CrossMark

The what, why, and how of born-open data

Jeffrey N. Rouder[1]

Abstract Although many researchers agree that scientific data should be open to scrutiny to ferret out poor analyses and outright fraud, most raw data sets are not available on demand. There are many reasons researchers do not open their data, and one is technical. It is often time consuming to prepare and archive data. In response, my laboratory has automated the process such that our data are archived the night they are created without any human approval or action. All data are versioned, logged, time stamped, and uploaded including aborted runs and data from pilot subjects. The archive is GitHub, `github.com`, the world's largest collection of open-source materials. Data archived in this manner are called *born open*. In this paper, I discuss the benefits of born-open data and provide a brief technical overview of the process. I also address some of the common concerns about opening data before publication.

Keywords Open science · Open data · Data integrity · Data sharing

Psychological science is beset by a methodological crisis in which many researchers believe there are widespread and systemic problems in the way researchers produce, evaluate, and report knowledge (Pashler & Wagenmakers, 2012; Yong, 2012). This crisis is precipitated by the publication of fantastic extra-sensory perception claims in mainstream journals (e.g., Bem, 2011; Storm, Tressoldi, & Di Risio,

2010), by the widespread belief that well-publicized effects may not be replicable (Carpenter, 2012; Kahneman, 2012; Roediger, 2012) and by several cases of outright fraud. Such a crisis is very worrying because if researchers and labs cannot have confidence in one another, then the core of the field is at risk.

This methodological crisis has spurred many proposals for improvement including an increased consideration of replicability (Nosek et al., 2012), a focus on the philosophy and statistics underlying inference (Cumming, 2014; Morey et al., 2013), and an emphasis on what is now termed *open science*, which can be summarized as the practice of making research as transparent as possible. Practitioners of open science make several elements of the research freely available at an archival repository. These elements include the details of data collection, the data themselves, and code for analysis. The rationale behind making science open is clear—as psychological science becomes more open, questionable practices such as fortuitous censoring of subjects becomes more detectable. Moreover, and more importantly, researchers who place their methods, data, and analyses in open sources have an incentive to use better judgment in their methodology and more carefully consider the ramifications of their decisions in analysis. Indeed, Wicherts et al. (2011) report that research with open science has stronger evidence and more sound analyses than research not in the open.

One critical part of open science is *open data*, the act of making raw data available on demand to anyone. Recently, there has been a grass-roots movement, The Agenda for Open Research (`agendaforopenresearch.org`), for researchers to make their data open and for journals to insist they do so.[1] The rationale and context for this

---

✉ Jeffrey N. Rouder
rouderj@missouri.edu
http://www.pcl.missouri.edu

[1] University of Missouri, Columbia, MO 65211, USA

---

[1] I am a signatory to this agenda and support it completely.

call are provided in Morey et al. (2014). Accordingly, a researcher may make her or his data open by publishing them at a sanctioned or curated web repository with some form of institutional support. Examples of such repositories dedicated to archiving open-source materials include GitHub (github.com), Figshare (figshare.com), Dryad (data dryad.com), and Open Science Framework (osf.io). Other curated sites include university-related archives and society-related archives. By placing data in one of these curated sites rather than on personal websites the community may be assured of its long-term accessibility. In contrast, data placed on personal websites may not last long or be properly documented (Klein et al., 2014; Krawczyk and Reuben, 2012; Vines, 2014; Wicherts & Bakker, 2012).

Open data and data-on-request

Many researchers I encounter endorse the notion that openness and transparency are defining features of science. Simply put, scientific claims should be open to peer scrutiny. One example is peer review, where our scholarship and the logic of our arguments are placed under intense scrutiny. Most researchers would endorse the idea that data and analyses should be open as well. To meet this requirement, many colleagues tell me that if asked for data, they would gladly comply. I call this mode of data sharing as *data-on-request*. If authors always responded timely and diligently to data requests then data are in effect open to scrutiny just like other research endeavors. Unfortunately, they do not. Wicherts, Borsboom, Kats, & Molenaar (2006) asked 149 author teams to release 249 data sets that had appeared recently in American Psychological Association journals. Only 11 % complied with the initial request, and only an additional 16 % with repeated requests. A full 73 % of author teams never complied.

To me, the dismal failure of the data-sharing experiment means that data are effectively closed. It is worth considering, at least briefly, why researchers may not respond to a data request: Perhaps the raw data are unavailable—they may exist on a computer that is no longer in service. Alternatively, perhaps the data are in an arcane format that is no longer readable. Perhaps the code to analyze the data is missing; perhaps the metadata that explains what the fields mean is missing. Sometimes investigators delegate data curation to graduate students and postdoctoral researchers whose tenure at the institution is necessarily limited. Investigators themselves may become ill, leave the institution, or even die. The bottom line seems to be that researchers who practice data-on-request are rarely prepared for such a request, making it difficult if not impossible to comply.

Wicherts et al.'s and others findings strike me as unnerving. Without a field-wide commitment to truly open data, it is difficult to have the utmost confidence in experimental findings in the literature.

Psychologists hardly ever practice open data. Just think of all the data behind all those publications and ask what proportion are available to you right now. This proportion is assuredly negligible. The lack of open data is rampant in other disciplines such as biology and ecology (Punewska, 2014). The good news is that there is increased awareness in psychology about the desirability of open data. For example, *Psychological Science* now notes when data are open in the associated research articles. Moreover, there is a new journal, *The Journal of Open Data in Psychology* that peer reviews data sets to insure they are open, the metadata is sufficient in quality, and the archive is curated (Wicherts, 2013), and there are an increasing array of services like Open Science Frameworks where data uploads are straightforward.

And yet, practicing open data remains hard. In fact, until recently I was an open-data hypocrite. Although I was committed to open data, I was not implementing it in practice. I made my commitments boldly in data-management plans for grant-funding agencies. My data were supposed to be archived at MoSpace, the digital archives of the University of Missouri. MoSpace is professionally maintained and supported; completely open; and searchable. Yet, sadly, much of the data were not archived. Why not? Some of it was a lack of effort. It was a pain to document the data; it was a pain to format the data; it was a pain to contact the library personnel; it was a pain to figure out which data were indeed published as part of which experiments. Some of it was forgetfulness. I had neither a routine nor any daily incentive to archive data. Even with the best of intentions, it seems that making data open took too much time, effort, and attention. No wonder people revert to making data available upon request. It is so much easier than making data open.

Born-open data

If we were to make our data open, neither I nor my students could be relied upon to do routine tasks like file uploads. Instead, we needed an automatized system. And that is what we came up with: Behavioral data are uploaded to GitHub (github.com, an open web repository where they may be viewed by anyone at any time without any restrictions. This upload occurs nightly, that is, the data are available within 24 hours of their creation. The upload is automatic—no lab personnel is needed to start it or approve it. The upload is comprehensive in that all data files from all experiments are uploaded, even those that correspond to aborted experimental runs or pilot experiments. The data are uploaded with

time stamps and with an automatically generated log. The system is versioned so that any changes to data files are logged, and the new and old versions are saved. In summary, if we collect it, it is there. And it is open and transparent. I call data generated this way as *born-open data* and hope born-open data become a standard.

To see a subset of data we have been collecting since the start of 2015, point your browser to `github.com/PerceptionCognitionLab/data1`. One folder is `bayesObserver`, which contains the data from a set of experiments designed to assess whether people combine information from the stimuli with base rates in an ideal manner. Here there are a few experiments, and let's explore `be1`. Each file corresponds to a different participant, and when clicked on, the raw data are available. Of course, it is difficult to understand what the data mean without column labels, and these are provided in a separate file, `be1.txt`. Even with the column labels, it is quite difficult to understand the experiment or the meaning of the data without a guiding publication. But with the publication, the data are easily understood and analyzed. Hence, we document when we collected the data, and it is open to scrutiny to anyone with the corresponding manuscript.

I have found that there are several advantages to using born-open data:

- The use of born-open data incentivizes me and my students to use the highest level of judiciousness in analysis. I suspect that we will have an increased awareness of our decisions and their consequences. This is the intended effect, and uncomfortable as it may be, I view it as an advantage.
- With born-open data, we do not make data-management mistakes. It is much easier to audit and document all data and all analysis steps. Gone are manually labeling data or wondering if the version on the memory stick or the hard drive is the latest. I load data into the analyses right from GitHub without any intermediary downloads or any proliferation of files. There are no more "raw" files and "cleaned" files. Instead, there is just data and code. Here is snippet of code that loads up data from Experiment be1. It can be run on any computer with R and the R package "RCurl."

- With born-open data, backup is automatic. The GitHub copy backups the local copy. Importantly, both copies are versioned. Consequently, there is no reason to worry about weekly incremental backups, monthly backups, restoring from backup and the like.
- Born-open data simplifies sharing of data within the lab and with collaborators. I often e-mail collaborators R code that loads the data from GitHub and cleans the data, and they can provided analyses as they wish.
- Born-open data increases the long-term availability of the data. Because data are archived at a site with institutional support, we can reasonably expect them to be available on a scale of several decades.

Technical elements

In this section, I describe the three technological elements I use to create born-open data. The three basic elements are: 1. a shared local storage file system, 2. the use of Git repository software and the GitHub open-source web repository, and 3. automation through a task scheduler. I provide a brief overview of each step. There is good and bad news here for those wishing to implement the system. The good news is that all the protocols are standard, well used, and available for all major operating systems including Windows, Mac-OSX, and Unix/Linux. Any skilled IT professional should know them well or be willing to learn them. The bad news is that implementation will require a bit of tinkering with the specifics dependent on machines, networks, and operating systems. Given that archiving data is a critical enterprise that may affect the researcher for the length of her or his career, it seems that investing in these or comparable technologies is warranted and beneficial.

Element 1: Shared local storage

In my lab, behavioral data are collected across several computers. One key problem is coordinating among these computers. Code to run the experiments must be placed on each, and the outputted data must be merged into a master set. These tasks must be done accurately to insure the

```
library(RCurl)
name1="https://raw.githubusercontent.com/PerceptionCognitionLab/data1/master/"
name2="bayesObserver/be1/be1.all"
intext=getURL(paste(name1,name2,sep=""))
dat=read.table(text=intext)
colnames(dat)=c('sub','trl','blk','tBlk','cond','ecc','stim','rt','resp','pts','ptsTot')
```

integrity of the data. In some labs, assistants move files from one machine to another, often via memory sticks. Not only is this approach labor intensive, it may not be reliable. The better approach is to use one shared drive. The lab server shares a drive where both the experimental code and the data are stored. This drive is the master drive, and the other computers that collect the data read and write to it rather than to their own internal drives. When behavioral data are created, then they are created only on the master drive and there is no need to move data files. Setting up a shared drive is not too difficult and most researchers have either the knowledge to do so in their preferred operating system or institutional support.

Element 2: Git and GitHub

Once data have been stored in a single location, they must be made into a *repository*. A repository is more than a set of files; it contains also versions, changes, and logs. Because repositories contain logs and versions, they provide for a digital audit trail. For me, the easiest approach is to use a single repository for all experiments in my lab. This repository spans several folders and files. Making a repository, versioning, and logging is performed by a dedicated software application. Over the years there have been several options, but there is a dominant application: *Git*. Git is dominant because in my opinion it is the most flexible and convenient. One advantage of Git is that it interfaces seamlessly with GitHub, a public website that hosts a ridiculously large number of open-source projects. Git is very easy to find and install, and the GitHub clients, such as GitHub-for-Windows and GitHub-for-Mac include it. Git is part of the Linux operating system and no further installation is needed for Linux servers.

At the heart of the system is GitHub, the largest web host for sharing open-source projects (`www.github.com`). GitHub may be used at no cost and when used in this mode, the information is freely and publicly available. GitHub was originally designed for the development of open-source code, but is now used for a wide range of publicly available projects. Anyone can make a GitHub account and use GitHub software to create a local repository and link it to GitHub. GitHub is designed to be fairly user friendly and provides extensive help and services. Git repositories may be uploaded to GitHub. The GitHub copy is then made available on the web either through the Git system or through the web on demand. Trained IT professionals should be familiar with Git and GitHub because many useful projects are archived on these systems.

Here is a very brief step-by-step example of how to set up Git and GitHub. I take the perspective of Kirby, my dog, who has never used Git or GitHub. Kirby will be storing cute photos of himself as data.

1. The first task is to set up a repository on GitHub server. The following steps are followed.

- Kirby goes to GitHub and signs up for a free account (last option). Once the account is set up (with user name KirbyHerby), a bedazzling screen with a lot of options for exploring GitHub appears.
- To create his first repository on the server, Kirby presses the green button that says "+ New repository" on the bottom left. Figure 1A provides a screen shot of the button.
- Kirby now has to make some choices about the repository as shown in Fig. 1B. He names it "data," enters a description of the repository, makes it public, initializes it with a README and does not specify which files to ignore or a license. He then presses the green "Create repository" button on the bottom, and is given his first view of the repository. Kirby's repository is now at github.com/KirbyHerby/data, and he will bark out this URL to anyone interested. The repository contains only the README.md file at this point.

2. The next task is to set up a version of the repository on Kirby's local computer and link the local and Github version:

- Kirby downloads the GitHub application for his operating system (mac.github.com or windows.github.com), and on installation, chooses to install the command-line tools (this will be helpful subsequently in task scheduling).
- Kirby enters his GitHub username ("KirbyHerby") and password. He next has to create a local repository and link it to the one on the server. To do so, he chooses to "Add repository" and is given a choice to "Add," "Create," or "Clone." Since the repository already exists at GitHub, he presses "Clone." A list of his repositories shows up, and in this case, it is a short list of one repository, "data." Figure 1C shows the screen shot. Kirby then selects "data" and presses the bottom button "Clone repository." The repository now exists on the local computer under the folder "data." There are two, separate copies of the same repository: one on the GitHub server and one on Kirby's local computer.

3. The final task is to add files so that others may see them:

- Kirby add his data files to his local data repository as follows: In this case, being that Kirby is a dog, his data are his favorite photos. Kirby copies the photos to the files in the usual way, which for Mac-OSX is by using the Finder. Figure 1D shows the Finder window in the foreground and the GitHub client window in the background. As can be seen, Kirby has added three files, and

these show up in both applications. Kirby has no more need for the Finder and closes it to get a better view of the local repository in the GitHub client window.

- Kirby is now going to save the updated state of the local repository, which is called *committing* it. Committing a local action, and can be thought of as a snapshot of the repository at this point in time. Kirby turns his attention to the bottom part of the screen, which is shown in Fig. 1E. To commit, Kirby must add a log entry, which in this case is, "Added three great photos." The log will contain not only this message, but a description of what files were added, when, and by whom. This log message is enforced—one cannot make a commit without it. Finally, Kirby presses "Commit to master."
- Kirby now has to push his changes to the repository to the GitHub server so everyone may see them. He can do so by pressing the "sync" button.

Kirby's additions are now available to everyone at github. com/KirbyHerby/data. Moreover, as Kirby gets new photos, he can add them by copying the files into the data directory on his local computer, committing a new version of the repository with a new message, and syncing up the local with the GitHub server version. After Kirby added his first three photos, he then added a fourth one by following these steps.

There is a lot more to Git and GitHub than this. Multiple people may work on multiple parts of the same project. Git and GitHub have support for branches, tagging versions, merging files, and resolving conflicts. Help for Git and GitHub is available in the online book plainly titled, "Git Book" at http://git-scm.com/book/en/v2. The system does take some time to learn, but there is a big payoff outside of data archiving. It can be used to version much of the academic process including analysis and manuscript preparation. I find it indispensable in keeping a reliable pipeline from data collection to final manuscript submission.

Element 3: Execution and scheduling

The last step is to execute the adding of the day's data files to the Github repository. The steps are as files: 1. All new files with certain filename extensions are added to the local version of the repository. Then the local version is committed, meaning that the state of all files is logged, and the commit is timestamped and labeled. Finally, the new state is uploaded to the GitHub site. I have written these steps in the following script, which is executed nightly.

```
#!/bin/bash
git add *.dat.*
git commit -am "automatic commit"
git push
```

This script is executed nightly by a *task scheduler*. Setting up the a task scheduler is the last step. I use CRON tables on my Linux server. Task scheduling is built into Linux, Mac OSX, and Windows.[2] Trained IT personnel should be capable of writing Git scripts and automatizing their execution.

Concerns about git and github

There are important technical concerns:

File sizes Git and GitHub do not place any size restrictions on the size of files. Nonetheless, Git does not manage very large files efficiently. I have read though I cannot find the citation that Git is not recommended for files larger than 100 MB.

Repository sizes Git and GitHub do not place any restriction on the size of repositories. Nonetheless, Git slows down considerably when repositories exceed a gigabyte. If our behavioral repository becomes too large, we will start a new one. There is no limit on the number of public repositories a user may have.

Permanence of Git & GitHub It is reasonable to wonder about the permanence of Git and GitHub. Git is an open-source application, much like emacs, R, Linux, and the like. It is too widely adopted, too useful, and too beloved to go away. GitHub is a different matter. GitHub is a private company much like DropBox or Google, and it theoretically can fail. It is more likely that GitHub would include advertisements or charge a small fee rather than go away.

Curation In most curated archives, reposited materials cannot be deleted or changes. The material is immutable, and indeed, most university and society archives work this way. Files on GitHub may be changed by the uploader. Fortunately, changes are logged and older versions of the same file are saved. GitHub should not be considered a properly curated archive. Instead, it is a useful workaround until properly curated archives are reconfigured to accept incrementally added materials by automatic processes. Because GitHub changes are logged and versioned, GitHub offers the community high confidence on the integrity of data.

---

[2]For help with task scheduling, see support.apple.com/en-us/HT2488 and windows.microsoft.com/en-US/windows/schedule-task for the Windows and Mac-OSX operating systems, respectively.
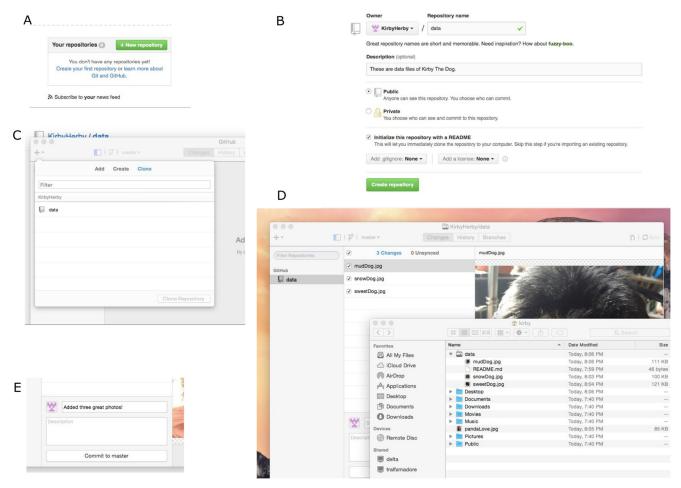
Fig. 1 Screen shots for using Git and GitHub in creating repositories. **a** After creating an account on GitHub, Kirby creates his first repository on the server. **b** Creating the repository requires a few choices, which are shown here. **c** Setting up a version of the repository on the local machine form the server is called "cloning." **d** Files may be added the usual way. **e** The local repository is updated by committing changes into a new snapshot. The new snapshot can be synced (pushed) with (to) the repository on the server

Concerns about open data and born-open data

I think the concept of open data in general and born open data in particular are needed in psychology to promote better science. The current culture of closed data and analysis serves us very poorly, and is a partial contributor to the current crisis in confidence. Unfortunately, many people I talk to do not think opening data is wise or needed. In my conversations, I have heard a number of concerns about open data, which may prevent some from adopting it. I think these concerns are misguided:

Concern: Privacy Some researchers cite the privacy of their participants as a primary concern. Privacy concerns are important and legitimate. The solution of course is to archive deidentified data. In my lab, we program the experiment-run scripts to generate files with deidentified and identified data. Only the deidentified files are added to the repository. Deidentification will often mean that not all

demographic data may be available as some of these data may identify the participant. There are some cases where data cannot ethically be made open, say those involving illegal activities which could put the participant at risk. As a rule, researchers should make open those data that may be ethically shared.

Concern: Permission IRBs must approve all data-sharing plans be they open data or not. Participants should also be informed that their data will be open. We include information in our written informed-consent materials to participants that (a) participants' data are archived at GiHub, and (b) these data are open to the public though on an anonymous basis.

Concern: Giving away something I own All of us are aware of the amount of time, care, and effort required to obtain good data, and none of us want our time, care, and effort to go unrewarded. As Morey et al. (2014) explain,
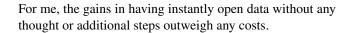
most of us think of data as something we own. If we own our data, then sharing it amounts to giving away something for free that we worked hard for. And that seems concerning to some people. My own view is that the concept of ownership is the wrong frame for scientific data. Instead, we should follow a cue from the legal community who views our data more as facts of the world that may not be eligible for copyright protection.[3] If data could be owned, there would be a tricky question establishing whether the local institution, the funding agency, or the research subjects themselves own the data. Instead, the better frame is that of data stewardship—the institution and the researcher share a joint responsibility to be good stewards of the collected data. It is in this context of stewardship rather than ownership does opening the data becomes attractive. Good stewards preserve and curate the data for the greater good of science.

Concern: The Fear of Being Scooped Some researchers support the notion of opening up their data after publication rather than before. The downside of this view, however, is that reviewers and journals neither have a method of scrutinizing the data in the peer-review process nor have any enforcement mechanism that the data will be made available after publication. Given the results of Wicherts et al. (2006), it is likely that researchers who insist on sharing after publications are far less prepared to share it at all.

The argument for post-publication sharing is that the researchers should be the first to publish their data. Publishing one's data first is necessary for a healthy science. I believe born-open data does not threaten the ability of researchers to publish first. First, scooping data is fraudulent and representing others' data as your own is a form of deceit. Few psychologists are truly fraudulent or deceptive. Second, it is exceedingly difficult to understand what archived data mean without a guiding publication. If you think it is easy, take a look at Experiment be1 discussed earlier. The column labels are provided; can you figure out what happened in the experiment? Researchers who are nervous that their column descriptors share too much can always add the descriptors after publication. The risk of being scooped is small, perhaps even microscopic. By any reasonable standard, the gain to the community of open data on submission more than outweighs the risk of any individual being scooped.

Making data open on submission is a perfectly reasonable model of data sharing. It is, however, not the same as born-open data. I prefer born-open data for my lab because of the nightly automatization. The benefits of automatization, detailed above, are critical to the reliability of my lab. I fear that without it, my data would not be made open.

For me, the gains in having instantly open data without any thought or additional steps outweigh any costs.

Concern: Professional vulnerability Perhaps a salient if not often discussed reason people are unenthusiastic about open data is a sense of risk, vulnerability, and fear. The worst-case scenario may be that someone else is going to find an indefensible mistake. As a consequence of sharing data, one might have to retract a paper, or even worse, look incompetent among colleagues. Amplifying these fears is the realization that those who take the time to look critically at born-open data do so as motivated skeptics. Moreover, there is no short-term gain for opening data.

The vulnerability and risk are evident in many of the authentic sentiments I have heard while talking about open data:

- One of my colleagues told me bluntly that she fears open data. She fears that someone will look at her data and her conclusions and find a mistake, and she will look and feel "stupid."
- One of my colleagues told me that his data are not well organized and he does not want the field to see this state of chaos. It would take him too much time and effort to get his data to the standards with which he would feel comfortable sharing.
- One of my colleagues has revealed that he doesn't trust the intentions of others. Another told me she distrusts the self-appointed replication police who are viewed in some quarters as trying to shame otherwise good researchers. I attribute these sentiments of ill will to a sense of vulnerability.

The fear of scrutiny is a deep and real one, and it is compounded for people who do not have the security of tenure. I have sympathy for people who are scared to open their data because, honestly, I am scared to open my data as well. Clearly, the practice of open data, and especially born-open data requires being comfortable with additional vulnerability and scrutiny. Yet, it is this very vulnerability that makes for better scientific practice. The very act of putting our data out there makes our pipelines more reliable and our decisions more judicious. And this gain holds even if nobody happens to scrutinize our data. This vulnerability should be harnessed as it is quite a good thing in the long run.

Summary

Born-open data has the potential to lead to a better, more self-reflexive, and more open state of psychological science. Computer technology has evolved to the point where many of us can implement it (though not without some effort). Perhaps it is incumbent of the most senior of us to

lead the effort to make data open. We have the most security, the most evolved perspective on being critiqued, and the least to fear from increase scrutiny. If we do so, then the younger people may follow. And as a result, we will all benefit.

References

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*, 407–425. doi:10.1037/a0021524

Carpenter, S. (2012). Psychology's bold initiative. *Science*, *335*, 1558–1561.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29.

Kahneman, D. J. (2012). Open Letter: A proposal to deal with questions about priming effects.

Klein, M., Van de Sompel, H., Sanderson, R., Shankar, H., Balakireva, L., Zhou, K., & Tobin, R. (2014). Scholarly context not found: One in five articles suffers from reference rot. *PLoS One*, *9*, e115253.

Krawczyk, M., & Reuben, E. (2012). (Un) available upon request: field experiment on researchers' willingness to share supplementary materials. *Accountability in Research*, *19*, 175–186.

Morey, R. D., Chambers, C. D., Lakens, D., Lewandowsky, S., Wagenmakers, E. J., & Zwaan, R. A. (2014). An agenda for open research: Incentivising open research practices through peer review. http://agendaforopenresearch.org/agenda10.pdf

Morey, R. D., Romeijn, J. W., & Rouder, J. N. (2013). The humble Bayesian: Model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, *66*, 68-75. doi:10.1111/j.2044-8317.2012.02067.x

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*, 615–631.

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*, 528–530.

Punewska, M. (2014). Scientists have a sharing problem. The Atlantic Monthly. Available from www.theatlantic.com/health/archive/2014/12/scientists-have-a-sharing-problem/383061/

Roediger, H. L. (2012). Psychology's woes and a partial cure: The value of replication. *APS Observer*, *25*.

Storm, L., Tressoldi, P. E., & Di Risio, L. (2010). Meta-analysis of free-response studies, 1992-2008: Assessing the noise reduction model in parapsychology. *Psychological Bulletin*, *136*, 471–485. doi:10.1037/a0019457

Vines, T. H. e. a. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*, 94–97.

Wicherts, J. M. (2013). Science revolves around the data. *Journal of Open Psychology Data*, *1*(1). http://openpsychologydata.metajnl.com/article/view/jopd.e1

Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish! Why not publish your data too? *Intelligence*, *40*(2), 73–76. http://www.sciencedirect.com/science/article/pii/S0160289612000050

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, *6*(11), e26828. http://www.plosone.org/annotation/listThread.action?root=19627

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. http://wicherts.socsci.uva.nl/datasharing.pdf

Yong, E. (2012). Replication studies: Bad copy. *Nature*, *485*, 298–300.