CHAPTER 19

Transparent Science: A More Credible, Reproducible, and Publishable Way to Do Science

David Mellor, Simine Vazire, and D. Stephen Lindsay

Background and Perspective

This is an exciting time to be a psychological scientist. There is a major new movement that seeks to promote the credibility and replicability of psychological research by enhancing its transparency, with scholarly societies promoting the principles (www.psychologicalscience.org/publications/open-science) and groups formed specifically to advance that mission (see http://improvingpsych.org/ and https://cos.io for two examples). While relatively low rates of replicability among scientific findings (Begley & Ellis, 2012; Chang & Li, 2015; OSC, 2015) inspired the existence of these groups, in this chapter we describe how striving to maximize transparency in your research can benefit both science and your career.

Traditional scientific publications present only a summary of the work conducted that led to the presented findings (Claerbout, 1994). This is at least partially a consequence of the limits of printed journals: In the past, there were simply no easy means of disseminating raw data, and sharing materials such as questionnaires or stimulus materials was costly and time consuming. Consequently, access to those details relied on personal communications. Some scientists also argued that it is more effective to tell a coherent story in a compelling way than to report the research as it was planned (e.g., Bem, 2004). The internet has largely eliminated obstacles to sharing materials, scripts, and data, and more and more researchers have come to understand that it is problematic to report analyses of serendipitous data patterns as though they had been predicted. Consequently, the scientific community is adopting new and improved practices.

We would like to thank Andy DeSoto for comments and improvements on an early version of this chapter.

Increased transparency has many benefits for the scientific community. Primarily, it enables an informed evaluation of scientific claims. Also, transparency enables others to build upon existing work more easily. And it allows errors to be spotted and corrections applied more quickly. Open access allows science to happen the way it is meant to happen.

Though increased transparency is better for science, it also carries potential risks for researchers. Such transparency may reveal errors, leaving the researcher open to criticism. It also may give advantage to other scientists, which can harm the individual in an ultra-competitive market for jobs. Transparency also increases the accountability of researchers - it becomes harder for a researcher to fool herself, and makes it easier for others to assess the basis for her claims. Thus, it is feared, ~~open access~~ could make it more difficult to publish.

On the positive side, being more transparent may be perceived as an indicator of confidence and credibility. By being more open with the data, research materials, and analysis plans, the individual researcher signals rigor and credibility, increasing the likelihood of publication. As these practices become more normative aspects of scientific conduct, the risks of those practices will lessen, and fewer people will be in a situation of "sticking their necks out" before others do so. As this reality emerges, the risk may shift toward those who are perceived as less open with important research details. Indeed, it is possible that norms will shift so much that research that is not reported transparently may not be considered credible.

Another benefit to being more open with research practices is to create final research outputs that are useful to future consumers of your work. In many cases, that future consumer will be yourself. Archiving your analysis code, study materials such as questionnaires, code books that give clear definitions of each variable, and of course the raw data, will help you in months or years ahead through later stages of write up or extension.

So how can you, as a researcher, use better practices to advance your career while conducting rigorous scholarship? This chapter will answer that question and will point to the tools and practices you can use to make your work more transparent. This will include transparent practices in:

- study planning and design with preregistration
- data analysis
- making outputs of your work more open
- use of preprints for transparency and quicker dissemination
- submission for publication and responding to reviewers.

Each section will focus on the need to create useful tools for you and other researchers. As you will see, this transparency will provide two self- and science-serving purposes: making your research more credible and reproducible, and making it more publishable

Preregistration

A preregistration is an immutable record of your plans for a research project that is created before you begin conducting that project (Lindsay, Simons, & Lilienfeld, 2016; Veer & Giner-Sorolla, 2016). The primary goal of a preregistration is to distinguish traditional hypothesis tests from serendipitous findings. This distinction is vital to how we interpret research. As the philosopher of science DeGroot (1969) has emphasized, the same data that were used to generate a hypothesis (i.e., exploratory results) cannot be used to test the hypothesis. Thus, if we want to use a study to test a prespecified hypothesis, we must make that clear ahead of time. It is also perfectly acceptable to use the same dataset to generate new hypotheses (i.e., conduct exploratory analyses), but it is important to be clear about this distinction. Without this clarity, tentative results can be unintentionally presented as confirmed findings, thus making them appear more credible than they actually are. Preregistering your planned analyses and predictions allows you to make it clear to yourself and others which findings were predicted and which were unplanned.

Prespecifying an analysis allows you to use and interpret inferential statistics to test hypotheses. These are sometimes called "confirmatory" analyses, because they are conducted to confirm (or disconfirm) a specific prediction. In contrast, the analyses that are inspired by unexpected patterns in a dataset are exploratory. The probability (p) values obtained in exploratory analyses cannot be interpreted in a straightforward way. The researcher may notice a trend in a dataset, and then conduct a specific analysis based on that observation to test the suggested hypothesis. When presented as a predicted result, this is known as HARKing, for "hypothesizing after the results are known" (Kerr, 1998). Likewise, making analytic decisions as to which variables or even which participants to include, after the results of the analyses are known, so that the p-value becomes significant, is another questionable research practice (QRP) known as p-hacking. The p-values yielded by inferential statistical tests (e.g., t-tests, ANOVAs) cannot be interpreted when p-hacking occurs.

When conducting confirmatory research (i.e., testing a hypothesis), creating a preregistration signals methodological rigor and gives additional credibility to the forthcoming results. Preregistrations address two major problems. First, when preregistrations are posted online in an open way, they increase discoverability of the work that is conducted but never published, known broadly as the "file drawer problem" (Pashler & Wagenmakers, 2012; Rosenthal, 1979), when only a biased subset (Franco, Malhotra, & Simonovits, 2014) of conducted work is ever reported. Second, they decrease the many subtle, post hoc decisions by the analyst, which may be biased to find the most "desirable" results (i.e., those that seem more surprising or more publishable, or that align with a favorite narrative). Simmons, Nelson, and Simonsohn (2011) showed how decisions such as continuing to collect data after undesirable findings or adding additional variables to a model can inflate the rate of finding a spurious, positive finding. Note that studies with no prespecified predictions or analysis plan can also be registered, to signal that you are committed to presenting the final results as exploratory.

The process of creating a preregistration can help you create more specific and reasonable plans and to think through details that you could otherwise be prone to miss. A preregistration will typically include the following items:

- the broad research questions that the proposed study will address
- a list of specific hypotheses to be tested
- a detailed study design
- a description of the data collection procedures
- your planned sample size
- a precise description of every variable included in the study
- a description of how any variable will be combined into more complex indices (e.g., "happiness" could be the mean of three responses to related survey questions), and how data will be preprocessed (e.g., how variables will be transformed, how missing data will be handled, and how outliers will be identified)
- any validity checks that will be conducted (e.g., checking for floor or ceiling effects, verifying that the manipulation had the intended effect using a manipulation check, etc.)
- a specific statistical model for every hypothesis listed: There should be a clear alignment for the reader between any hypothesis and a specific model; it should be clear which results will be taken as confirming and disconfirming the hypothesis, and how unexpected results will be interpreted
- the criteria you will use to make any inference (e.g., if using null hypothesis significance testing, your alpha; if planning a Bayesian

approach, your assumptions about the prior odds of each hypothesis and your criteria for interpreting Bayes factors).

A preregistration does not have to contain any background information or context about your proposed research. Its purpose is not to put the research into any wider context, but simply to make a very clear statement about any a priori plans and predictions. See http://datacolada.org/64 for arguments for keeping preregistrations short and to the point.

Writing a list of standard operating procedures (SOP) can help keep preregistrations more concise, while still showing that analytical decisions were a priori (Lin & Green, 2016). Standard procedures, exclusion criteria, thresholds, or criteria for using parametric over nonparametric statistical tests, etc., can be kept in one or more SOP documents on a public repository.

Having to consider the above details before beginning your work will improve your planning and preparation. The preregistration can also be the object of critique and improvement within your research group prior to beginning the study.

Another benefit of preregistrations is that they provide very credible justification for the use of one-tailed hypothesis tests. One-tailed tests specifically look for a difference or a relationship in a single direction instead of looking for any difference (e.g., group A will be taller than group B). The use of a one-tailed test implies that the direction was predicted before seeing the data and that only the single direction was tested (and therefore no conclusions can be drawn about differences in the opposite direction, no matter how large); preregistration satisfies both of those implications.

Preregistrations can be kept private while the work is ongoing (and some preregistration archives allow preregistrations to be kept private indefinitely). Authors can provide editors/reviewers with anonymized read-only links if they wish, so that those individuals can read the preregistration as part of the review process.

It is also possible, at an increasing number of journals, to submit a proposed research plan as a Registered Report (https://cos.io/rr). In this publishing format, the peer review occurs before data are collected, and the journal could decide to guarantee publication for work that is conducted and that meets prespecified quality checks. Decisions to publish are determined by the importance of addressing the proposed questions and the anticipated ability of the proposed methods to answer those research questions.

Registered Reports provide two additional benefits to the researcher. The first is the more rigorous and formalized peer review on the research plan itself. There is substantial opportunity for improving research design during the first stage of peer review in a Registered Report. Journals are loathe to publish shoddy work, so there is abundant motivation to point out methodological errors that would be unfixable in a research project that had already been conducted.

The second additional benefit is the simple guarantee to publish work that meets the prespecified quality checks. Authors benefit from knowing that their work will be published after meeting their commitment to conduct and analyze the study as promised.

Finally, consider reaching out to colleagues in order to start a research collaboration. Finding relevant collaborators with the necessary skills and coordinating the details of multiple locations of data collection require substantial effort. However, the collaboration may help increase your sample size, which is often smaller than it should be (Button et al., 2013). Also, by working through the details required to begin a multisite project, other experts from your discipline will have the opportunity to scrutinize the design and procedures and point out unanticipated errors. See, for example, the case study by Lithgow, Driscoll, and Phillips (2017) of a collaboration between three life-science labs that uncovered many unexpected pitfalls to clear and reproducible science. Increased transparency allows for increased scrutiny and increased opportunity for improvement. Though challenging, this increased scrutiny benefits the researcher, the publishability of the work, and ultimately the scientific process.

Planning for Sample Size, Power, Measurement Sensitivity

One challenge in planning a study is deciding how many participants to include. If your aim is to test a hypothesis about an effect, then sample size planning should be informed, in part, by an estimate of the size of that effect. Statistical power refers to the probability of rejecting the null hypothesis when it is false. If an effect is very large, then power to detect that effect will be high even with modest sample size. If an effect is very small, then power will be low unless sample size is large.

Researchers are often encouraged to set sample size to attain statistical power of 80 percent. But doing this is difficult for two reasons. First, most studies test numerous hypotheses (e.g., a main effect and an interaction, mediated effects, multiple planned comparisons, etc.), so the researcher must predict which of the key effects is likely to be the weakest. Second, it is often difficult to come up with a well-justified, specific estimate of effect size (Anderson, Kelley & Maxwell, 2017). If we already knew the effect sizes we wouldn't need to conduct the study.

Sometimes researchers estimate effect size based on small pilot studies, but such estimates are by their nature imprecise and even small differences in assumed effect size during power analyses can lead to very different recommended sample sizes. Also, the small samples that are likely to spark a researcher's interest in conducting follow-up research tend to be those that by chance yielded large effects.

One common basis for decisions about sample size is previous publications. The main problem with this approach is that *p*-hacking, together with publication bias (the tendency for journals to want to publish significant – and impressive – results), lead to inflated effect sizes in the literature. Even meta-analyses that attempt to correct for publication bias typically cannot accurately estimate the true effect size under a particular set of conditions (Pereira & Ioannidis, 2011), because that requires making assumptions about the extent of *p*-hacking and publication bias, something that is unknown and probably unknowable. One compromise is to find some studies or meta-analyses that test similar effects, then assume the real effect size is about half of what the published effect sizes are (e.g., in the Reproducibility Project: Psychology, the preregistered replications typically produced effect sizes about half as large as the original, unregistered studies).

Another solution is to decide the smallest effect size that would be of interest to you on theoretical grounds. That is, you could decide that an effect smaller than "X" would be inconsistent or uninformative with regard to your prediction or theory, and choose your sample size to have adequate statistical power to detect that effect size. This approach is sometimes called the "Smallest Effect Size of Interest (SESOI)". This approach is often ideal, because if you have high statistical power to detect the SESOI and you do not obtain a significant effect, then you can be fairly confident that any effect that exists is likely to be too small to be consistent with your prediction. However, this approach typically requires quite large samples (e.g., to attain 80 percent power to detect an effect of Cohen's $d = .10$ in a two-tailed test of two independent groups requires 3,142 subjects).

When your resources are limited, or recruiting participants from your population of interest is difficult, there are a few other options. One is to use sequential analysis (Lakens, 2014a). This allows you to check your results as you are collecting data, but adjust your inferential statistics so that you keep a constant alpha level (typically .05). This approach is a good compromise for maximizing efficiency, but the main drawback is that the effect size should be interpreted with caution.

If you are truly limited to a small sample size, it is essential that your results be tempered by the consequences of low statistical power. Even if you cannot draw conclusions from your results, the collected data could provide an incremental contribution to existing or future studies, which is another strong argument for data and materials sharing. Submitting to a data journal is a great way to maximize the chances that your data will be used for such purposes in the future.

One approach that should be avoided is to do a post hoc power analysis. In a post hoc power analysis, the researcher conducts the study, then uses the effect size obtained in the study to calculate how much power he or she had to detect an effect that size. There are many papers explaining the problems with post hoc power analyses (see Hoenig & Heisey, 2001; Lakens, 2014a).

Finally, another important question to ask yourself when choosing your sample size is whether you want to be able to draw conclusions if your result is not statistically significant. Ideally, studies would be designed such that any result is informative. Indeed, if you design your study such that you can interpret a significant result as evidence for your prediction, but you cannot interpret a nonsignificant result as evidence against your prediction, then you are setting yourself up to fall prey to confirmation bias.[1] That is, you are giving yourself permission to ignore any result that is not consistent with your prediction, and thus your hypothesis is not really falsifiable. Thus, you should ideally plan your sample size so that any result can be interpreted as evidence for or against your hypothesis. This requires going beyond power analyses, because a power analysis only applies when there is a true (non-zero) effect. Two more comprehensive approaches to planning your sample size (that take into account the possibility that the true effect is zero) are Bayesian statistics (e.g., https://jasp-stats.org/ see Wagenmakers et al., 2017) and equivalence testing (Lakens, 2017).

Data Analysis

If you created a preregistration prior to collecting or accessing your data, data analysis will be greatly simplified. There may be some deviations

[1] Nonsignificant results could of course also arise from poor study design. Ruling out such interpretations requires planning: How will you demonstrate that your manipulation worked, without looking at your main outcome variable? See the above section on preregistration for recommendations.

from your preregistered plan. If you encounter unexpected minor deviations from your prespecified plans, such as having to remove some participants who provided nonsensical responses, simply note the deviation and your justification for it, and provide the information necessary for others to verify it (e.g., the shared dataset). If the deviation is substantial, such as realizing that your proposed analysis does not test the precise hypothesis, you are under the general obligation to report the proposed analysis and also the better, more appropriate test. You can justify the improved analysis in your methods transparently. This new analysis should be presented as exploratory, because it was not constrained by the preregistration.

If you did not preregister your analysis, it is important to be aware of the biases that lead to HARKing, *p*-hacking, and cherry-picking your results, so that you can do your best to avoid them. However, it is important to remember that preregistering analysis plans and predictions is the best way to obtain interpretable *p*-values, so these results should always be taken as less certain than the results of preregistered analyses. Good practices in reporting unregistered analyses include:

- Start documenting your analysis at the data-cleaning stage. Use a version control system such as GitHub to track your progress from data cleaning through final analysis.
- If your *N* is sufficient, randomly split your dataset into two parts (Anderson & Magruder, 2017). Conduct as many exploratory tests as you want on one part of the data. When you find something interesting or significant, create a preregistration with that suggestive finding, uncover the other half of the data, and confirm your finding on that pristine dataset.
- Use statistical software that allows you to write precise instructions as a script, which permits others to see the exact analysis you conducted. Other statistical software, where instructions are entered through a menu, leaves a less precise record of the tests that you conducted. The open source software R is widely used, but others such as SPSS are also possible. If you're not yet experienced with these tools, there are many free online resources for learning, and many university departments are offering training in these exciting new tools.
- Report all relevant descriptive statistics, including associations among your measured variables.
- Present your data graphically in ways that maximize the depiction of individual data points. For example, instead of using a bar chart,

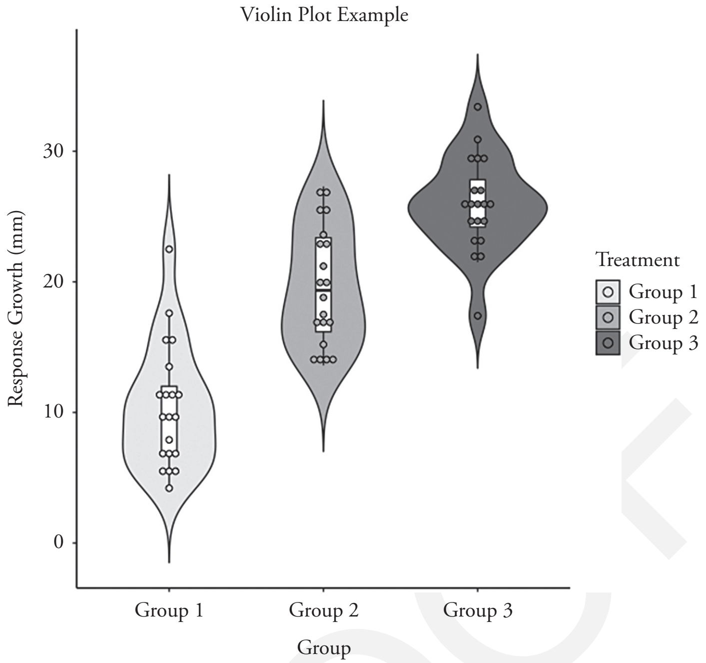

Figure 19.1 An example violin plot using data modified from Crampton (1947). The width of each plot represents the probability density of data distributed within each group. In this example, individual data points are also plotted, which is possible for relatively small datasets.

use a violin plot or frequency histograms in order to present a more full distribution of the data (see Figure 19.1 for an example). If the relationship between two variables is of interest, display that relationship in a scatterplot.

- Adding new moderating variables or covariates to a statistical model represents a new hypothesis that you are testing. Do this with caution, as the number of such hypotheses may be much larger than you initially thought. If you try any such tests, report all of their results as purely exploratory that require preregistered confirmation prior to drawing any conclusion.
- Comprehensively report your design and analysis to help add clarity. Use a reporting checklist relevant to your study design to ensure that your work is reported in sufficient detail. The Equator Network (http://equator-network.org/) includes reporting guidelines for most types of studies, and many journals do as well.

- Calibrate your conclusions to the strength and robustness of your evidence. If your results are weak or do not hold up under other, reasonable analysis specifications, describe your evidence as suggestive or inconclusive.
- If possible, preregister a direct replication of any interesting findings you uncover through this process. It will add substantial credibility to your paper, and is the best way to provide a real test of the new hypothesis.

Open Research Products

One of the most substantial ways to be more open with your research is to freely share all relevant outputs of a project. This improves the overall presentation of work submitted for publication, which can be influential in the editorial and peer review process. As is true with other open research practices, they benefit both the researcher and the wider research community. In particular, they allow others to reproduce the results in your manuscript (i.e., reanalyze your data) and to attempt a new replication study. As noted by the US NIH (https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm), sharing these research outputs encourages open scientific inquiry, promotes alternative explanations and hypotheses, enables new knowledge and data synthesis (such as by meta-analysis), improves education of early-career researchers, allows for exploration into new and unanticipated topics, and allows for the creation of new and larger datasets.

Materials

Sharing the materials used in your research will help readers better understand what you measured or manipulated and how. Access to these items may address many potential questions readers will have that could not be addressed thoroughly in the manuscript itself (e.g., the specific items on a questionnaire, or the specific language used by an experimenter or confederate). This will help readers evaluate the design of your study, and interpret the meaning of the results. While some measures are copyrighted in such a way that researchers are not free to reprint them, most materials can and should be shared completely.

Thorough descriptions of all materials and procedures also enables others to attempt to replicate your study. These descriptions will make it easier for others to build on your work, and will save you time you may have had to spend answering other researchers' questions about how to copy or adapt your materials or procedures. Thus, when sharing materials, you should include any details that you think should be held constant if another researcher wishes to attempt a direct replication (e.g., If the gender of the experimenter matters, that should be described, but if not, it can be omitted – see also the section below on "Constraints on Generality").

Data and Scripts

Openly sharing data from research projects can be a major signal of confidence in your reported outcomes. Indeed, it is odd that not having access to data underlying claims has been a normal part of the culture of science, where skepticism and verification of evidence-based claims are core to its success (Vazire, 2017). Even so, most researchers do not share data when requested (Savage & Vickers, 2009; Wicherts et al., 2006) or even required (Alsheikh-Ali et al., 2011) to do so, contact information changes (Vines et al., 2014), and even the most compliant scientists can only satisfy the requests if they are alive.

Perhaps the most important justification for not sharing data is for ethical considerations. Most social scientists use data collected from humans. Collecting information that may be embarrassing or simply private is sometimes necessary, and the risk of that information being linked back to the participant is a risk that the scientific community must not take lightly. Not only would harm be done to the affected participant, but future participants would be less likely to take part in research if they did not trust the scientific community to take their privacy seriously.

Balancing the privacy concerns of individuals with the benefits of sharing data need not be an overwhelming task. Creating a detailed plan for collecting and storing data, separating any personally identifiable information from the variables used in an analysis, and frequent communication with your university ethics boards can ensure that you are following best practices.

When data cannot ethically be shared publicly, there are options for making the data easily accessible to other scientists who meet certain criteria. For example, some data repositories will provide security guarantees, and can vet requests for data and make sure that only people with appropriate qualifications (e.g., ethics training) gain access to the data. This has the advantage of guaranteeing the security and longevity of the data, and relieves the original researchers from the burden of responding (or not) to individual requests.

In addition to sharing their data, researchers should also share the code or scripts used to analyze the data. This code should be executable (including information about what software and version is needed, any settings that should be used, etc.) and should be sufficient to reproduce all the results presented in the manuscript.

Codebooks

Sharing a large, raw dataset is possibly useful to a future researcher investigating related claims, or to a contemporary reviewer. However, the meaning behind every value is only preserved if it is well documented at the time of collection. A "data dictionary," "codebook," or the omnibus phrase "metadata" all describe the same thing – a way to decode the otherwise meaningless jargon contained in your files. A good data dictionary will define the following items within a data file for every variable:

- variable name, as written in the data file
- units
- allowed values, which are necessary for both quality checks and context; for example, if you are measuring height, then the allowed values have to be numerical and greater than zero
- a definition of the variable
- a synonym or other descriptor of the variable, if necessary and not covered by the definition.

Constraints on Generality

Traditional incentives encourage researchers to present their findings as general truths with broad relevance. One might, for example, find that undergraduates exposed to misleading suggestions regarding a crime video often later reported having witnessed what was in those suggestions, and cast that result as evidence that real witnesses are likely to form false memories of suggested details. The more general and universal the claim, the more impressive it may seem. But what is really impressive is being right – that is, making claims that stand the test of time and turn out to usefully advance understanding of psychology.

The findings of an individual study may have limited generality beyond the conditions of that study (in terms of the participants, time period, materials, and procedures). It is standard practice to acknowledge that possibility, writing something along the lines of "further research is needed to establish the generality of these results." But very often such caveats are vague or empty and are undermined by other parts of the manuscript that insinuate wide generality.

As one indicator of the reality of this problem, consider cases in which a research group fails to replicate the results of another group. It is common, in such cases, for the authors of the original finding to point to differences between the original and the replication study. Such defenses may well have merit, but much less so when the points of difference were not highlighted, in the original work, as important for obtaining the effect.

Simons, Shoda, and Lindsay (2017) argued that authors should include in their reports detailed expositions of their beliefs regarding constraints on the generality of their findings. If you have empirical, theoretical, or even intuitive grounds for believing that the results of your study are specific to particular characteristics of your subjects, materials, and/or procedures you should make that clear. Specifying likely constraints on generality is important for theory development and lays the groundwork for follow-up research.

The benefits of identifying anticipated constraints on generality should not be purchased at the cost of undermining interest in your research. Don't oversell your work, but don't undersell it either.

Preprints

Preprints are a way to quickly share the findings of any study before the results have been evaluated through peer review. The most well established preprint server is ArXiv (https://arxiv.org/), hosted by Cornell University, but there are new preprint servers being established, such as PsyArXiv in psychology, SocArxiv and SSRN for the social sciences, and bioRxiv for the life sciences. The recent proliferation of these services underscores their appeal to the research community.

Their appeal lies in their myriad benefits. These include:

- quicker dissemination of findings
- the opportunity to receive feedback sooner
- a way to show evidence of productivity as an interim research output
- a means to bring transparency into the research process by surfacing changes that occurred between rounds of revision and improvement
- a way to quickly share research findings that is typically compliant with most copyright agreements as long as the final, copyedited version is not shared as a "preprint"

- providing open access to research reports regardless to those who lack subscriptions to traditional scientific journals.

While popular, there are some risks with such rapid dissemination of findings. You may wish to consider soliciting feedback from a few colleagues before you post a manuscript on a preprint server; that way, if there is an obvious problem you can find out about it before telling the world about it. Peer review is the gold standard in scholarly credibility because of its utility in catching errors or oversights by an original author. While errors can and do slip through the peer review process, it is nonetheless often effective at improving output. Preprints should be considered preliminary reports subject to change (though, arguably, so should published articles).

Submission and Responding to Reviewers

Most journals invite authors to write a cover letter that provides background information about the submitted manuscript. Use this opportunity to communicate to the editor the ways in which transparent processes add to the rigor and credibility of your manuscript. Make sure to take the time to point out features such as openly available data or materials, and of course to link to any preregistration included in the study. By including these points in the cover letter, it becomes clearer what actions you are taking that may be above and beyond what is expected.

The other main avenue for correspondence during the publication process is when responding to the editor's and reviewers' comments when submitting a revision. In this venue, it is appropriate and expected to point to possibly overlooked evidence of rigor. One fear about increased transparency is exposing oneself to critiques that could harm publishability of results. You may, for example, face unwanted "encouragement" to *p*-hack or HARK. Simple and direct responses to such encouragements should maintain the spirit of preregistration: "The presented work includes all of the a priori hypotheses, but I can include these additional analyses as exploratory work that deserves later confirmation." Then, simply add those additional analyses as preliminary results.

Finally, editors or reviewers may express disappointment that your results are less perfect and less groundbreaking (because they are less counterintuitive or less exaggerated) than they expect. In that case, it may be useful to point out that results from unregistered studies often lead to inflated effect sizes and findings that may not be as replicable.

Preregistered and transparently reported findings may be less surprising than findings that result from less constrained research practices (e.g., when undisclosed flexibility, such as *p*-hacking and HARKing, occurs), but the trade-off is that the findings from preregistered and transparently reported studies are more credible. Making this point effectively requires considerable diplomatic skill and tact.

Since most people are not in a position to set policy that would reduce such biases from affecting publishing decisions, what are the best steps you can take if confronted with these biases? Submit to journals that adhere to and reward open science practices. Journals that implement policies designed to increase the openness and reproducibility of science are more likely to be familiar with the importance of these practices. Journals that use policies covered by the Transparency and Openness Promotion (TOP, https://cos.io/top) Guidelines (Nosek et al., 2015), that accept Registered Reports, or that issue badges that signal open science practices (https://cos.io/badges) are all taking steps to address these biases and may be more open minded about the benefits of a more open science. Finally, read the journal's most recent editorial to find out what the current editorial team considers best practices, and submit to journals that emphasize transparency and reproducibility.

Conclusion

The primary benefit of transparency to the researcher is the ability to increase the credibility of one's scientific claims. Sharing data, minimizing bias toward novel or significant findings, and improving documentation of the research process clearly improve the credibility of one's claims, and this can be conveyed to the editor, reviewer, and your readers. The technological hurdles to being more open with research materials have been removed, and so now is an ideal time to stand out in a crowded field by taking simple steps to be more transparent.

One risk of being more open is the risk of uncovering errors that would otherwise be missed. We cannot prevent you from making such errors, but we can assure you that the best way to respond to such occurrences is with an equal measure of transparency. If an error is spotted in your work, take a bit of time to swallow any defensiveness, and then thank the person. Depending on the reach of the work, the error may require some public correction in the scientific literature. If you take the correction seriously, your reputation is likely to be improved.

Another common fear with a more open psychological science is of generating a body of work that is less creative and too focused on credibility over making new, unexpected discoveries. We believe that this situation is unlikely to occur, given the very strong bias we all share toward valuing new discoveries. The motive to find something previously unexpected would not be diminished under a system in which preregistration or data sharing is commonly practiced. There will simply always be a desire to find something unknown, and our ideal scientific community is one in which rigorous confirmation and credible evidence is valued more, not one in which discoveries are valued less. For a more complete discussion on the proper balance between creativity and credibility, see Vazire (in press) and Wagenmakers et al. (2018).

Our hope is that the benefit of increased rigor and replicability will clearly outweigh any possible costs. All else being equal, journals should prioritize publishing findings that are likely to be true over findings that grab people's attention or make bold claims, and researchers should value finding the truth over getting attention or confirming one's predictions.

Our expectation is that more openness in research will become increasingly expected. However, you can work to make this transformation happen more quickly through your work, and, in the process, benefit from the increased credibility of your research.

References

Alsheikh-Ali, A. A., Qureshi, W., Al-Mallah, M. H., & Ioannidis, J. P. A. (2011). Public availability of published research data in high-impact journals. *PLoS ONE*, *6*(9), e24357. https://doi.org/10.1371/journal.pone.0024357

Anderson, M., & Magruder, J. (2017). *Split-sample strategies for avoiding false discoveries. w23544*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w23544

Anderson, S. F., Kelley, K., & Maxwell, S. E. (2017). Sample-size planning for more accurate statistical power: A method adjusting sample effect sizes for publication bias and uncertainty. *Psychological Science*, *28*(11), 1547–1562. doi:10.1177/0956797617723724

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. doi:10.1038/483531a

Bem, D. J. (2004). Writing the empirical journal article. In M. P. Zanna & J. M. Darley (Eds.), *The compleat academic: A practical guide for the beginning social scientist* (2nd edn., pp. 185–219). Washington, DC: American Psychological Association.

Board of Governors of the Federal Reserve System, Chang, A. C., & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say "usually not." *Finance and Economics Discussion Series, 2015*(83), 1–26. doi:10.17016/FEDS.2015.083

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, Emma S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. doi:10.1038/nrn3475

Claerbout, J. (1994). Seventeen years of super computing and other problems in seismology. *Paper presented at the National Research Council meeting on High Performance Computing in Seismology*. http://sepwww.stanford.edu/sep/jon/nrc.html

Crampton, E. W. (1947). The growth of the odontoblast of the incisor teeth as a criterion of vitamin C intake of the guinea pig. *Journal of Nutrition, 33*(5), 491–504. http://jn.nutrition.org/content/33/5/491.full.pdf

Degroot, A. (1969). *Methodology: Foundations of inference and research in the behavioral sciences*. The Hague: Mouton.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*(6203) 1502–1505. doi:10.1126/science.1255484

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power. *American Statistician, 55*(February), 19–24. doi:10.1198/000313001300339897

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*(3), 196–217. doi:10.1207/s15327957pspr0203_4

Lakens, D. (2014a). Performing high-powered studies efficiently with sequential analyses: Sequential analyses. *European Journal of Social Psychology, 44*(7), 701–710. doi:10.1002/ejsp.2023

(2014b, December 19). The 20% statistician: Observed power, and what to do if your editor asks for post-hoc power analyses. Retrieved November 11, 2017, from https://web.archive.org/web/20170711191030/http://daniel-lakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html

(2017). Equivalence tests: A practical primer for tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. doi:10.1177/1948550617697177

Lin, W., & Green, D. P. (2016). Standard operating procedures: A safety net for pre-analysis plans. *PS: Political Science & Politics, 49*(3), 495–500. doi:10.1017/S1049096516000810

Lindsay, S., Simons, D., & Lilienfeld, S. (2016). Research preregistration 101. *Observer, 29*(10), 14–16.

Lithgow, G. J., Driscoll, M., & Phillips, P. (2017). A long journey to reproducible results. *Nature, 548*(7668) 387–388. doi:10.1038/548387a

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. doi:10.1126/science.aab2374

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. doi:10.1126/science.aac4716

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528–530. doi:10.1177/1745691612465253

Pereira, T. V., & Ioannidis, J. P. A. (2011). Statistically significant meta-analyses of clinical trials have modest credibility and inflated effects. *Journal of Clinical Epidemiology, 64*(10), 1060–1069. doi:10.1016/j.jclinepi.2010.12.012

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641. doi:10.1037/0033-2909.86.3.638

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. Ed. Chris Mavergames. *PLoS ONE*, *4*(9), e7078. doi:10.1371/journal.pone.0007078

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. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, August, 174569161770863. doi:10.1177/1745691617708630

Vazire, S. (2017). Quality uncertainty erodes trust in Science. *Collabra: Psychology, 3*(1), 1. doi:10.1525/collabra.74

(in press). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*.

Veer, A. E. van 't, & Giner-Sorolla, R. (2016). Pre-registration in social psychology – A discussion and suggested template. *Journal of Experimental Social Psychology, 67*(November), 2–12. doi:10.1016/j.jesp.2016.03.004

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., … Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*(1), 94–97. doi.org/10.1016/j.cub.2013.11.014

Wagenmakers, E.-J., Dutilh, G., & Sarafoglou, A. (2018). The creativity-verification cycle in psychological science: New methods to combat old idols. *PsyArXiv*. doi:10.17605/OSF.IO/37NTP

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., et al. (2017). Bayesian inference for psychology. Part II: Example Applications with JASP. *Psychonomic Bulletin & Review,* July. doi:10.3758/s13423-017-1323-7

Wicherts, J. M., Borsboom, D., Kats, J., Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726–728. doi:10.1037/0003-066X.61.7.72