Truth and Transparency in Expertise Research

Rachel A. Searston[1], Matthew B. Thompson[2], Samuel G. Robson[3], Brooklyn J. Corbett[3], Gianni Ribeiro[3], Gary Edmond[4], and Jason M. Tangen[3]

[1]School of Psychology, University of Adelaide, Australia
[2]Discipline of Psychology, Murdoch University, Australia
[3]School of Psychology, University of Queensland, Australia
[4]School of Law, University of New South Wales, Australia

Correspondence: Rachel A. Searston, rachel.searston@gmail.com

Abstract

Across research areas, general issues of low statistical power, publication bias, undisclosed flexibility in data analysis, and researcher degrees of freedom, can be recipes for irreproducibility. To address the problem, a reform movement known as the "credibility revolution" emphasizes the need for greater transparency in how research is conducted. In this article, we describe a general approach to creating a culture of openness—tailored for expertise researchers—and describe how and why practices such as preregistration, open notebooks, open data, open materials, and open communication, might be applied to research on experts. We argue that adopting these practices helps to connect end-users with the entire research lifecycle and helps reconnect researchers with the process of gaining knowledge. By sharing notes about our predictions and plans along the way, we are forced to confront their merits. By documenting design and data analytic decisions ahead of time, and by sharing data and materials, we make errors and insights more discoverable. And by inviting research partners, expert practitioners, and the public into the lab, we stand the best chance of successfully translating research into practice.

Keywords

Expertise, open science, preregistration, reproducibility, cognitive psychology

Introduction

Research on the nature and development of expertise has revealed a great deal about what makes some people exceptional in their domain. Indeed, expertise studies have explored the skills of a wide range of professionals, including firefighters (Klein, Calderwood, & Clinton-Cirocco, 1985), forensic scientists (Towler et al., 2018), pianists (Krampe & Ericsson 1996), tennis players (Abernethy & Russel, 1987), chess champions (Chase & Simon, 1973), and surgeons (Norman, Eva, Brooks, & Hamstra, 2006). Outside of these domains, expertise researchers are making headway on more general questions about expertise, such as how to operationalize deliberate practice and

"effortful activities" (Ericsson, Krampe, & Tesch-Römer, 1993), how to quantify their relative contribution to expertise among countless other individual differences (Macnamara, Hambrick, & Oswald, 2014), or how to capture the limits of generalizing highly specialized skills (Sala & Gobet, 2017).

Expertise researchers have honed methodological skills that provide insights into expert performance (e.g., experimental design, data curation, visualisation, and analysis). But if rates of reproducibility in psychology—the degree to which consistent results are observed when studies are repeated—are an indication of "expertise" in methodology, then our

performance seems to be far from optimal. Only half of the cognitive psychology studies included in the reproducibility project (21 out of 42), for example, produced significant results in the direction of the original study (Open Science Collaboration, 2015). This rate was even lower for social psychology studies (14 out of 55). It is tempting to infer that the reproducibility rates of expertise research will be closer to those observed in cognitive studies, given that many fundamental questions about expertise are cognitive in nature (Campitelli & Gobet, 2010; Miller, 2003). But small and specialist populations, retrospective reports, and natural comparisons of experts and novices that lack random assignment and random sampling procedures are common to studies of expert performance (McAbee, 2018). While this combination of methodological quirks may increase fidelity, external validity, and practical application, it comes at the cost of reduced control over random error variance and reduced statistical power. Low statistical power, combined with a bias to publish positive results (the "file-drawer problem," Rosenthal, 1979), undisclosed flexibility in data analytic choices or "questionable research practices" (John, Loewenstein, & Prelec, 2012; Simmons, Nelson & Simonsohn 2011), and interpreting results as if they were expected all along (hypothesizing after the results are known or "HARKing"; Kerr, 1998) is a recipe for irreproducibility (Bishop, 2019; Nosek, Spies, & Motyl, 2012) in any field of psychology, including expertise research.

Problems of irreproducibility and mounting evidence of questionable research practices (for a list see John, Loewenstein, & Prelec, 2012) have given rise to a reform movement known as the "credibility revolution" in psychology (Vazire, 2018). At the heart of the movement lies the need for greater transparency in how research is conducted, and the proposed "open science" reforms are designed to make it easier for others to evaluate, reproduce, and use research findings (Spellman, Gilbert & Corker, 2018). For example, sharing data, code, and materials in public online repositories, preregistering research plans and predictions ahead of data collection on public registries, open peer reviewing with commentaries published alongside research articles, and open access publication models that help make research freely available to the public are a few of the practices that fall under the umbrella term of Open Science. In this article, we focus on why and how expertise researchers might adopt these practices.

Expertise researchers have already begun to recognize the benefits of open science for combating the combination of methodological challenges inherent to research on experts (see McAbee, 2018). We extend the discussion by describing a general approach to adopting open science practices tailored for expertise researchers. We start with a "primer on forensic expertise research" as an example domain before explaining how adopting openness, as a first principle, can foster collaboration between researchers and practicing experts in "open culture." We then describe how and why practices such as preregistration, open notebooks, open data, open materials, and open communication might be applied to expertise research. These practices were chosen because we have found them to be beneficial in our own research with police collaborators and forensic experts, and we share some of our experiences in adopting them throughout. While the examples used are by no means representative of research in every expert domain, we aim to show that it is possible to overcome some common barriers to openness, with several benefits.

A Primer on Forensic Expertise Research

One broad goal of expertise research is to pinpoint cognitive and perceptual processes that distinguish experts from novices. Several of our experiments have demonstrated that qualified fingerprint examiners are consistently more accurate than novices. We used a range of tasks as windows into their expertise and found that examiners were more capable than novices at recognizing fingerprints spaced in time and presented in noise (Thompson & Tangen, 2014), and they could judge whether, say, a left little

and middle fingerprint were made by the same individual or not (Searston & Tangen, 2017b). These perceptual skills also appeared to be domain-specific and developed over time with several months of on-the-job training (Searston & Tangen, 2017a, 2017b, 2017c).

Publicizing the extreme capabilities of expert performers is gratifying but revealing their limits can be a challenge. When judging, without time constraints, if two prints belonged to the same finger or not, a basic fingerprint comparison task, the performance of fingerprint examiners was impressive relative to novices, but it was not error free (Tangen, Thompson, & McCarthy, 2011; Thompson, Tangen, & McCarthy, 2014). These results contradicted widely accepted and longstanding beliefs about the infallibility of latent fingerprint examiners. As an example, when research demonstrating bias in forensic experts' decision-making was first published (e.g., Dror, Charlton, & Péron, 2005; Dror & Charlton, 2006; for a critical overview, see Searston, Tangen & Eva, 2016) the Chair of the Fingerprint Society in the UK, Martin Leadbetter, made the following remark:

> Any fingerprint examiner who comes to a decision on identification and is swayed either way in that decision making process under the influence of stories and gory images is either totally incapable of performing the noble tasks expected of him/her or is so immature he/she should seek employment at Disneyland (Leadbetter, 2007).

Those beliefs have gradually been abandoned as fingerprint examiners and researchers adopted more human-centric perspectives inspired by research results (Tangen, 2013). A similar shift occurred in diagnostic medicine after the alarming rate of misdiagnoses and mistreatments was publicly exposed (Institute of Medicine, 2000). These discoveries have contributed to the development of systems that are resilient to error, systems that make it harder for people to do something wrong and easier for them to do it right (Woods, Dekker, Cook, Johannesen, & Sarter, 2010). Drawing on examples from forensic expertise research, we aim to show that fostering a culture of openness can help to bypass many challenges that can arise when investigating and communicating the capabilities of experts, as well as their limits.

Open Culture

An open research culture embraces the Mertonian principle of "communalism" (Merton, 1973) embodied by the following aphorism: "If I have seen further it is by standing on the shoulders of giants" (for an account of the origins of this metaphor, see Merton, 1965). In short, full and open communication of scientific discoveries allows others to verify them more easily and build on them. As such, any field of scientific enquiry aiming to accumulate knowledge would benefit from a culture of openness. This imperative to be open extends not simply to a narrow scientific audience, but to all those interested or impacted by the research findings (e.g., the public, expert practitioners, and research partners).

Embracing a culture of openness by seeking input from non-academic collaborators right from the outset is one way to address some of the challenges researchers face when communicating their work to experts. While not all research on expertise involves working directly with experts or non-academic collaborators, many studies on expertise do nonetheless impact a variety of end-users outside of academia: expert practitioners, athletes, musicians, managers, coaches, trainers, organizations, governments, and members of the public. In our research with police and forensic scientists, for example, we have found that including expert practitioners in the research process from the beginning provides them with the opportunity to contribute ideas about the suitability and fairness of the performance measures, manipulations, and materials used, and the ways in which their expertise is being operationalized. This open exchange of ideas at every step has given rise to new lines of enquiry of mutual interest, new discoveries, and new ways of capturing and communicating expertise.

Preregistration

Publicly registering predictions and plans for research and analysis *before* the results are known—a practice known as "preregistration"—can guard against a particularly insidious form of hindsight bias where we researchers fool ourselves into thinking that our data analytic decisions were planned in advance and that we predicted the results all along (Fischhoff & Beyth, 1975; Munafò et al., 2017; Nosek, Ebersole, DeHaven, & Mellor, 2018; Wicherts et al., 2016). For example, in a study of more than 2,000 psychology researchers, 58% indicated that they had peeked at the results of a study before subsequently deciding whether to collect more data (John, Loewenstein, & Prelec, 2012). In addition, 35% claimed that they had reported *unexpected* findings as though they were expected when first conceiving of and specifying the hypotheses. Preregistration is a partial antidote to many of these problems as it helps researchers maintain distinctions between prediction and postdiction.

Put simply, preregistration is a commitment to an analysis plan without advanced knowledge of the outcomes (Nosek, Ebersole, DeHaven & Mellor, 2018). Such an analysis plan constrains how the data will be used to address research questions. All the hard work and decision-making is moved upfront prior to data collection: spelling out recruitment strategies, stopping rules, exclusion criteria, materials, procedures, predictions and data analysis plans before testing a single participant. An advantage of moving the bulk of the work before data collection or analysis (Gelman & Loken, 2013) is that it helps us distill fuzzy ideas and ill-defined research questions into a clear set of predictions. Only by explicitly predicting the outcome of an experiment—in our case actually predicting a number for each condition and estimating effect size—can we appreciate how closely our beliefs map on to reality. The practice of preregistration naturally extends to research that involves collaborating with experts from specialist populations. If the results of an experiment differ from what experts themselves predicted ahead of time, it is far more difficult

to explain them away or to dismiss them out of hand.

Preregistration may add work in the short term, but it can help ensure all the kinks, such as less than sufficient sample sizes or poorly framed definitions of expert performance (McAbee, 2018), are ironed out before testing the first participant. There are several online registries for recording research plans (e.g., https://osf.io, https://aspredicted.org or https://clinicaltrials.gov). New article formats that encourage preregistration are also emerging across publication outlets. The registered report format, for instance, is characterized by a two-step review process: (1) a pre-data collection review of research plans with the potential for in principle acceptance based on the merits of the research questions and proposed methodology at hand, and (2) a secondary post-data collection review to ensure the proposed plan has been adequately followed (Chambers, 2019). Some outlets have even begun incentivizing preregistration by awarding badges that signal the use of such open science practices; since their implementation, there have been marked increases in the uptake of such practices (Kidwell et al. 2016). These solutions are designed to reduce the sort of undisclosed flexibility in data analyses that can lead to false positive discoveries (Simmons, Nelson, & Simonsohn, 2011). Especially when practical applications are inspired by research findings, it is important to disclose the design and data analytic choices that may impact the decisions made by expert practitioners, research partners, and members of the public who depend on expert systems.

Open Notebooks

Not all expertise research is focused on testing predictions, and many online platforms (e.g., Open Science Framework, Figshare, GitHub) support version-controlled documentation of lab notes and research plans beyond mere hypothesis testing. We refer to this broader practice of dynamically posting updates and changes to research plans throughout the research lifecycle as "open notebooks." Open notebooks can help non-academic collaborators

and research end-users to better understand the nature of the end-to-end research process, and to raise questions or concerns early, *before* the results are communicated. This aspect of open notebooks is well-suited to research with expert populations. For instance, we can better explain the reasons for particular research design choices that may seem unnecessary or artificial to experts at first glance but are important for addressing particular research questions (e.g., the use of forced-choice tasks and controls to understand decision processes). And revisions to research plans based on early feedback from collaborators can be openly documented. Keeping an open notebook, in consequence, can facilitate genuine engagement with the research process.

We have found that there are many benefits to open notebooks in our research with forensic experts. Whenever we venture into a new domain there is often a need to create new measures of expert performance, but experts' time is invaluable and finite, and we don't want to waste it testing unsound or insensitive measures and underpowered research designs. To ensure that we use participants' time wisely, we have adopted the approach of estimating the number of trials and participants we need for our smallest effect size of interest (a principled estimate of the smallest effect you would care about based on a formalized model, prior research results, the purpose of the research; Lakens, Scheel & Isager, 2018). We then generate all our participant event sequences in advance to check that the counterbalancing and randomization of trials, and random sampling of materials, is working as planned. We program our own experiments as open-source computer applications and simulate participants' responses under different conditions. In some cases, we simulate random responses in an experiment to provide a model of chance performance. If the simulated data reflect impossible or skewed levels of performance, then we know that we have made a coding error. Another strategy is to simulate various levels of performance (e.g., selecting a correct response on 66% of the trials for novices and 84% trials for experts). Simulating participants' responses in this

way not only provides an end-to-end test of our experiments, but it also allows us to prepare analysis scripts and data visualizations ahead of time. While these general steps of planning, predicting, simulating, and pilot-testing may differ from one research program to the next, they are all components of the research process that can be publicly shared and revised as a part of a dynamic "open notebook," where each version of the research plan is recorded and stored.

Open Data

The public sharing of data in an online repository is another practice we can adopt to ensure that others can verify, extend, and build on our results (Ceci & Walker, 1983; Fecher, Friesike, Hebing, 2015). Many psychology researchers are reluctant to share data in public repositories for a variety of reasons, including the perception that it is not common practice (Houtkoop et al., 2018). Nevertheless, open data sharing appears to be on the rise across disciplines (e.g., Federer et al., 2018). Better access to published and unpublished data may help correct for bias towards publishing positive results (although such biases remain a concern even in unpublished literature searches; Ferguson & Brannick, 2012). Providing data analytic scripts (e.g., commented code, syntax, details about software and version numbers) in addition to the raw data also allows peer reviewers to reproduce the results themselves with the code provided by the authors. Open data provides a feedback mechanism to researchers by which we can evaluate how well calibrated our analytic techniques really are.

With free access to an assortment of online repositories—such as the Open Science Framework—moving towards open data practices would seem like an achievable first step. But in expertise research, the confidentiality or anonymity of expert participants is an ethical challenge when it comes to sharing raw data. Someone has to be the best and someone has to be the worst performing participant, but when dealing with professionals who make critical decisions (e.g., medical practitioners and forensic scientists) it is often tempting to focus on the active errors

that occur at the level of the frontline operator (i.e., the "bad apple" problem), rather than on problems that are latent in the system (Institute of Medicine, 2000). Since many experiments necessarily involve a small number of expert participants, it may be easy to identify individual performers in a dataset if demographic information or metadata are included in the raw data files.

It is possible to blind the experts, the investigators, or everyone to the origin of each data point by assigning each participant a unique random code that cannot be traced back to the participant. But occasionally the very measures that we use to understand expertise, such as amount of experience, can reveal participants' identities. One possible solution that we've used in such cases is to substitute the raw values (e.g., 37 years of formal experience) with rankings that are not easily traced to the individual (e.g., a rank of 2 to denote the second-most experienced expert in the sample). The feasibility of this approach will depend on the sample at hand and whether participants' identities can still be gleaned from their rankings. The issue of anonymity in extremely small datasets might be side-stepped altogether with many labs working together to recruit larger pools of expert participants from diverse populations (e.g., the Psychological Science Accelerator; Moshontz et al., 2018; McAbee, 2018). Ultimately, removing barriers to open data is imperative to building a more accurate, cumulative picture of expertise.

Open Materials

Open materials is the practice of making publicly available the components needed to reproduce the research procedure: the stimuli, measures, questionnaires and experimental software, the participant and experimenter instructions, and other instruments used to conduct the research. Open materials has many of the same virtues as open data, but the focus is on enabling others to repeat the steps taken to produce it. In other words, the public sharing of materials takes some of the guesswork out of reproducing procedures from methodological descriptions (Spellman, 2013; Spellman, Gilbert

& Corker, 2018). Interestingly, publication outlets that have awarded badges for open science practices have seen an increase in data and materials sharing, and the incentive of earning a badge as a signal to openness has also seemingly increased the completeness and usability of the datasets and materials that are made public (Kidwell et al. 2016).

Open materials is in many respects critical to error detection in science because confounds introduced by research procedures (e.g., improper counterbalancing, randomization, and blinding) can influence the reported results, and such confounds can be difficult to detect from the methodological descriptions alone, much less from the data themselves. Studies of expertise tend to rely heavily on natural stimuli (e.g., forensic science evidence, medical images, musical notation, chess configurations). We gain fidelity by using such rich naturalistic materials in our experiments, but they can also introduce artefacts that reduce generality and confound results. For example, without careful attention to proper counterbalancing or control measures, seemingly trivial features such as the dimensions of an image can produce spurious effects (Vokey et al. 2004), and these errors can only be detected by accessing the original materials.

Randomly sampling a variety of materials for each participant from large pools or repositories can alleviate concerns about artefacts and generality (Searston et al., 2019; Zech et al., 2018), but such large repositories are scarce, they are resource-intensive to generate (e.g., finding truly challenging ground-truth stimuli for experts), or restricted by copyright and other legal issues. We have encountered difficulties in accessing and sharing genuine forensic biometric materials (e.g., crime-scene fingerprints), for instance, as there are often necessary ethical and legal constraints in place to protect the identities of those involved. One solution to such barriers is to generate new stimulus sets with volunteers who consent to share their data from the outset. For example, Harvard's Personal Genome Project is a public repository of genomic, health, and trait data in which more than 5,000 volunteers have

consented to their data being used openly and freely for scientific and commercial purposes (The Harvard Personal Genome Project, n.d.). Another emerging solution that may be viable for some areas of expertise research is to use synthetic materials that have been generated using neural network models as a proxy for high fidelity stimulus sets (e.g., Google's FaceForensics++ dataset contains thousands of "deep fake" videos of human faces; Rossler et al., 2019). If similar patterns of results are obtained with the synthetic versions, we can be confident in using them and sharing them online for others to verify and build upon. These solutions are no panacea, but they go some way to resolving the tension between our goals of advancing reproducible expertise research and retaining a degree of external validity (National Academy of Sciences, Engineering, and Medicine, 2018).

Open Communication

Research on the nature of human expertise is more likely to reside in the category of "applied" or "use-inspired" research than in the category of "pure" or "basic" research (Wolfe, 2016). Expertise researchers, then, are more likely to collaborate with research end-users outside of academia. End-users include expert practitioners, professionals, the media, governments, lawmakers, and the public. In this context, collaboration and communication with research end-users is often routine. There currently exists a perverse incentive for academics to favor outlets that are closed or specialized, such as academic journals (Nosek, Spies, & Motyl, 2012). And there are few incentives to create the kind of communications that would benefit those outside academia, such as research translation papers for impact in practice, primers for an introduction to a subject, and academic reports for impact on policy. The rise of preprint repositories may help. By making research outputs available on open repositories, such as PsyArXiv, all stakeholders can easily access research that would ordinarily be locked behind a paywall.

We have found it difficult to keep our research partners abreast of our progress and,

more importantly, to provide opportunities to meaningfully collaborate in the research process. To combat this issue, we use the Open Science Framework as more than just a repository and preregistration platform. For each of our experiments, we have written a comprehensive project description that includes enough information for readers to understand the research. This description includes the usual journal-article type information, such as participants, procedure, predictions, and planned analyses, but also includes information for non-academic audiences such as rationale, project description, and even video screen captures of a participant's-eye-view of the experiment. Because these project pages are substantial pieces of work that can be allocated a digital object identifier, experts and research partners can use them as evidence to satisfy institutional performance indicators. In order to better communicate with our research partners, we produce short video trailers, depicting our research program and results, that are tailored for specific audiences, such as governments and the public. We also use social media to communicate publicly by posting photographs of conference presentations, data collection with experts, and even data analysis.

These aspects of the research process are largely undervalued at present, but we think they will be increasingly valuable in the future as they help to foster productive research partnerships outside of academia. The broad goal that we seek to achieve with this open approach to communication—which could be called radical transparency—is to invite those outside academia into every aspect of the data collection and analytic process to see how the research sausage is made.

Conclusion

Open science practices promise to connect end-users with the entire warts-and-all research process, not just the glossy end-result (Grand et al. 2012). By sharing our notes, data, materials, code, and the evolution of our thinking over time, researchers can reconnect with the process of finding things out. By making our predictions ahead of time, and preregistering them, we are

forced to confront their merits when the results are revealed. By documenting our research and data analytic plans, we open ourselves to criticism and improvement. Our errors are more discoverable and our findings are more easily built upon. By inviting research partners and the public into the lab, we give ourselves the best chance to successfully translate research into practice.

Expertise researchers stand to gain substantially by embracing open science practices. Our work resonates with people ranging from novices looking to learn a new skill to experts looking to gain an edge over highly skilled peers. Our measures and psychometric tests are used by organizations looking to select the best person for the job, and our insights about the development and optimization of expertise are used to train competent practitioners in safety- and security-critical systems. If we do not show the methodological and analytical steps that we have taken, then our scientific contributions may be lost in translation to practice or, equally problematic, misinterpreted and misapplied. Robust and reproducible findings are desirable in any field of research, but in expert contexts where lives and livelihoods are at stake—such as medicine (Norman & Eva, 2010; Carrigan et al. 2019), border security (Heyer, Semmler, & Hendrickson, 2018; Towler, Kemp & White, 2017), and forensic science (Chin, Ribeiro & Rairden, 2019; Edmond et al., 2016)—it is imperative that human systems are built on a foundation of accurate research findings. We have found that there are few downsides to being more open with our peers, partners, and the public about the end-to-end research process involved in understanding expertise.

Author Contributions

RAS was the coordinating author and created the outline for the paper. Authors who contributed substantially to the drafting of each section: Abstract (MBT), Introduction (RAS), A primer on forensic expertise (JMT & RAS), Open culture (JMT & RAS), Preregistration (RAS, SGR & GR), Open notebooks (RAS & BJC), Open data (RAS & JMT), Open materials (RAS, BJC & GR), Open communication (MBT), Conclusion (RAS). All authors contributed critical revisions and edits to each version of the manuscript.

Acknowledgments

Author's Declarations

The authors declare that they have no conflicts of interest with respect to their authorship or the publication of this article.

References

Abernethy, B., & Russel, D. G. (1987). Expert-novice differences in an applied selective attention task. *Journal of Sport Psychology*, *9*(4), 326-345. doi: 10.1123/jsp.9.4.326

Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature*, *568*(7753), 435. doi: 10.1038/d41586-019-01307-2

Campitelli, G., & Gobet, F. (2010). Herbert Simon's decision-making approach: Investigation of cognitive processes in experts. *Review of General Psychology*, *14*(4), 354-364. doi: 10.1037/a0021256

Chambers, C. What's next for Registered Reports? (2019). *Nature 573*, 187-189. doi: 10.1038/d41586-019-02674-6

Carrigan, A.J., Curby, K.M., Moerel, D, & Rich, A. (2019). Exploring the effect of context and expertise on attention: is attention shifted by information in medical images? *Attention, Perception & Psychophysics, 81*(5) 1283–1296. doi: 10.3758/s13414-019-01695-7

Chase, W. G., and Simon, H. A. (1973). *The mind's eye in chess.* In W. G. Chase (Ed.), *Visual Information Processing* (pp. 215–28). New York, NY: Academic Press.

Ceci, S. J., Walker, E. (1983). Private archives and public needs. *American Psychologist, 38*, 414–423.

Chin, J. M., Ribeiro, G., & Rairden, A. (2019). Open forensic science. *Journal of Law and the Biosciences*, advanced access publication, 1-34. doi: 10.1093/jlb/lsz009

Dror, I. E., Charlton, D., & Péron, A. E. (2005). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, 156(1), 74–78. doi: 10.1016/j.forsciint.2005.10.017

Dror, I. E., & Charlton, D. (2006). Why experts make errors. *Journal of Forensic Identification*, 56(4), 600–616.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. doi: 10.1037/0033-295X.100.3.363

Fecher, B., Friesike, S., Hebing, M. (2015). What drives academic data sharing? *PLOS ONE, 10*(2), Article e0118053. Doi: 10.1371/journal.pone.0118053

Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L. et al. (2018). Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLOS ONE,* 13(5): e0194768. doi: 10.1371/journal.pone.0194768

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17(1), 120-128. doi: 10.1037/a0024445

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance*, 13(1), 1-16. doi: 10.1016/0030-5073(75)90002-1

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem even when there is no "fishing expectation" or "p-hacking" and the research hypothesis was posited ahead of time. Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf

Grand, A., Wilkinson, C., Bultitude, K., & Winfield, A. (2012). Open science: A new "trust technology"? *Science Communication*, 34(5), 679-689. doi: 10.1177/1075547012443021

Heyer, R., Semmler, C., & Hendrickson, A. (2018). Humans and algorithms for facial recognition: The effects of candidate list length and experience on performance. *Journal of Applied Research in Memory and Cognition*, 7(4), 597-609. Doi: 10.1016/j.jarmac.2018.06.002

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V. M., Nichols, T. E., & Wagenmakers, E.-J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science, 1*(1), 70–85. https://doi.org/10.1177/2515245917751886

Institute of Medicine (2000). *To err is human: Building a safer health system*. Washington, DC: The National Academies Press. doi: 10.17226/9728

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. doi: 10.1177/0956797611430953

Kerr, R. A. (1998). Scientists ponder deep slabs, small comets, hidden oceans. *Science*, 280(5370), 1694-1695.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S. et al. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology 14*(5): e1002456. doi: 10.1371/journal.pbio.1002456

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1985). *Rapid decision making on the fire ground* (Tech. Rep. No. TR-85-46-12). Alexandria, VA: U.S. Army Research Institute.

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General*, 125(4), 331-359. doi: 10.1037/0096-3445.125.4.331

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259-269. doi: 10.1177/2515245918770963

Leadbetter, M. (2007). Letter to the Editor. Fingerprint World, 33, 231.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science*, 25(8), 1608-1618. doi: 10.1177/0956797614535810

McAbee, S. T. (2018). Statistical and methodological considerations for research on expertise: Remedies from open science and the reproducibility crisis. *Journal of Expertise*, 1(3).

Merton, R. K. (1965). On the shoulders of giants: A postscript. New York: The Free Press.

Merton, R. K. (1973). Science and the social order. In Storer, Norman W. (Ed*.), The sociology of science: Theoretical and empirical investigations* (pp. 254–266). Chicago, IL: The University of Chicago Press.

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141-144. doi: 10.1016/S1364-6613(03)00029-9

Moshontz, H., Campbell, L., Ebersole, C. R., Ijzerman, H., Urry, H. L., Forscher, P. S., … Chartier. C. R. (2018). The psychological science accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science, 1*(4), 501-515. doi: 10.1177/2515245918797607

Munafò, M., Nosek, B., Bishop, D., Button, K., Chambers, C., & Percie du Sert, N. et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1). doi: 10.1038/s41562-016-0021

National Academy of Sciences, Engineering, and Medicine (2018). Open science by design: Realizing a vision for 21st century research. Retrieved from: https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education*, *44*(1), 94-100. doi: 10.1111/j.1365-2923.2009.03507.x

Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 339-353). New York, NY, US: Cambridge University Press.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606. doi: 10.1073/pnas.1708274114

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(6), 615–631. Doi: 10.1177/1745691612459058

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 943. doi: 10.1126/science.aac4716

Robson, S. G., Corbett, B. J., McCarthy, D. J., Ribiero, G., Searston, R. A., Tangen, J. M., Thompson, M. B., (2019, July 12). How to collaborate with industry using open science [Blog post]. Retrieved from: https://cos.io/blog/how-to-collaborate-with-industry-using-open-science/

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638-641. doi: 10.1037/0033-2909.86.3.63

Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Niessner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. arXiv preprint arXiv:1901.08971.

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*(4), 671-685. doi: 10.1037/dev0000265

Searston, R. A., Tangen, J. M., & Eva, K. W. (2016). Putting bias into context: The role of familiarity in identification. *Law and Human Behavior, 40*(1), 50–64. doi: 10.1037/lhb0000154

Searston, R. A., & Tangen, J. M. (2017a). Expertise with unfamiliar objects is flexible to changes in task but not changes in class. *PLoS ONE*, *12*(6), p.e0178403. doi: 10.1371/journal.pone.0178403

Searston, R. A., & Tangen, J. M. (2017b). The style of a stranger: Identification expertise generalizes to coarser level categories. *Psychonomic Bulletin & Review*, *24*(4), 1324-1329. doi: 10.3758/s13423-016-1211-6

Searston, R. A., & Tangen, J. M. (2017c). The emergence of perceptual expertise with fingerprints over time. *Journal of Applied Research in Memory and Cognition*, *6*(4), 442-451. doi: 10.1016/j.jarmac.2017.08.006

Searston, R. A., Thompson, M. B., Vokey, J. R., French, L. A., & Tangen, J. M. (2019). How low can you go? Detecting style in extremely low resolution images. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(5), 573-584. doi: 10.1037/xhp0000628

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

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083-2101. doi: 10.3758/s13423-018-1451-8

Spellman, B. A. (2013). There is no such thing as replication but we should do it anyway: Comment on "Recommendations for increasing replicability

in psychology." *European Journal of Personality, 27*, 120–144. doi: 10.1002/per.1920

Spellman, B. A., Gilbert, E. A., & Corker, K. S. (2018). Open science. In J. Wixted & E.-J. Wagenmakers (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 5: Methodology* (4th ed., pp. 729–776). New York, NY: John Wiley.

Tangen, J. M. (2013). Identification personified. *Australian Journal of Forensic Sciences*, *45*(3), 315-322. doi: 10.1080/00450618.2013.782339

Tangen, J. M., Thompson, M. B., & McCarthy D. J. (2011). Identifying fingerprint expertise. *Psychological Science*, 22(8), 995–997. doi: 10.1177/0956797611414729

The Harvard Personal Genome Project (n.d.) About. Retrieved from: https://pgp.med.harvard.edu/about

Thompson, M. B., & Tangen, J. M. (2014). The nature of expertise in fingerprint matching: Experts can do a lot with a little. *PLoS ONE*, *9*(12), p.e114759. doi: 10.1371/journal.pone.0114759

Thompson, M. B., Tangen, J. M., & McCarthy, D. J. (2014). Human matching performance of genuine crime scene latent fingerprints. *Law and Human Behavior, 38*(1), 84–93. doi: 10.1037/lhb0000051

Towler, A., Kemp, R. I., & White, D. (2017). Unfamiliar face matching systems in applied settings. In M. Bindemann, & A. M. Megreya (Eds.), *Face processing: Systems, disorders and cultural differences,* New York: Nova Science Publishing, Inc.

Towler, A., White, D., Ballantyne, K., Searston, R. A., Matire, K. A., & Kemp, R. I. (2018). Are forensic scientists experts? *Journal of Applied Research in Memory & Cognition*, *7*(2), 199–208. doi: 10.1016/j.jarmac.2018.03.010

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411-417. doi: 10.1177/1745691617751884

Vokey, J. R., Rendall, D., Tangen, J. M., Parr, L. A., & de Waal, F. B. M. (2004). Visual kin recognition and family resemblance in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *118*(2), 194-199. doi: 10.1037/0735-7036.118.2.194

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

Wolfe, J. M. (2016). Rethinking the basic-applied dichotomy. *Cognitive Research: Principles and Implications*, *1*(1), 1. doi: 10.1186/s41235-016-0011-x

Woods, D. D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2010). *Behind human error*. Boca Raton, FL: CRC Press LLC

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., et al. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLOS Medicine 15(11): e1002683. https://doi.org/10.1371/journal.pmed.100268