Open Minded Psychology

Wouter van den Bos, Mirjam A. Jenny
and Dirk U. Wulff

Center for Adaptive Rationality, Max Planck Institute
for Human Development, Berlin, Germany

Introduction

Psychology is a young and dynamic scientific discipline, which has a history of closely scrutinizing its own methods. For example, in the sixties, experimental psychology improved its methods after researchers became aware of the experimenter effect, that is, experimenters may inadvertently influence experimental outcomes (Kintz et al. 1965). The introduction of new technologies such as neuroimaging in the late nineties also raised several unique methodological issues (e.g. reverse inferences and double dipping: Poldrack, 2006; Kriegeskorte et al., 2009). Finally, debating and improving our statistical toolbox has always been

an integral part of the field: many psychology departments have methods departments and there are several dedicated journals (e.g. *Behavior Research Methods* since 1969). Currently, advancements of online technologies hold the potential to transform the field regarding the reporting and sharing of data.

There has been a shift from "paper only" to the digital presence of scientific journals, which has lifted the physical limits of research reports, allowing for publication of much more extensive supplementary material. Open Access journals like those of PLOS and *Frontiers* are on the rise, and open access options become the standard. Finally, online repositories and collaborative tools (e.g. openscienceframework.org) allow for effortless and free storing of data, experimental designs, analysis code, and additional information needed for successful replication or meta-analyses.

Psychology as a field has always been quick to integrate new technologies into their experimental design and measurement, such as computerized experiments and neuroimaging techniques. However, as David Johnson observed in 2001, "psychological science has largely taken a pass on optimizing knowledge production and integration through use of electronic communication" (Johnson 2001). Now almost 15 years later, with a few notable exceptions, this still largely rings true. To openly share material, over the past decades physicists, mathematicians, and computer scientists used arXiv.org; molecular biologists used the Protein Data Bank; and GenBank, geoscientists, and environmental researchers have Germany's Publishing Network for Earth and Environmental Science (PANGAEA). However, nothing of the like has been developed in psychology.

As a result, the collection of data is still surprisingly cumbersome. According to one study, around 73% of corresponding authors failed to share data from their published papers upon request (Wicherts 2013) Luckily, one of our authors has been more successful;

For our meta-analyses, we aimed to combine the results of 20 papers, all published between 2004 and 2013 (Wulff, Hertwig & Mergenthaler, in prep.). Out of those papers, 16 were first-authored by 11 different researchers outside our own research group and thus needed formal contacting to request the data. After first contact in April 2012, it took nearly three months (84 days) to either completely retrieve the requested data or have certainty over its unavailability (three datasets). A total of 68 emails were exchanged and 12 reminders needed to be sent (see **Figure 2**).
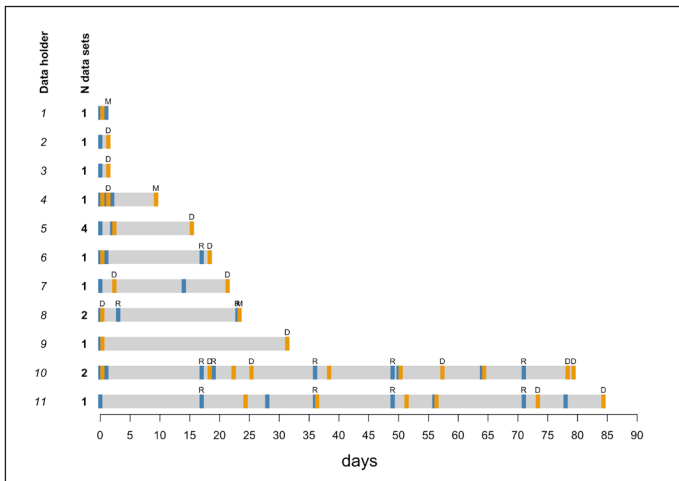

Figure 2: Correspondence timeline for retrieving a total of 16 data sets from 11 data holders.

D = Data; M = Missing (data that was eventually declared missing); R = Reminder. Blue marking indicates emails of the requester, orange of the data holder. Multiple data indicators may result from requests for multiple datasets, but also from incomplete data submission.

(Box continued from previous page)

The two main reasons for the delay seemed to be the unavailability of the data for the researcher herself and low prioritization. An exemplary response was this: "I'm quite busy for the next few days but will send this on to after that. Please remind me again by the end of the month." Like this one, most correspondences made very apparent that providing data meant a substantial amount of work for the providing researcher. This was further illustrated by three cases where data was provided in separate chunks or required later supplement due to initially incomplete data.

A second and often more bothersome obstacle arose after the data was retrieved: bringing the data into a coherent organization scheme. For (now) obvious reasons, this work remains with the requester. Usually data has not only been collected in slightly different paradigms and with different tools, they also come in different formats (e.g. long or wide). The restructuring requires a lot of manual labor, but also a significant amount of intellectual work to understand the data structure. Here, the presence and quality of accompanying data documentation took an important moderating role. In the study, the level of documentation ranged from not being there to elaborate and easily intelligible descriptions of how the data correspond to the elements of the published paper. Clearly, some of the descriptions were crafted for this instance, which pointed again to the merits of making documented data available upon publication.

Box 1: Meta-analyses: A Case Study.

he managed to receive over 85% of the requested datasets for this meta-analysis. However, collection of only 16 datasets took about three months, and this does not include the time spent on the subsequent organization of the data for analysis (see **Box 1**). Had these datasets been stored in a repository in a standardized format, their collection would probably have taken five minutes (which is approximately 25,000 times faster). Such slow and incomplete dataset collection clearly hinders academic progress. For this and other reasons, there is a growing call for increased openness in sharing data in psychology (Miguel et al. 2014; Pitt & Yang 2013; Wicherts 2013).

In this chapter we aim to make a case for the need of a common data sharing policy for psychological science, discuss what such a policy should address, and hope to make some practical suggestions along the way. First, we summarize the reasons for open data and what the advantages could be specifically for psychological science. Next, we will address in more detail what it means for data to be truly open, as well as some concerns about open data. Finally, we discuss how we could move toward a more open minded psychology.

Why Open Data?

One argument for open data that has received a lot of attention recently has been a number of cases of data fraud in science. Although it is likely that open data requirements may reduce fraudulent behavior (Simonsohn 2013), we do not think that an open data policy should be based on the motivation of exposing fraudulent behavior. Instead we deem it more successful to highlight the numerous benefits of data sharing, in general and for psychology specifically.

To start with a very straightforward benefit, data sharing leads to better data preservation. Technological advancements (or planned

obsolescence) quickly make our storage media obsolete and unusable (floppy drive anyone?), rendering the data stored on it inaccessible. In addition, scientists move locations in many stages of their careers, each time introducing the danger of the data getting lost. Of course, many researchers think this would not happen to them, but the results of published data requests do suggest that lost data is probably one of the main causes of non-compliance. Luckily, most online repositories have structured institutional funding and make use of professional servers that provide continuous backups of stored data. As such, there really is no reason for data to get lost; it can now be potentially stored forever.[1]

Crucially, when data is openly available it can be used in many ways; it can be combined with other datasets, used to address questions that were not thought of by the authors of the original studies, analyzed with novel statistical methods that were not available at the time of publication, or used as an independent replication dataset.

One very successful example of such a project is the 1000 Functional Connectomes Project.[2] This project, explicitly modeled upon the successful collaborative efforts to discover the human genome, was formed to aggregate existing resting state functional magnetic resonance imaging (R-fMRI) data from collaborating centers throughout the world.[3] The initiators of this project were

---

[1] This is a lot longer than the mere five years that is currently indicated in the publication manual of the American Psychological Association (APA manual sixth edition) as a reasonable time to keep your data, more on this below.

[2] See: http://fcon_1000.projects.nitrc.org/

[3] Imaging the brain during rest reveals large-amplitude spontaneous low-frequency (<0.1 Hz) fluctuations in the fMRI signal that are temporally correlated across functionally related areas. Referred to as functional connectivity, these correlations yield detailed maps of complex neural systems, collectively constituting an individual's "functional connectome."

able to gather and combine R-fMRI data from over 1200 volunteers collected independently at 35 centers around the world (Biswal et al. 2010). Using this large dataset, the researchers were able to establish the presence of a universal functional architecture in the brain and explore the potential impact of a range demographic variables (e.g. age, sex) on intrinsic connectivity. An additional benefit from such a collaborative effort is the size of the dataset that is created in the process. Due to high costs and limited access to facilities, studies in the cognitive neurosciences currently have rather small sample sizes (Button et al. 2013), which may result in overestimates of effect sizes and low reproducibility of results. Thus, combining efforts to create larger sample sizes would be one way to address this issue.[4] Of course, re-analysis may also entail much more straightforward secondary analyses such as those that may be raised during the review process (e.g. how about using variable X as a covariate?), which may provide readers with more insight into the impact of published results (Pitt & Tang 2013).

Finally, science is and should be a cumulative enterprise. Optimally, scientists digest the cumulated relevant literature and incorporate the extracted knowledge in designing their own new and potentially better experiment. Still, a single experimental setup is often repeated by different scientists under only mildly different conditions. In supplement to the accumulation of theoretical knowledge, such repetitions enable an interested researcher to actively cumulate existing evidence by means of combined statistical analyses, i.e. meta-analyses. Although meta-analyses can be

---

[4] It is commonly believed that one way to increase replicability is to present multiple studies. If an effect can be shown in different studies, even though each one may be underpowered, many will conclude that the effect is robust and replicable. However, Schimmack (2012) has recently shown that this reasoning is flawed.

done on group level statistics that are extracted from papers, such as effect sizes or foci of brain activity, there are several benefits to using the raw data for meta-analyses (Cooper & Pattall 2009; Salimi-Khorshidi et al. 2009). As we pointed out in our example (**Box 1**), open data would greatly facilitate meta-analyses.

Of course, data could also be used for purposes other than analysis. For instance, data can be used in courses on statistics and research methods (Whitlock 2011), as well as in the development and validation of new statistical methods (Pitt & Tang 2013).

Finally, one could argue that the results of publicly funded research should, by definition, be made publicly available. Reasoning along these lines, many funding bodies are increasing the degree to which they encourage open archiving. However, there should not be two classes of data: publicly funded open access and privately funded "hidden" datasets. Ideally, all data should be publicly available at the latest after the first study with them has been published.

What Is Open Data?

> A piece of data or content is open if anyone is free to use, reuse, and redistribute it—subject only, at most, to the requirement to attribute and/or share-alike.
>   (The Open Knowledge Foundation)

This is the short version of the Open Definition provided by the Open Knowledge Foundation.[5] Although open data sounds very straightforward, it may actually be more complicated than you think. Open data does not just mean storing your data on your personal website. For it to be open, the data also need to be usable by others. There are several criteria that should be met for data to be truly usable.

---

[5] ofkn.org; for the full length definition see: http://opendefinition.org/

First and foremost, other people need to be able to find the data; it should be discoverable. Archiving data at online repositories (see **Box 2**) significantly increases discoverability of data. These repositories will most likely be around longer than personal websites and often also allow for the storage of additional materials (e.g. code). Many of these repositories have good search functions so related datasets will be found with a single search. As an added bonus, several online repositories, like figshare, provide all datasets with a DataCite digital object identifier (DOI). As a result, these datasets can be cited using traditional citation methods and citations can be tracked. In fact, Thomson Reuters has recently launched a Data Citation Index.[6]

Second, if your data is discoverable, it should also be usable. As pointed out in our own meta-analyses case study, usability can be a matter of degree. Psychologists make use of a whole suite of different software tools to collect their data, many of which are proprietary such as MATLAB or E-Prime, or are dependent on proprietary software such as SPM or psychtoolbox. Subsequently, data is often also organized in proprietary software packages such as SPSS, SAS or Microsoft Excel. Output files from these software packages are not truly open because you first need to buy a (sometimes) expensive program to be able to read them. Currently, not many psychologists seem to be aware of this. To illustrate this, we made a random draw of two 2013 issues of the journal of the Society for Judgment and Decision Making, a journal with probably the best data sharing culture in psychology. It revealed that more than two thirds of the shared data was in proprietary format.[7] The solution here is simple, all of the software packages have the

---

[6] See: http://thomsonreuters.com/data-citation-index/

[7] Datasets of issues 1 and 2 of 2013 in order of frequency: 6 SPSS, 5 CSV, 3 Excel, 1 STATA, and 1 MATLAB.

Where to share your data

Repositories

- openfmri.org
- figshare.com
- openscienceframework.org
- psychfiledrawer.org

Data publications

- PsychFileDrawer
- *Journal of Open Psychology Data*
- Nature's *Scientific Data*

Licensing your data

When licensing your data, *JOPD* recommends any of these for licenses:

- Creative Commons Zero (CC0)
- Open Data Commons Public Domain Dedication and License (PDDL)
- Creative Commons Attribution (CC-By)
- ODC Attribution (ODC-By)

All of the above licenses carry an obligation for anyone using the data to properly attribute it. The main differences are whether this is a social requirement (CC0 and PDDL) or a legal one (CC-By and ODC-By). The less restrictive

(Box continued from previous page)

> your license, the greater the potential for reuse. In general, it is not recommended to use licenses that impose commercial or other restrictions on the use of data (for more on licensing see Chapter 3.

Box 2: Putting Your Data Online—a Practical Guide.

option to export the data to formats that are readable by every machine or operating system (e.g. CSV or TXT).

Next, for data to be usable, it must be completely clear how to read the data files. When a published paper is accompanied by open access data it may be easy to understand the content of the data file just from the header information. However, for some more complex datasets, such as neuroimaging data, this may not be the case. In this instance, it is important to make sure that others can use the data. Of course good standards for structuring complex datasets further increases usability (e.g. OpenfMRI standards for fMRI data https://openfmri.org/).

Finally, it is important to license your data when you share it to make sure it is as open as you want. When there is no license, it is not clear to what extent the data is open, and it is thus effectively unusable. Luckily, Creative Commons and the Open Knowledge Foundation made it very easy for scientists to decide how open they want their data to be. In addition, the *Journal for Open Psychology Data* (*JOPD*) set up recommendations for psychologists (see **Box 2** for more information). To summarize, these licenses make sure that people can use the data but also obligates users to properly attribute it.

Concerns: Privacy and consent

When dealing with personal data on public repositories it is of the utmost importance to protect the privacy of the participants. Of course there are already very good rules in place, but a mistake is easily made and with the development of new technologies (and, ironically, more open data) it becomes easier than ever to identify persons from just small pieces of data. For behavioral experiments in psychology it often seems enough just to replace participants' names with codes. But here is one sobering statistic from the Sweeney's Data Privacy Lab: About half of the US population (53%) are likely to be uniquely identified by only place, gender, or date of birth. This goes up to 87% when place is specified as a zip code (Sweeney 2000).[8] Date of birth and gender are of course very general measures, and place can often be derived from the university where the researchers are based. An interest in social economic status may lead researchers to store zip codes too. It is customary to report age at time of the study instead of birth date, but this illustrates the consequences when dates of birth are accidentally shared.

The increasing use of biological measures (brain, hormones, DNA) in psychology not only further increases the challenge to keep participants data anonymous but also makes anonymous data storage more pressing. For instance, when submitting brain imaging data to a public repository it is very important to extensively de-identify your images. There are three important sources of identifiable information. The most obvious is of course the filename. Less obvious are the information stored in the file headers, and the three-dimensional image of the participant's face that is

---

[8] If you live in the US we encourage you to check this out for yourself using the Sweeney's web app at http://aboutmyinfo.org/

often part of the image (see **Figure 1**). Luckily, there are several tools that help remove this data (e.g. LONI De-identification Debablet). However, de-anonymization for other types of biological data can be more difficult, maybe even impossible. Single gene
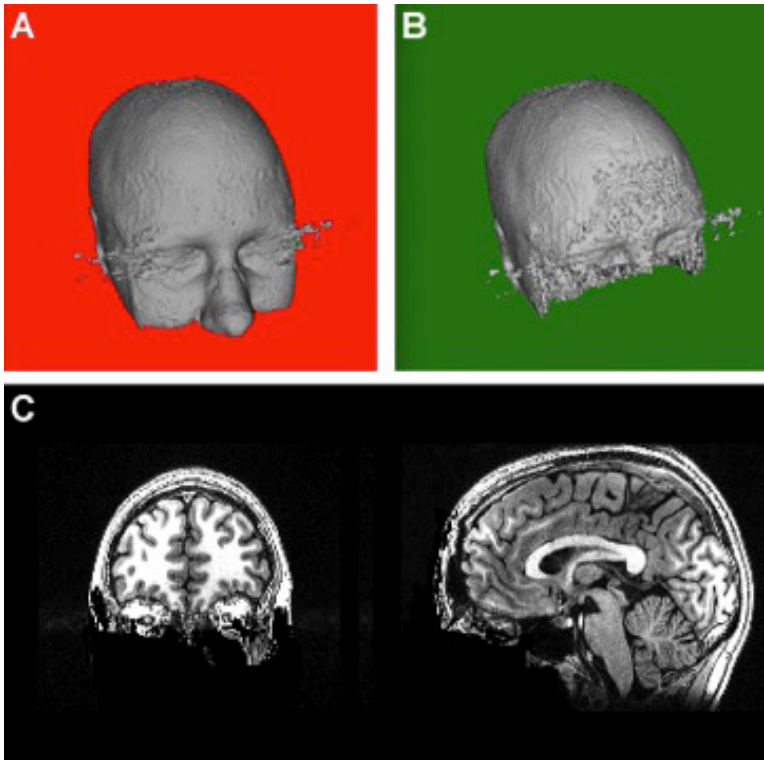

Figure 1: Structural Image before (A) and after defacing (B). It is clear that facial features are no longer recognizable but (C) essential brain data is still fully accessible. Removing the face of a structural image can be done using the Mbrin defacer package (http://www.nitrc.org/projects/mri_deface/). Alternatively, FSL brain extraction tool could be used to remove all non-brain tissue, although this might be more time intensive. In both cases is it essential to check whether deidentification is successful while brain data is still 100% intact.

mutation data is not very revealing, but genome-wide association data, or even just parts of it, can lead to identifiable data. At the moment it is not even clear whether such data can ever be shared anonymously (Hayden 2013).

Finally, it is important that both the local ethics committee or institutional review board and the participants fully agree that the data is eligible for posting on an open repository. Whereas some ethical committees allow for existing data to be posted in repositories, most will require explicit statements to this effect in the consent form. Thus, to enable data to be shared, consent forms must inform participants about the researcher's wish to post their anonymous data on public repositories.[9]

Quo Vadis? Intentions, Integration and Incentives

One of the authors recently showed that psychologists appear to be less in favor of mandatory conditions of publication than standards of good practice (Fuchs, Jenny & Fiedler 2012). Before we answer the question if good standards will suffice, let us first examine the existing standards for psychology.

In most fields, the standards are set by the journals, grant agencies and professional societies or associations. In their publication manual, the American Psychological Association (APA) encourages the open sharing of data (APA 6th ed., p.12), but has surprisingly little to say on the matter. The whole section is not even a page and no longer seems to be in line with current sentiments in the field. It mainly suggests several limiting factors on sharing, and relieves the publishing scientist of most responsibilities. First,

---

[9] Because of the potential risk with DNA data, these consent forms have become very long and include mandatory examinations to make sure all details are understood.

there is the surprisingly short period of five years that researchers are recommended to keep their data. Next, APA suggests that sharing is done only with qualified researchers (?), that all the costs should be borne by the requester, and they stress that both parties sign a written agreement that further limits data use. There is no mention of public repositories, data formatting or licensing.

In general, most journals have followed guidelines similar to those suggested by the APA. That is, they encourage but do not require data sharing.[10] However, as we mentioned, this encouragement has not resulted in data sharing on public repositories, and direct data requests are not met with great enthusiasm. Thus, on several issues, good standards are absent, and current guidance has not spurred researchers to publish or even share their data. In other words, there is room for a new data sharing policy for psychological science, one that is even more open minded. What should such a policy look like? Such a policy should of course represent the whole community. And it should also seriously address the concerns of its constituents, the scientists. Here, we briefly address issues that we think should be considered.

Concerns

One of the main concerns is that other researchers might publish research done with the data that the original researchers were also planning to work on. This worry applies to longitudinal studies especially. We view this as a valid and understandable concern and would like to propose a few conditions for the publication of data, which could resolve these worries. As long as research is still

---

[10] Exceptions being the *Journal of Cognitive Neuroscience* (2000 to 2006) and *Judgment and Decision Making*.

ongoing, authors could define embargoes or publish only part of the data. How such embargoes would be set up would have to be agreed upon in the community, however, and the embargoes should not hinder research progress. As is currently the case with patents, there could be a fixed term after which data could no longer be embargoed. Building on the existing guidelines, the rule could be that after the five-year period suggested by APA for researchers to keep their data, the data must then be shared. This embargoed period preferably should be shorter. Another concern is copyright for data that was funded privately. Here, there is probably no single solution that would always be effective, but research institutions should negotiate to make the data publicly available whenever possible.

More importantly, research in other fields (Pitt & Tang 2013) suggests that one major reason for scientists not being willing to share their data is because it is too much work. But once data sharing is the norm and researchers plan ahead, this argument does not hold. Yes, searching for old data on a bunch of hard disks, CD-ROMS or outdated laptops may be tedious, but usually researchers have easy access to their own data. Organizing the data in a self-explanatory fashion poses a bit of extra work but also further ensures data quality as the data is double-checked. Furthermore, platforms such as the Open Science Framework are extremely helpful for organizing, storing and making your data accessible. Together with a good data management plan, which should be taught at the latest in graduate school, data sharing can be made quick and simple.

Incentives—badges, data publication, citations—versus enforcement

Even if sharing data is not a large time investment for most (early career) scientists, the time can be better spent writing a paper.

So it seems that even if many scientists agree that data should be made openly available, and sharing data can be done almost automatically, the right incentives are still required to get them to actually do it.

First of all there is of course enforcement. Journals, especially high impact journals, and grant agencies could simply make open data obligatory. Following other fields, psychology could simply force open data on itself. Although enforcement is probably the most effective it is the least attractive strategy (Fuchs, Jenny & Fiedler 2012) and we therefore would like to consider some alternatives.

Recently, several journals (including flagship *Psychological Science*) adopted the badges provided by the Center for Open Science to further encourage data sharing. If authors post their data and other material online, they receive an open data badge. However, it is unclear how much improvement these badges will bring given that researchers are mostly evaluated by high impact papers and the number of citations. Of course, the journals could make the badges more powerful by, for example, ensuring increased attention or promotion for badged articles. In addition, grant agencies could amplify the effect of badges by taking them into consideration when evaluating grant applications. Another more traditional way of encouraging open data that is currently being implemented is turning data sharing into citable data publications (**Box 2**). Finally, it is worth pointing out that several studies have now shown there is a general citations advantage for papers that are accompanied by open data (Piwowar & Vision 2013).

Data storage

It must be clear who is responsible for the storage of and access to the data. The publications must indicate where

the raw data is located and how it has been made permanently accessible. It must always remain possible for the conclusions to be traced back to the original data.

(Levelt, Noort and Drenth Committees, 2012, p.58)

This quote from the report on the fraudulent psychologist Diederik Stapel highlights two important issues that need to be addressed. First, it suggests that data should be permanently accessible, which is a much more than the five years the APA currently recommends. We do agree that we should opt for much longer data retention; however, this raises the question of how long exactly (given permanently means forever, and that is a mighty long time). Second, it raises the question of who is responsible for storage and what are sustainable models for storing data for such long periods of time? Currently some of the online repositories are commercially funded whereas others are backed by universities. For instance, figshare is backed by the company Digital Science, but what happens with the data if Digital Science goes bankrupt? How should the costs of sharing data (in terms of time and money) be distributed?

Conclusion

For truly open science, not only should data be openly accessible but also the code for experiments and data analysis. This makes it easier to completely understand and replicate analyses, and prevents researchers from having to repeat the same workload by reprogramming an already existing task. As with the data, the codes should ideally be published in openly available languages such as R and Python (e.g. using PsychoPy or OpenSesame).

To make psychological data available and openly accessible, we must work toward a shift in researchers' minds. Open science is

simply more efficient science; it will speed up discovery and our understanding of the world. It is good to remind ourselves of this bigger picture when we are writing papers and grant proposals. We think the time is ripe for more open minded psychology, and hope with this chapter we contribute to the ongoing discussion and work toward a common data policy, and at the same time we have tried to point out several tools that are already available.

References

American Psychological Association 2010 *Publication Manual of the American Psychological Association* (6th ed.). Washington DC: APA.

Button, K S, Ioannidis, J P, Mokrysz, C. Nosek, B A, Flint, J, Robinson, E S and 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.

Biswal, B B, Mennes, M, Zuo, X N, Gohel, S, Kelly, C, Smith, S M, Beckmann, C F, Adelstein, J S, Buckner, R L, Colcombe, S, Dogonowski, A M, Ernst, M, Fair, D, Hampson, M, Hoptman, M J, Hyde, J S, Kiviniemi, V J, Kötter, R, Li, S J, Lin, C P, Lowe, M J, Mackay, C, Madden, D J, Madsen, K H, Margulies, D S, Mayberg, H S, McMahon, K, Monk, C S, Mostofsky, S H, Nagel, B J, et al. 2010 Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10): 4734–4739. DOI:10.1073/pnas.0911855107.

Cooper, H and Patall, E A 2009 The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychol Methods,* 14(2): 165–176. DOI: 10.1037/a0015565.

Fuchs, H M, Jenny, M and Fiedler, S 2012 Psychologists are open to change, yet wary of rules. *Perspectives on Psychological Science*, 7(6): 639–642. DOI:10.1177/1745691612459521.

Hayden, E C 2013 Privacy protections: the genome hacker. *Nature,* 497(7448): 172–174. DOI: 10.1038/497172a.

Johnson, D H 2001 Sharing data: it's time to end psychology's guild approach. *APS Observer,* 14(8). Available at https://www.psychologicalscience.org/index.php/uncategorized/sharing-data-its-time-to-end-psychologys-guild-approach.html [Last accessed 12 August 2014].

Kintz, B L, Delprato, D J, Mettee, D R, Persons, C E and Schappe, R H 1965 The experimenter effect. *Psychological Bulletin*, 63(4): 223–232.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. Retrieved from http://www.nature.com/doifinder/10.1038/nn.2303.

Levelt, Noort and Drenth Committees 2013 *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel.* Available at https://www.commissielevelt.nl/wp-content/uploads_per_blog/commissielevelt/2013/01/finalreportLevelt1.pdf [Last accessed 12 August 2014].

Miguel, E, Camerer, C, Casey, K, Cohen, J, Esterling, K M, Gerber, A, Glennerster, R, Green, D P, Humphreys, M, Imbens, G, Laitin, D, Madon, T, Nelson, L, Nosek, B A, Petersen, M, Sedlmayr, R, Simmons, J P, Simonsohn, U and van der Laan, M 2014 Promoting transparency in social science research. *Science*, 343(6166): 30–31. DOI: 0.1126/science.1245317.

Pitt, M A and Tang, Y 2013 What should be the data sharing policy of cognitive science? *Topics in Cognitive Science*, 5(1): 214–221. DOI:10.1111/tops.12006.

Piwowar, H A and Vision, T J 2013 Data reuse and the open data citation advantage. *PeerJ,* 1: e175. DOI: 10.7717/peerj.175.

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1364661305003360.

Salimi-Khorshidi, G, Smith, S M, Keltner, J R, Wager, T D and Nichols, T E 2009 Meta-analysis of neuroimaging data:

a comparison of image-based and coordinate-based pooling of studies. *NeuroImage*, 45(3): 810–823. DOI:10.1016/j.neuroimage.2008.12.039.

Schimmack, U 2012 The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17(4): 551–566. DOI:10.1037/a0029487.

Simonsohn, U 2013 Just post it: the lesson from two cases of fabricated data detected by statistics alone. *Psychological Science*, 24(10):1875–1888. DOI:10.1177/0956797613480366.

Sweeney, L 2000 *Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3*. Pittsburgh PA: Carnegie Mellon University. Available at http://impcenter.org/wp-content/uploads/2013/09/Simple-Demographics-Often-Identify-People-Uniquely.pdf   [Last accessed 12 August 2014].

Wicherts, J M 2013 Science revolves around the data. *Journal of Open Psychology Data* 1(1): e1. DOI: 10.5334/jopd.e1.

Whitlock, M C 2011 Data archiving in ecology and evolution: best practices. *Trends in Ecology and Evolution*, 26(2): 61–65. DOI:10.1016/j.tree.2010.11.006.