0.18	0.19	0.25
QRP21	**0.92**	0.42	0.40	-0.17	0.21	0.22
QRP22	**0.90**	0.43	0.36	-0.17	0.17	0.24
QRP09	0.48	**0.81**	0.45	-0.11	0.17	0.15
QRP16	0.36	**0.93**	0.39	-0.07	-0.01	0.30
QRP18	0.41	**0.89**	0.39	-0.14	0.02	0.23
QRP04	0.29	0.38	**0.82**	-0.05	0.10	0.19
QRP05	0.35	0.46	**0.81**	0.02	0.13	0.16
QRP06	0.42	0.42	**0.87**	-0.05	0.16	0.24
QRP10	0.38	0.32	**0.87**	-0.06	0.16	0.21
QRP11	0.31	0.39	**0.91**	0.02	0.16	0.22
QRP14	-0.26	-0.17	0.08	**0.79**	-0.10	-0.33
QRP15	-0.10	-0.10	-0.03	**0.83**	-0.09	-0.32
QRP17	-0.08	0.02	-0.15	**0.68**	-0.15	-0.22
FEAR01	0.13	-0.06	0.12	-0.08	**0.86**	-0.05
FEAR02	0.19	-0.04	0.21	-0.12	**0.84**	0.10
FEAR03	0.23	0.02	0.16	-0.12	**0.84**	-0.02
SEV01	0.21	0.09	0.17	-0.17	**0.95**	0.09
SEV02	0.20	0.11	0.22	-0.14	**0.94**	0.08
SEV03	0.15	0.10	0.16	-0.13	**0.95**	0.05
SEV04	0.16	0.04	0.08	-0.16	**0.82**	-0.03
SEV05	0.20	0.05	0.06	-0.06	**0.82**	-0.01
SEV06	0.20	0.04	0.04	-0.07	**0.69**	0.14
RCST01	0.29	0.20	0.26	-0.24	0.18	**0.75**

RCST02	0.17	0.21	0.24	-0.28	-0.04	**0.83**
RCST03	0.17	0.19	0.13	-0.33	0.00	**0.86**
RCST04	0.21	0.27	0.16	-0.40	0.02	**0.85**

About the Authors

Han-fen Hu is an Associate Professor in the Department of Management, Entrepreneurship, and Technology, University of Nevada, Las Vegas. Dr. Hu receives her Ph.D. at the David Eccles School of Business, University of Utah. Her research interests include electronic commerce, online consumer behavior, and organizational technology strategy. She has papers published in *MIS Quarterly; Journal of Management Information Systems; Information and Management; Journal of Business Research; International Journal of Contemporary Hospitality Management; Psychology and Marketing; Journal of the American Society for Information Science and Technology; Current Issues in Tourism; Communications of the Association of Information Systems* and other IS and Marketing Journals.

Gregory D. Moody (Ph.D., University of Pittsburgh; Ph.D., University of Oulu) is currently a Lee Professor of Information Systems and a Troesh Scholar in the Management, Entrepreneurship and Technology Department in the Lee Business School at the University of Nevada, Las Vegas and Director of the Graduate Cybersecurity program. He has published in *ISR, MISQ, JMIS, JAIS, EJIS, ISJ* and other journals. His interests include IS security and privacy, e-business (electronic markets, trust) and human–computer interaction (Web site browsing, entertainment). He is currently a senior editor for ISJ and THCI, the previous president of SIGHCI, and the Managing Editor for THCI.

Dennis F. Galletta is Thomas H. O'Brien Chair of Business Administration at the University of Pittsburgh. He is an AIS lifetime achievement awardee (LEO), an AIS Fellow, and AIS distinguished Member Cum Laude. He serves as PhD Program Director and DBA Program Academic Director for the Katz Business School. He has published 59 journal articles, with 20 in Financial Times outlets, 58 conference papers and five books. He served as a Senior Editor at MIS Quarterly, is an editorial board member at Journal of MIS and Preeminent Editor at Data Base. He won a "Developmental Associate Editor Award" at MISQ in 2006. He won a Provost's Mentorship Award in 2016. He was program co-chair for ICIS 2005 and AMCIS 2003, chaired the first AMCIS in 1995, and co-chaired ICIS 2011. In the past, he served as AIS President, ICIS Treasurer, AIS Council Member, and Editor-in-Chief of AISWorld. He was founding co-Editor in Chief of AIS Transactions on HCI from 2008 to 2018, and established the concept and framework for Special Interest Groups in AIS in 2000.