Designing for Preregistration in Practice: Multiple Norms and Purposes

Xiaoying Pu

Computer Science and Engineering,

University of Michigan

Ann Arbor, MI 48109 USA

xpu@umich.edu

Licheng Zhu

School of Information, University of Michigan

Ann Arbor, MI 48109 USA

zlicheng@umich.edu

Matthew Kay

School of Information, University of Michigan

Ann Arbor, MI 48109 USA

mjskay@umich.edu

Frederick Conrad

Institute for Social Research, University of Michigan

Ann Arbor, MI 48109 USA

fconrad@umich.edu

Abstract

The replication crisis—a failure to replicate foundational studies—has sparked a conversation in psychology, HCI, and beyond about scientific reliability. To address the crisis, researchers increasingly adopt preregistration: the practice of documenting research plans before conducting a study. Done properly, preregistration reduces bias from taking exploratory findings as confirmatory. It is crucial to treat preregistration, often an online form/template, as a user-centered design problem to ensure preregistration achieves its intended goal. To understand preregistration in practice, we conducted 14 semi-structured interviews with preregistration users (researchers) who ranged in seniority and experience. We found that norms for using preregistration are uncertain, and identified four purposes people have for using preregistration. We describe how the different purposes could explain conflicting designs of existing preregistration templates, and how they may even undermine the original purpose of preregistration. We suggest opportunities to explicitly resolve these conflicts to support the different purposes of preregistration.

Designing for Preregistration in Practice: Multiple Norms and Purposes

Introduction

A replication crisis shook the psychological science, where many foundational studies, such as behavioral priming and ego depletion, have failed to be independently replicated Carter and McCullough (2014); Doyen, Klein, Pichon, and Cleeremans (2012). Among others, preregistration has been proposed as a *tool* for addressing the replication crisis and for improving the scientific process Nosek, Ebersole, DeHaven, and Mellor (2018). One definition of preregistration is “the action of confirming an unalterable version of one’s research plan prior to collecting data”, thus distinguishing any preregistered, confirmatory findings from exploratory findings Kidwell et al. (2016); Nosek et al. (2018). Preregistration constrains the researcher’s flexibility in analysis and experimental design—the very flexibility that arguably contributed to the replication crisis Ioannidis (2005); Wagenmakers, Wetzels, Borsboom, van der Maas, and Kievit (2012).

We approach preregistration as a *user-centered design problem*: to investigate how and whether it achieves its proposed goals in practice, and how the process might be improved. The idealization of preregistration described in the literature is a process by which authors can clearly distinguish between exploratory and confirmatory analyses, and peer reviewers and readers can verify which analyses are which Nosek et al. (2018). In practice, several high-profile cases have emerged in psychology where prominent researchers have investigated published, preregistered papers and found that the papers did not follow their preregistrations or did not report significant deviations from protocol Simons (2015); Wagenmakers and Gronau (2017, 2018). This raises the question: is the current implementation of preregistration in practice—the design of preregistration forms, how preregistration integrates into research practice and peer review, etc.—effectively supporting its intended goals?

We offer a broad overview of how preregistration is actually used based on a qualitative interview study of 14 preregistration users—primarily academic psychologists, a

field that is at the forefront of preregistration adoption. We aim to understand current practice and the norms and purposes for preregistration that actual users have. This allows us to contrast these purposes with the original goal of preregistration—distinguishing exploratory and confirmatory findings—to understand how the purpose of preregistration may have evolved, and how design might better support it. We find:

- *Norms* around how to use preregistration are uncertain and vary amongst its users.
- Users have a variety of *purposes* for adopting preregistration, some of which deviate from the original purpose of distinguishing exploratory and confirmatory findings.
- The same preregistration document has different use cases for authors and reviewers/editors, but these different use cases are not explicitly designed for.
- There exist conceptual barriers to adopting preregistration, including users from different disciplines feeling that preregistration is not for them.

We discuss how the different purposes ascribed to preregistration may result in conflicting designs of preregistration formats, and how these conflicting formats may undermine the overarching goal of preregistration: improving the reliability of scientific findings. We also propose how the mismatch between a *designer's intention* for preregistration and *users' purposes* for preregistration might be resolved explicitly through design.

Background

Giofre et al. show that as of 2015, only 2% of the papers published in the journal *Psychological Science* were preregistered and *Journal of Experimental Psychology: General* was at 4% Giofr, Cumming, Fresc, Boedker, and Tressoldi (2017). While these adoption rates are low, the number of preregistrations on file are quickly increasing Kupferschmidt (2018); Nosek and Lindsay (2018). In the OSF registry alone, the number of registrations, though not all in psychology, grew from 38 in 2012 to 18,000 in 2018 Kupferschmidt (2018); Nosek and Lindsay (2018). A new journal, *Advances in Methods and Practices in*

Psychological Science, has been created specifically to be “a forum for demonstrating best practices for open science, including preregistration” in Methods and in Psychological Science (2018). With the low but rapidly increasing adoption of preregistration, it is crucial to understand how preregistration is unfolding in practice.

Preregistration process

We illustrate the preregistration process in Figure 1. The user preregisters their study design before data collection and analysis. After completing the preregistration, the user executes the project and produces a final paper describing their findings. Both the paper and the preregistration are made available to the reviewers, editors and other readers, so that readers can *check* the protocol was followed. If any deviations from the preregistration occur, these deviations should be clearly marked in the paper so that readers can make informed judgments about the impact of those deviations on the results.

Preregistration formats and registries. While not the only approach, preregistrations are typically made using web forms submitted to online registries, which support archiving and anonymous links for peer review. Two common registries are the Open Science Foundation (OSF; osf.io) and AsPredicted (aspredicted.org). These templates vary in length and content, see Figure 2. Besides the general-purpose sites, different research fields have their own preregistration sites, such as the The American Economic Association’s registry for randomized controlled trials (AEA RCT Registry) (www.socialscienceregistry.org) for economics and ClinicalTrials.gov for medical research.

Journals also increasingly endorse an approach called *registered reports* Center for Open Science (2018). As shown in Figure 1, in contrast to typical pre-registrations, a registered report is peer-reviewed *before* the data is collected and can earn an in-principle acceptance regardless of the study outcome Chambers (2013), so long as the protocol is followed. This allows a study to be accepted on its merits regardless of the outcome, which may help reduce publication bias.

Existing scientific norms

Preregistration is a relatively new open science practice Nosek and Lindsay (2018), so it may be helpful to connect it to existing norms for scientific research. There is a body of literature that describes the *norms of science*—the normative research standards among scientists Anderson, Martinson, and De Vries (2007). One such formulation is Merton’s norms, consisting of four aspects Anderson (2000); Merton (1973) which easily translate into the current open science practices:

- *universalism*: the validity of research work should be the same everywhere, independent of the researcher;
- *communality*: scientific findings are shared with everyone in the open;
- *disinterestedness*: the researcher’s practices are not motivated by personal interests but by the pursuit of truth only; and
- *organized skepticism*: research findings should be critically examined.

By stating in advance the study plans, preregistration can be said to curtail motivated reasoning (promoting disinterestedness).¹ Preregistration, ideally conducted, should also aid organized skepticism by allowing independent researchers to assess how closely a preregistration was followed and to understand clearly what findings are confirmatory or exploratory—the purpose of preregistration according to Nosek and Lindsay Nosek and Lindsay (2018). It is worth asking, then, what the state of preregistration quality is in the wild.

Assessing preregistration quality

Several prominent psychology methodologists have written blog posts commenting on preregistered publications. These critiques are sometimes called *post-publication peer review*, also shown in Figure 1. For example, Simons critiqued a paper in which the original

¹Merton’s norms also coincide with other open science practices, such as independent replication of findings (universalism, organized skepticism), open-access data and material (communality) and *post-publication peer review* (organized skepticism).

authors did not recruit as many participants as specified in the preregistration and the participants did not undergo the specified length of training Simons (2015). Tullett critiqued the experiment in a different paper, finding it did not reach the sample size for 80% power determined in the preregistration Tullett (2018); as a result, the paper used an under-powered analysis to accept a null hypothesis.

Besides more direct violation of preregistered protocols, researchers/readers noted non-ideal reporting of confirmatory *v.s.* exploratory results (see Table 1), given Nosek et al.'s definition of the purpose of preregistration Nosek et al. (2018). By comparing against its preregistration, Wagenmakers and Gronau found that the confirmatory section of a paper was “infiltrated” with exploratory analysis Wagenmakers and Gronau (2018). The original authors of that paper also based strong conclusions on their exploratory analysis. Wagenmakers and Gronau also investigated another paper, finding that the authors only selected significant results to report in the main text and left preregistered but not significant results in the supplement, which Wagenmakers and Gronau argued leaves the reader a “skewed impression of the evidence” Wagenmakers and Gronau (2017).

These blog posts show that not all preregistered studies conform to their preregistrations or are inconsistent in reporting deviations.² Ideally, these problems should have been caught during peer review, where the preregistration should have been available to reviewers and should have been evaluated against the final paper. Part of the purpose of this paper is to better understand why the current design of preregistration can fail in its intended purpose, and what design might do to fix it.

Interview study

Interview goal and participants

To understand the use of preregistration in research, writing, and the peer-review process, we conducted fourteen ($N = 14$) semi-structured interviews. Participants were

²It is important to stress that deviations from a preregistration are considered acceptable by many advocates of the practice, so long as such deviations are clearly reported.

recruited through word-of-mouth and at open science workshops in psychology. All participants were Ph.D. students or faculty members and had self-reported experience or interest in preregistration. The participants were from eight different universities in the United States and the majority of participants were in psychology. We had an even distribution in job title/seniority and gender. Table 2 shows more details about participant demographics.

We primarily recruited in psychology because researchers in that field have increasingly adopted, advocated for, and extensively discussed Nosek et al. (2018); Wagenmakers et al. (2012) preregistration practice. Thus from psychologists, we should expect to learn more about how preregistration use unfolds in practice.

Interview protocols

The interview sessions lasted 43 minutes on average ($min = 23, max = 57, sd = 11.2$). We conducted the interviews in person and through video calls. Audio recordings of the interviews were transcribed to text for analysis. The interview questions targeted different roles in which a participant might encounter preregistration; e.g., as an author of a preregistration or as an editor/reviewer of a preregistered manuscript. We separately asked about experiences with preregistration under each role (some participants had experience with preregistration under multiple roles). We also asked more general questions about preregistration, such as motivations to adopt preregistration and opinions about preregistration formats.

Our interview questions evolved over time as we encountered participants with different levels of experience with preregistration. For instance, we quickly added questions specific to journal reviewers and editors to our protocol. With the permission of P12 and P13, we were also able to walk through the version history of their preregistration on Google Docs and discuss how they had developed it in detail. Throughout the interviews, we asked follow up questions for details and clarification.

Coding protocol/Data analysis

Two independent coders developed codes from the ground up and then summarized themes with affinity diagrams. For each of the first two interviews, the coders independently assigned codes using NVivo 12 and then compared the two sets of codes qualitatively. The codes were largely consistent, and any disagreements were resolved by discussion. The remaining interviews were coded once by one of the two coders separately for efficiency. After all interviews had been coded, the coders grouped codes into higher level themes as they emerged. The coders also drew affinity diagrams to uncover connections among themes.

Supplementary material practices. Our final interview protocols and codebook are available in our supplementary material. As noted above, the interviews were semi-structured, so the protocols were not followed strictly for every participant (particularly where they had less experience with a particular role or topic).

Interview findings

Based on the results of our qualitative coding, we found a variety of *norms* and *purposes* for preregistration, different aspects of its *use*, and several *barriers* to its adoption.

Uncertain norms around preregistration

If we view *norms* as a set of ideal behaviors, norms around preregistration can range from whether to conduct power analysis to which preregistration format to use. Unlike the high-level Merton's norms that researchers "subscribed strongly to" Anderson (2000), a recurring theme across interviews is that the norms around creating and using preregistration are *uncertain*.

As a specific example of this uncertainty, P6 said that the norm around which preregistration format to choose was "in the middle of it evolving". P7 wanted clearer norms around how to mark the preregistered from the non-preregistered parts of a paper.

P5 was unsure about the norms around whether to preregister power analysis (Table 1). From this uncertain norm, P5 commented that preregistration had “a lot of moving parts.” Indeed, there is a disagreement about whether or not power analysis should be included in a preregistration (P6): (we revisit this disagreement in Discussion):

There is some argument about whether or not power analyses should be included. I’m in the group that is, “You should absolutely include power analyses”. But there is a group that... And actually, it’s Uri Simonsohn’s group who doesn’t think that you need to include power analyses. [...]
Preregistration is not fully formed as to exactly what it [the norm] is.

The uncertain norms can be difficult to keep up with, or a source of confusion especially for new users—P14 described it as “things evolving too fast”. P5 pointed out that as a student who just started doing preregistration, they often heard conflicting information about how things should be done. In this case, evolving norms might have made doing preregistration more difficult than it needs to be.

We also found that there is no single set of norms about how to evaluate studies that have used preregistrations—e.g., when and how such studies should be compared against their preregistrations to ensure the study has been executed as planned, and how to report deviations. When asked the question of how to do the evaluation, P6 asked back, “What is quality?”, and stated that the *status quo* on evaluation is in the “wild, wild west”. Specifically, it is unclear as to who is doing the evaluation, if at all, raising concerns about the quality of the match between the paper and the preregistration. As P6 put it:

Some people, like at our journal, will actually go back and check and see what was pre-registered and what actually came through on the paper. But at the big journals, I’m not entirely sure they’re doing that. So people can just get these pre-registration badges, but did they follow? [It could be that] people

didn't understand pre-registration and then what they did in the paper they came out with was very different from each other.

Another issue is that the evaluation of preregistrations is not and should not be simple. For instance, P14, like P6, had concerns about giving out badges (defined in Table 1) to paper just for including a preregistration, which might be accidentally applied to poor-quality research if the preregistration was not properly evaluated.

Purposes of preregistration

Four broad purposes that participants ascribed to preregistration emerged from our interviews. For each purpose, we describe what participants perceive as *benefits* of using preregistration related to that purpose.

Purpose: delimiting flexibility. *Delimiting flexibility* subsumes the purpose of marking the distinction between exploratory (flexible) and confirmatory studies, defined as the purpose for preregistration in the literature Nosek et al. (2018); Wagenmakers et al. (2012). We use the word *delimit* specifically for its connotations of both separation (to separate exploratory from confirmatory) and containment (as some participants discussed limiting, or containing, flexibility—or *researcher degrees of freedom*).

This purpose of preregistration was endorsed by several participants. P9 thought that preregistration was “a way to manage flexibility in analyses”, or as P10 put it, “decision” points. P10 also emphasized that it was the time-sensitive decisions that should be preregistered and those only.

For example, P10 considered power analysis irrelevant to preregistration, because the rationale for why a given sample size is adequate is not time-sensitive; preregistration should only state the predetermined sample size itself. That is, just the sample size is necessary to be stated in advance in order to determine if the final analysis followed the plan (i.e., to delimit flexibility), not the *reason* that sample size was chosen.

As authors, participants felt their research workflow improved by having

preregistration as “a forcing function” (P13). For example, writing down study plans ahead of time might prompt them to think more thoroughly and to anticipate flexibilities/researchers’ degree of freedom in advance. P11 said that when they authored papers, they habitually preregister analysis code, “so there’s no place for that [flexibility] to happen”. Consequently, P11 was not concerned about unintentionally exploring other un-preregistered analyses—their flexibility was contained. Similarly, P13 reported that they would have peeked at their data and analyzed the data on the go if their study hadn’t been preregistered; preregistration prevented them from making decisions contingent on data.

On the reviewers’ side, participants report that preregistration makes the review process “easier” (P6). Participants, as reviewers, usually wonder about decisions with study design and analysis, especially “potential sources of flexibility” (P8). P6 reported using preregistration to understand the decisions made behind the paper under review:

I’ve been very impressed with ones [studies] who stuck with it [preregistration] and then also noted changes they had to make across time. It’s so much easier to understand and review papers when you understand the process that they went through, because often the papers have cut them out, so you don’t get to see decision-making . . .

Purpose: increasing transparency. Another purpose for preregistration that emerged from our interviews was that of increasing transparency. Lupia and Elman define *research transparency* to be production transparency (open access data or data collection procedures) and analytic transparency (“a full account of how they draw their analytic conclusions from the data”) Lupia and Elman (2014).

In contrast to participants describing *delimiting flexibility* as a purpose, P6 had a different view on the norms around power analysis. As described above (and in Figure 2), power analysis (or sample size rationale) is not strictly necessary for reducing researchers’ flexibility prior to data collection, only the sample size may be. However, P6 still suggested power analysis should be a part of preregistration:

I think [statistical] power is one of the biggest mistakes that are made in psychology, so not asking people to actually prove that they have the power of the test is a mistake, I think, to do.

This may reflect a desire to use preregistration to force researchers to “prove” they are following good practices by making deviations from good practices more transparent.

Also in contrast to the purpose of *delimiting flexibility*, P11 used the purpose of *increasing transparency* to justify preregistering exploratory analyses. P11 said:

It [preregistering exploratory analyses] just makes it clear that you’re not planning... Nothing was *a priori*, nothing was a prediction when you do that. And that’s just being honest, it’s just being transparent.

This may reflect a desire to be honest and transparent about an exploratory analysis—yet, this is an example of *increasing transparency* and not *delimiting flexibility*. If the sole purpose of preregistration were to delimit confirmatory from exploratory analyses, anything not preregistered would by definition be exploratory; thus, preregistering exploratory studies would not be necessary for the delimitation purpose alone.

P13 said that one benefit of preregistration was that it “makes me a better scientist or researcher, by being more transparent, forthcoming, ...”. Authors also thought that transparency brought them credibility, a positive signal to reviewers or readers in general. They might be perceived to be “thoughtful, open” (P4), their studies “followed-through” (P6). P10 thought that transparency reduced problems such as motivated reasoning (a researcher seeking results they want to see) and cherry picking. A natural consequence of increased transparency is that preregistrations become longer as more possible decisions are made transparent: *increased transparency implies increased length*. This created disagreements among participants: P14 said, “I don’t have a complaint that they [preregistrations] are too detailed”, perhaps considering the length of preregistration

necessary for proper use/evaluation, or perhaps subscribing more to a purpose of transparency.

It's important to note that the above purposes are not mutually exclusive: explicitly *delimiting* (or stating) *flexibility* is in a way being transparent. That said, the preference of one purpose over another can have tangible effects on the design of preregistration, see Discussion section.

Purpose: designing and implementing experiments. Within an experimental lab setting, participants used preregistration as a design tool and communication protocol/device. For example, P11 refined their experimental setup because preregistering made them consider more thoroughly. They realized that wanted to prevent experiment participants to switch response keys and have protocols for invalid responses. P2 called preregistration record “incredibly helpful when actually implementing the study”. They used their preregistration to instruct research assistants about experiment logistics such as how to set up the lab.

Purpose: learning about /teaching preregistration and research. Participants also used existing preregistrations for learning about or teaching about the practice. P9 assigned preregistrations as reading assignments; students were asked to evaluate and define what a good preregistration was. In summarizing the assignment, P9 said that it “really reflected my understanding of preregistration as a way to manage flexibility in analyses”.

When participants first started preregistering, existing preregistrations also helped them write their own. P13 used the term “meta-commentary” to describe how an existing preregistration helped them determine the level of detail they should use, another reflection of the uncertain norms around authoring preregistrations. Also, they benefited from published preregistrations by learning what information went into a preregistration itself and what went into the final paper.

Using preregistration

Creating preregistration. The process of creating preregistration is difficult to reconstruct with interview questions alone, so we asked P12 and P13 to walk us through a preregistration on which they were collaborating around the time of our interview. It was their first preregistration, and for easy collaboration, they had copied the AsPredicted template (see Figure 2) into a Google Doc. Prior to the two interviews, the interviewer reviewed the revision history on their Google Doc and designed specific interview questions about it.

From the two interviews, we learned that P12 and P13 completed the “easier” questions, such as title and hypotheses, first. Then they spent most of their time working on the Analysis section. When they revised their preregistration, they focused on clarity: clear hypotheses and the corresponding analysis appropriate for them. They also sought feedback from colleagues, because the example preregistration on [AsPredicted.org](https://aspredicted.org) has a simpler statistical model than theirs; this feedback helped them to refine the level of detail in their responses.

Reviewing the revision history of the Google Doc, we found that the responses to many questions changed from vague to specific over time. Specifically, phrases such as “the primary research question”, “exploratory analysis” and “confirmatory analysis” were added to later versions, reflecting participants’ updated awareness to use preregistration vocabulary (see Table 1). For the “Analyses” question, an initial vague description (“look at change in likelihood of [outcome]”) changed to describe a “logistic regression” with a description of independent variables. In the latest version, the two participants added R code to fit this model with informative variable names. This evolution—and particularly its focus on the analysis section—may reflect an increasing awareness of the participants’ parts of the need to delimit flexibility in the analysis.

Choosing between different preregistration formats. On the surface level, preregistration formats differ in lengths: P10 summarized preregistration formats as a

“spectrum from AsPredicted being very short and standardized to formats that are very loose and often longer”. For example, P5 used PROSPERO (<https://www.crd.york.ac.uk/prospero/>), a registry for health-related systematic reviews that contains 40 questions, while there are only nine questions in AsPredicted. As the most common formats in use in psychology, we compare the AsPredicted and OSF (Preregistration Challenge), the two formats popular among participants. Figure 2 lists all the questions in the two templates.

OSF, as a format, is often associated with detailed questions; for this reason, P14 stated that it is not “an entry-level product”. P14 regarded it as widely used and would recommend it for others to adopt. A possible criticism of the OSF format is that it covers perhaps too much and becomes too long: in talking about “formats that are very loose and often longer”, P10 preferred shorter formats, such as AsPredicted.

AsPredicted, on the other hand, is considered to be simple and faster to complete, be that a benefit or shortcoming. P11 said that “the interface is beautiful. It’s really simple, easy to use.” P6 also pointed out that AsPredicted was originally designed to be simple to make it easier and more likely for people to adopt preregistration. Not all participants thought this simplicity was an advantage, however. P6 thought that the AsPredicted format is “too watered down . . . to reduce the problems in science.”. P11, as an advocate for preregistering analysis code, considers AsPredicted “an executive summary of an actual preregistration”, meaning that AsPredicted was not detailed enough without analysis code preregistered. Length also plays a significant role in assessing preregistration quality, as we discuss next.

Checking preregistered studies. We roughly define *checking* as comparing a manuscript to its preregistration and verifying that the study and the analyses are conducted as the preregistration has specified. Authors, reviewers, and editors may all engage in checking. Our participants reported similar physical actions when checking: placing the two documents, the final paper and the preregistration, side-by-side, either on a

computer screen or on paper. Although the action of checking might be similar, we describe a *checking problem*: to what detail should or does an author or a reviewer/editor check?

According to the authors, their checking process was *linear and thorough*, probably because they would take full responsibility for their papers. P2 considered checking to be “more time-consuming than challenging” but not “cognitively demanding”. P7 said that while writing the paper, they “went back to it [the preregistration] multiple times and said, ‘Okay. I did exactly what I said here’”. P14 compared their paper and registered report “point-by-point”, likening the checking process to doing paper revisions.

Reviewers and editors, on the other hand, were not as exhaustive when checking compared to authors; instead, they had heuristics to *query* a preregistration. One reason to query the preregistration during reviewing was “when something looks out of place” (P6), such as an unlikely, large-but-significant p-value.

There are several reasons for which reviewers and editors were not exhaustive with checking. As a reviewer/editor, P4 was not even “bothered” to check preregistrations, because preregistrations are in the public record and therefore could be examined with post-publication peer review (see background section for examples of such review). Some reported trusting the preregistration to a degree: “I guess if somebody said that [they reported all decisions] in their paper, I would trust that, and I wouldn’t necessarily look to the preregistration.” (P8). In some cases, checking may also be unnecessary if the paper is very specific about referring to parts of the preregistration: “Here are our pre-registered exclusion rules.” (P9). Time constraints in the review process are also likely a factor.

Given the *checking problem*, P6 mentioned a checklist used by editors and reviewers at Journal A. It was a document that pointed to where the data and the preregistration were, if available, mediating the communication between editors and reviewers and reflected “everything we preach [about preregistration]” (P6).

Time commitment. One way to understand the difference in adoption and use of preregistration is in terms of time commitment. For example, we wanted to know whether

using preregistration takes extra time, compared to conducting research without preregistration.

Some authors and reviewers did not think preregistration took them extra time in research. The authors consider doing preregistration as a “time shift”, bringing the “decision points” before data collection instead of during analysis (P11) Some reviewers even think that a preregistration saves them the time they might have spent wondering about undisclosed flexibilities when working on an un-preregistered paper. Other participants thought that preregistration added time to their research, but consider the extra time and effort worth it: that “It’s just correcting a bad habit” (P13), and that “the benefit is a worthwhile trade-off” (P3).

Barriers to adopting and using preregistration

Not all participants considered preregistration useful for their work or necessary to do good research. We summarize these *barriers* to adopting or using preregistration.

First, there were participants who considered preregistration alienating. Researchers who do non-experimental studies expressed a “not for me” sentiment. For example, P5 worked with secondary data analysis and had concerns about their idea being scooped from a public preregistration. P6, who did not work in the experimental, hypothesis testing framework, complained:

I can’t go into the OSF and register a correlation study. We should be able to pre-register anything. And the fact that things are set up really for experiments makes it difficult because you’re like, “Well, that doesn’t really fit what I’m doing.” And it can turn you off so then you just stop, and so it’s turning off people from using it.

The many concepts related to preregistration and open science practices (see Table 1) also led to conceptual barriers. For example, some participants considered preregistration not useful for exploratory work as it limits flexibility. Some of the sentiments towards

preregistration included it being “too rigid . . . can backfire” (P4), or it “limit[ing] flexibility when doing the data analysis” (P2).

In addition, participants used different terminology—and sometimes even the same terminology to refer to subtly different concepts. For example, P9 preferred to distinguish between “focal hypotheses” and “non-focal” hypotheses in place of making a distinction between confirmatory and exploratory analyses. P14 used “preregistration” primarily to refer to registered reports and “self-registration” to refer to what we term “preregistration” in this paper.³ These differences in vocabulary may reflect different conventions among subfields or different understandings of the preregistration concepts at play.

P10 stated that misunderstandings about what preregistration is for—its purpose—might allow people to “game the system”. They alluded to the distinction between “the letters of the law” versus the spirit of the law: that is, following the steps of conducting a preregistration without understanding their purposes, versus understanding the purpose of preregistration, and so being able to better follow the steps. Consider, for example, the progression in understanding that P12 and P13 went through in conducting their first preregistration (described above), and the refinement in their understanding of what detail is necessary where, and why (e.g., the detail in the analysis section aimed at delimiting flexibility). Researchers who are not guided in the preregistration process by someone with a clear understanding of its purpose may not arrive at a similar understanding.

An undue emphasis on preregistration as a mark of quality is also viewed by some as potentially counterproductive. P3 was concerned about junior researchers “latching onto” preregistration as “extra boost”: “if it [preregistration] gets weighted far more than testing interesting and worthwhile ideas, then that’s a problem for the field”. In a similar vein, P14 felt frustrated, stating that publications with preregistration badges should not become “*de facto* higher quality research” (P14). These concerns were not exactly preventing our

³In both of these cases we have attempted to use the more common terminology in this paper.

participants from using preregistration,⁴ but we believe that they might become other researchers' barriers to adoption.

Discussion

Preregistration practice currently reflects the varying norms and purposes its users have. There is an opportunity to more explicitly reify different preregistration purposes in the design of preregistration tools, and to more explicitly support the needs of different user roles (authors, readers, and reviewers/editors) and disciplines in the design of the preregistration process.

How preregistration purposes translate into formats

We found that the interview participants had distinctive impressions about AsPredicted and OSF formats, but the link between the format differences and the purpose of using preregistration was not immediately clear. We speculate that preregistration formats (typically web forms) implicitly reflect what their designers consider the primary *purpose* of preregistration to be, through the inclusion and exclusion of questions (shown in Figure 2) related to each purpose.

For example, if a preregistration format, say AsPredicted, is primarily for *delimiting flexibility*, it would explain why its designer included a question for *Sample Size* but not for *Sample Size Rationale*, where justifications such as a power analysis might be placed (as mentioned in the Findings Section). Strictly speaking, only the sample size, not its rationale, need be recorded for *delimit flexibility*; the rationale is more important for the *transparency* purpose. By contrast, the OSF format explicitly prompts for sample size rationale; the length and comprehensiveness of that format perhaps suggest that its designers' purpose might have been maximizing *transparency*. Another possible explanation was offered by P7, who explained that the OSF template evolved out of a

⁴Though most of our interviewees are advocates and early adopters.

series of focus groups organized by OSF—perhaps the length of the OSF template is partially a result of a less directed purpose being enforced across multiple focus groups.

Purposes may undermine each other and result in conflicting designs

Our interviews suggest a tension between *delimiting flexibility* and *transparency*. Some of the most pointed remarks made by participants came when criticizing preregistration formats, either for lacking some desired question (e.g., a complaint that AsPredicted does not ask for power analysis) or for being too long and tedious (OSF). We believe, as a starting point, resolving this tension may be aided simply by making the intended purpose of a preregistration form explicit: is the primary goal transparency, delimiting flexibility, or something else? Current designs represent a failure of designers to make their internal model of preregistration clear to users: the purpose of each question, and indeed the entire form, is often opaque. If the purpose were clearer, users might better be able to self-select which form is most appropriate to their needs.

The tension between *increasing transparency* and *delimiting flexibility* is critical to resolve, as having designs based on both purposes can create confusion, and may result in preregistration being used in sub-optimal ways. For example, a long, detailed, maximally transparent preregistration may verge on the length of a full paper. It should then not be surprising that paper reviewers routinely fail to thoroughly review manuscripts with attached preregistrations, as evidenced both by blog posts examining published papers and descriptions from our participants: it amounts to asking for two reviews in one from already-overworked reviewers. If a user's (or indeed, journal's) primary purpose is delimiting flexibility, then, they may wish to adopt a shorter format in order to aid reviewers in assessing the match between paper and preregistration.

Designing for multiple purposes

Besides simply making purpose explicit, it should be possible to more directly design to support preregistration's multiple purposes. The existing preregistration tools (i.e.,

templates) were constructed according to the varying purposes that their different designers had in mind. However, an ordinary user of preregistration, especially a new adopter, may not be aware of the purposes behind the web forms they see. When the purpose ascribed to preregistration varies between, say, an author and a reviewer, this can create conflict: an author’s preregistration might be thought of as being too detailed, or not detailed enough, depending on the reviewer or reader.

The preregistration template interface could explicitly tell the user for which purpose it is designed. Taken further, preregistration templates could be dynamic, allowing users to adjust their purpose: e.g., if they wish to be more transparent, they could be shown additional questions or asked for additional detail. Then another user looking at the same preregistration who is focused on delimiting flexibility (perhaps during review) might switch to a view that hides content related solely to transparency. Explicit statements of purpose—for the entirety or for subsections—could be especially useful to new users to help them navigate preregistration. Without such guidance (as now), new users experience confusion about which questions are important, what level of detail is necessary for which aspects of the preregistration, and how to report results—as P13 put it, “what the right level of nuance is”. With a better understanding of the purpose of preregistration, a user could make more informed decisions about which preregistration format to use and how to use it.

Designing for different user roles explicitly

Since an author, a reviewer/editor, and a reader use preregistration differently, we propose to design preregistration targeting these user roles separately.

For *authors*, they design, conduct, and write about their studies. From our findings, the authors had checked their preregistrations thoroughly and used them to *design and implement experiments*. However, we suspect that one source of observed inconsistencies between final papers and preregistrations is that authors forget about or fail to check all

details in the preregistered study plan when writing up their reports. There is an opportunity to better integrate preregistration into existing tools that aid the experimental design and analysis process in order to improve report quality. Interactive tools for experimental design and/or analysis (for example, TouchStone Mackay et al. (2007), which aids in experimental design, or StatSplorer Wacharamanotham, Subramanian, Völkel, and Borchers (2015), which aids in analysis—even suggesting text to report a result in a paper), could have support for preregistration explicitly designed into them. Such a tool could aid a researcher in designing an experiment, preregistering that design, submitting that preregistration to an online registry, analyzing the results (while keeping track of any protocol deviations), and even give suggested text for writing up results with clear indications of what is exploratory/confirmatory and where protocol deviations occurred.

For *reviewers/editors and readers*, they check (if at all) the match between the completed study and the preregistration. An important constraint on this checking task is time and energy. According to our findings, reviewers/editors *query* the preregistration when their internal heuristics raise questions about a paper under review. While reviewers/editors reported that preregistrations made it easier for them to evaluate papers, checking by query might miss discrepancies between the paper and preregistration (which might explain the publication of papers with flawed preregistrations). Existing preregistration formats are designed primarily for authors to quickly input the desired information, but may verge on being write-only media: they are ill-suited to the checking task that reviewers, paper readers, and in some cases even authors themselves might perform. Thus, we recommend exploring designs that (1) make relevant preregistration content easier to *query* and (2) encourage wider *coverage* of preregistration content during review.

Some tools have already been proposed to make it easier for reviewers and readers to *query* relevant preregistration content from the paper. For example, Hardwicke’s SMART pre-registration Hardwicke (2018) proposes adding tooltips containing pre-registration

content into the relevant sections of the online version of a published paper. This makes it easy for a reader to look up relevant information from the preregistration, so long as it has been tagged by authors. We believe this kind of support for paper-to-preregistration lookup (and vice versa)—or something like it—could also aid the paper review process.

However, SMART pre-registration does not guarantee that a reader will check all (or any) preregistered information; thus it does not explicitly improve *coverage*. An alternative might be to offload part of the checking process onto authors: Since authors are most familiar with their research, they could quickly go question-by-question through their preregistration and indicate consistency between paper and preregistration, describing any protocol deviations. Such a *declaration of match* could be submitted with a paper, similar to cover letters describing revisions practiced in some venues. Then, reviewers need not manually do lookups between paper and preregistration to assess match, but could simply review the authors' declaration of match. Combined with tool support for pre-registration during the design and analysis phases (described above), generating this declaration could even be partially or fully automated.

Designing for different user disciplines

Since existing preregistration formats tend to be experimental study-oriented (AsPredicted) or discipline-specific (PROSPERO), users not within those research paradigms found preregistration alienating or not readily usable. For example, some of our participants are in developmental psychology and could not find an existing template on OSF for this field. The lack of template can be easily addressed by creating new ones; for example, at the Society for the Improvement of Psychological Science (SIPS) conference in 2018, a group of psychologists collaboratively drafted a template for secondary data analysis (<https://osf.io/bpuw3/>). However, creating new templates for every sub-field and discipline seems potentially wasteful; there may be more effective ways to overcome barriers to adoption.

Overcoming preregistration barriers. One potential barrier described by participants is that the *delimiting flexibility* purpose is sometimes misinterpreted as indicating that exploratory analysis should not be conducted at all: that preregistration is an indictment of exploratory analysis altogether. On the contrary, our participants were clear that the purpose is not to stop exploratory analysis, but to clearly distinguish it from confirmatory analysis. Researchers that value exploration may feel preregistration is “not for me”. Perhaps dynamic preregistration forms that allow researchers to self-select purpose could include exploratory analysis as a purpose, then guide newcomers in how to delimit exploratory from confirmatory analyses.

Another potential way to lower the adoption barrier is to make preregistration *simple*. P6 mentioned that the AsPredicted format was designed to be simple so that it would be easier for people to adopt. There may be advantages and disadvantages to using simplicity to promote adoption. For example, it is possible that a simple template might not effectively communicate the purpose(s) of preregistration. A researcher might be able to quickly adopt the AsPredicted template due to its simplicity, but without guidance could fail to specify a detailed enough and sufficiently constrained analysis plan, defeating the purpose of *delimiting flexibility*.

Limitations

Our recruitment focused on psychologists, many of whom are already interested in using preregistration. It is possible that a different group of psychologists or researchers in other fields will have alternative views about preregistration. We also focused on AsPredicted and OSF Preregistration Challenge templates, as our participants frequently used and commented on them. Other templates may compare differently.

Conclusion

In this paper, we investigated preregistration practices as a user-centered design problem. We uncovered two main purposes—increasing transparency and delimiting

flexibility—that drive preregistration use, and which may explain seemingly contradictory preregistration formats. We suggest that future preregistration design should make these purposes explicit, and explore ways to support preregistration continually across the research pipeline (from experimental design to analysis to write-up to paper review to paper reading). Without resolving these design tensions, the design of the preregistration process may undermine its primary goal of improving the reliability of scientific findings.

References

- Anderson, M. S. (2000). Normative Orientations of University Faculty and Doctoral Students. *Science and Engineering Ethics*, 6(4), 443–461. doi: 10.1007/s11948-000-0002-6
- Anderson, M. S., Martinson, B. C., & De Vries, R. (2007). Normative Dissonance in Science: Results from a National Survey of U.S. Scientists. *Journal of Empirical Research on Human Research Ethics: An International Journal*, 2(4), 3–14. Retrieved from <http://www.jstor.org/stable/info/10.1525/jer.2007.2.4.3> doi: 10.1525/jer.2007.2.4.3
- Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: has the evidence for ego depletion been overestimated? *Frontiers in Psychology*, 5(July), 1–11. Retrieved from <http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00823/abstract> doi: 10.3389/fpsyg.2014.00823
- Center for Open Science. (2018). *Registered report*. Retrieved Sep 17, 2018, from cos.io/rr
- Chambers, C. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49(3), 609–610.
- Chuang, L. L., & Pfeil, U. (2018). Transparency and Openness Promotion Guidelines for HCI. In *Extended abstracts of the 2018 chi conference on human factors in computing systems - chi '18* (pp. 1–4). New York, New York, USA: ACM Press. Retrieved from <http://dl.acm.org/citation.cfm?doid=3170427.3185377> doi: 10.1145/3170427.3185377
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, 7(1). doi: 10.1371/journal.pone.0029081
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r*. Sage publications.
- Giofré, D., Cumming, G., Fresc, L., Boedker, I., & Tressoldi, P. (2017). The influence of

- journal submission guidelines on authors' reporting of statistics and use of open research practices. *PLoS ONE*, 12(4), 1–15. doi: 10.1371/journal.pone.0175583
- Hardwicke, T. E. (2018). *SMART Pre-registrations*.
in Methods, A., & in Psychological Science, P. (2018). *Advances in Methods and Practices in Psychological Science Submission Guidelines*. Retrieved from
<https://www.psychologicalscience.org/publications/ampps/ampps-submission-guidelines>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), 0696–0701. doi: 10.1371/journal.pmed.0020124
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., . . . Nosek, B. A. (2016, may). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biology*, 14(5), e1002456. Retrieved from
<http://dx.plos.org/10.1371/journal.pbio.1002456> doi: 10.1371/journal.pbio.1002456
- Kupferschmidt, K. (2018). A recipe for rigor. *Science*, 361(6408), 1192–1193. Retrieved from <http://science.sciencemag.org/content/361/6408/1192> doi: 10.1126/science.361.6408.1192
- Lupia, A., & Elman, C. (2014). Openness in political science: Data access and research transparency. *PS - Political Science and Politics*, 47(1), 19–42. doi: 10.1017/S1049096513001716
- Mackay, W. E., Appert, C., Beaudouin-Lafon, M., Chapuis, O., Du, Y., Fekete, J.-D., & Guiard, Y. (2007). Touchstone: exploratory design of experiments. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1425–1434).
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago press.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration

- revolution. *Proceedings of the National Academy of Sciences*, 2017, 201708274.
Retrieved from <http://www.pnas.org/lookup/doi/10.1073/pnas.1708274114>
doi: 10.1073/pnas.1708274114
- Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, 31(30/3).
- Peng, R. D. (2009). Reproducible research and Biostatistics. *Biostatistics*, 10(3), 405–408.
doi: 10.1093/biostatistics/kxp014
- 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. doi: 10.1177/0956797611417632
- Simons, D. (2015). *HI-BAR: A gold standard brain training study?* Retrieved 2018-08-31, from <http://blog.dansimons.com/2015/11/hi-bar-gold-standard-brain-training.html>
- Teixeira da Silva, J. A., & Dobránszki, J. (2015). Problems with Traditional Science Publishing and Finding a Wider Niche for Post-Publication Peer Review. *Accountability in Research*, 22(1), 22–40. doi: 10.1080/08989621.2014.899909
- Tullett, A. (2018). *Guest post by Alexa Tullett: a not-so-hypothetical HIBAR*. Retrieved from <http://sometimesimwrong.typepad.com/wrong/2018/08/hibar.html>
- Wacharamanotham, C., Subramanian, K., Völkel, S. T., & Borchers, J. (2015). Statsplorer: Guiding novices in statistical analysis. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 2693–2702).
- Wagenmakers, E.-J., & Gronau, Q. (2017). *Two Pitfalls of Preregistration: The Case of Ego Depletion*. Retrieved Aug 31, 2018, from <https://www.bayesianspectacles.org/two-pitfalls-of-preregistration-the-case-of-ego-depletion/>
- Wagenmakers, E.-J., & Gronau, Q. (2018). *Musings on Preregistration: The Case of the*

Facial Feedback Effect.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A.

(2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science*, 7(6), 632–638. Retrieved from

<http://journals.sagepub.com/doi/10.1177/1745691612463078> doi:
10.1177/1745691612463078

Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., van Aert, R. C., & van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology*, 7(NOV), 1–12. doi: 10.3389/fpsyg.2016.01832

Table 1
Glossary of preregistration and open science related terms

Term	Explanation	Reference
Researcher degree of freedom (Analysis) flexibility	Particular choices researchers make during the study	Wicherts et al. (2016)
Replicability	Degrees of freedom in choosing dependent variables, sample size, covariates, and reporting experimental conditions	Simmons, Nelson, & Simonsohn (2011)
Reproducibility	The replication of scientific findings using independent investigators, methods, data, equipment, and protocols	Peng (2009)
Transparency	A "minimum standard" where "data sets and computer code be made available to others for verifying published results and conducting alternative analyses"	Peng (2009)
Exploratory studies	Sharing of "data, analytic methods, research materials, design and analysis"	Chuang & Pfeil (2018)
Confirmatory studies	"Finding hypothesis in the data by post-hoc theorizing"	Wagenmakers et al. (2012)
Preregistration Challenge	Preregistered studies such that data collection and data analyses leave no room whatsoever for exploration	Wagenmakers et al. (2012)
Post-publication peer review	A challenge by Center for Open Science (COS, cos.io/prereg), incentivizing preregistration with cash prizes. It also has a preregistration template.	
Power analysis	The peer review process that "takes place after a paper has been published"	Teixeira da Silva & Dobránszki (2015)
Badge	For significance testing, calculating the sample size needed to detect an effect of certain size with a given probability	Field, Miles, & Field (2012)
	A token awarded for a paper with open science practices such as preregistration, open data, and open materials (osf.io/tvyxz/).	

Table 2

Participant demographics. Most participants have written or were writing a preregistration, indicated in the “Author” column. Among faculty member participants, the majority had experience reviewing/editing preregistered studies, indicated in “Reviewer/editor”.

Participant #	Title/Seniority	Discipline	Gender	Author	Reviewer/editor
1	PhD student	Psychology	Male	Yes	No
5	PhD student	Psychology	Female	Yes	No
12	PhD student	Human-computer Interaction	Male	Yes	No
13	PhD student	Human-computer Interaction	Male	Yes	No
2	Assistant professor	Psychology	Female	Yes	Unclear
8	Assistant professor	Psychology	Female	Yes	Yes
14	Assistant professor	Psychology	Female	Yes	Yes
3	Associate professor	Psychology	Male	Yes	No
9	Associate professor	Psychology	Male	Yes	No
10	Associate professor	Psychology	Female	Yes	Yes
4	Professor	Behavioral Economics	Female	Yes	Yes
6	Professor	Psychology	Female	No	Yes
7	Professor	Psychology	Male	Yes	Yes
11	Professor	Psychology	Male	Yes	Yes

Figure 1. Preregistration in the research process. Steps related to registered reports are highlighted in color. Adapted from Center for Open Science at cos.io/rr.

Figure 2. Comparing two preregistration templates, OSF and AsPredicted (question numbers original). Questions shared by the two templates are highlighted in grey.