DISCUSSION PUBLISHERS
Research Center · Open Access Edition

Replicability and Reproducibility in Comparative Psychology

Jeffrey R. Stevens

Discussion Research ID	DPR-2026-1225
Year	2017
Field	Psychology
DOI	10.3389/fpsyg.2017.00862
Author ORCID	0000-0003-2375-1360
Licence	CC-BY
Original source	Frontiers in Psychology
Permanent link	https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92
Publisher	Discussion Publishers discussionpublishers.com

Abstract

Psychology faces a replication crisis.The Reproducibility Project: Psychology sought to replicate the effects of 100 psychology studies.Though 97% of the original studies produced statistically significant results, only 36% of the replication studies did so (Open Science Collaboration, 2015).This inability to replicate previously published results, however, is not limited to psychology (Ioannidis, 2005).Replication projects in medicine (Prinz et al., 2011) and behavioral economics (Camerer et al., 2016) resulted in replication rates of 25 and 61%, respectively, and analyses in genetics (Munaf, 2009) and neuroscience (Button et al., 2013) question the validity of studies in those fields.Science, in general, is reckoning with challenges in one of its basic tenets: replication.Comparative psychology also faces the grand challenge of producing replicable research.Though social psychology has born the brunt of most of the critique regarding failed replications, comparative psychology suffers from some of the same problems faced by social psychology (e.g., small sample sizes).Yet, comparative psychology follows the methods of cognitive psychology by often using within-subjects designs, which may buffer it from replicability problems (Open Science Collaboration, 2015).In this Grand Challenge article, I explore the shared and unique challenges of and potential solutions for replication and reproducibility in comparative psychology.

Republished by Discussion Publishers under the original CC-BY licence. Copyright remains with the original authors; please cite the original work.

Cite as: Jeffrey R. Stevens (2017). Replicability and Reproducibility in Comparative P. Discussion Publishers, DPR-2026-1225.

frontiers in Psychology

SPECIALTY GRAND CHALLENGE
published: 26 May 2017
doi: 10.3389/fpsyg.2017.00862

Replicability and Reproducibility in Comparative Psychology

Jeffrey R. Stevens *

Department of Psychology and Center for Brain, Biology and Behavior, University of Nebraska-Lincoln, Lincoln, NE, United States

Keywords: animal research, comparative psychology, pre-registration, replication, reproducible research

Psychology faces a replication crisis. The Reproducibility Project: Psychology sought to replicate the effects of 100 psychology studies. Though 97% of the original studies produced statistically significant results, only 36% of the replication studies did so (Open Science Collaboration, 2015). This inability to replicate previously published results, however, is not limited to psychology (Ioannidis, 2005). Replication projects in medicine (Prinz et al., 2011) and behavioral economics (Camerer et al., 2016) resulted in replication rates of 25 and 61%, respectively, and analyses in genetics (Munafò, 2009) and neuroscience (Button et al., 2013) question the validity of studies in those fields. Science, in general, is reckoning with challenges in one of its basic tenets: replication.

Comparative psychology also faces the grand challenge of producing replicable research. Though social psychology has born the brunt of most of the critique regarding failed replications, comparative psychology suffers from some of the same problems faced by social psychology (e.g., small sample sizes). Yet, comparative psychology follows the methods of cognitive psychology by often using within-subjects designs, which may buffer it from replicability problems (Open Science Collaboration, 2015). In this Grand Challenge article, I explore the shared and unique challenges of and potential solutions for replication and reproducibility in comparative psychology.

OPEN ACCESS

Edited and reviewed by:
Axel Cleeremans,
Free University of Brussels, Belgium

***Correspondence:**
Jeffrey R. Stevens
jeffrey.r.stevens@gmail.com

Specialty section:
This article was submitted to Comparative Psychology, a section of the journal Frontiers in Psychology

Received: *09 January 2017*
Accepted: *10 May 2017*
Published: *26 May 2017*

Citation:
Stevens JR (2017) Replicability and Reproducibility in Comparative Psychology. Front. Psychol. 8:862. doi: 10.3389/fpsyg.2017.00862

1. REPLICABILITY AND REPRODUCIBILITY: DEFINITIONS AND CHALLENGES

Researchers often use the terms replicability and reproducability interchangeably, but it is useful to distinguish between them. *Replicability* is "re-performing the experiment and collecting new data," whereas *reproducibility* is "re-performing the same analysis with the same code using a different analyst" (Patil et al., 2016). Therefore, one can replicate a study or an effect (outcome of a study) but reproduce results (data analyses). Each of these three efforts face their own challenges.

1.1. Replicating Studies

Though science depends on replication, replication studies are rather rare due to an emphasis on novelty: journal editors and reviewers value replication studies less than original research (Neuliep and Crandall, 1990, 1993). This culture is changing with funding agencies (Collins and Tabak, 2014) and publishers (Association for Psychological Science, 2013; McNutt, 2014) adopting policies that encourage replications and reproducible research. The recent wave of replications, however, has resulted in a backlash, with replicators labeled as bullies, ill-intentioned, and unoriginal (Bartlett, 2014; Bohannon, 2014). Much of this has played out in opinion pieces, blogs, social media, and comment sections, leading some to allege a culture of "shaming" and "methodological intimidation" (Fiske, 2016). Nevertheless, replication studies are becoming more common, with some journals specifically soliciting them (e.g., *Animal Behavior and Cognition*, *Perspectives on Psychological Science*).

Discussion Publishers edition DPR-2026-1225 https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92

1.2. Replicating Effects

When studies are replicated, the outcomes do not always match the original studies' outcomes. This may result from differences in design and methods between the original and replication studies or from a false negative in the replication (Open Science Collaboration, 2015). However, it also may occur because the original study was a false positive; that is, the original result was spurious. Unfortunately, biases in how researchers decide on experimental design, data analysis, and publication can produce results that fail to replicate. At many steps in the scientific process, researchers can fall prey to confirmation bias (Wason, 1960; Nickerson, 1998) by focusing on positive confirmations of hypotheses. At the experimental design stage, researchers may develop tests that attempt to confirm rather than disconfirm hypotheses (Sohn, 1993). This typically relies on null hypothesis significance testing, which is frequently misunderstood and misapplied by researchers (Nickerson, 2000; Wagenmakers, 2007) and focuses on a null hypothesis rather than alternative hypotheses. At the data collection phase, researchers may perceive behavior in a way that aligns with their expectations rather than the actual outcomes (Marsh and Hanlon, 2007). When analyzing data, researchers may report results that confirm their hypotheses while ignoring disconfirming results. This "p-hacking" (Simmons et al., 2011; Simonsohn et al., 2014) generates an over-reporting of results with *p*-values just under 0.05 (Masicampo and Lalande, 2012) and is a particular problem for psychology (Head et al., 2015). Finally, after data are analyzed, studies with negative or disconfirming results may not get published, causing the "file drawer problem" (Rosenthal, 1979). This effect could also result in under-reporting of replication studies when they fail to find the same effects as the original studies.

1.3. Reproducing Results

"An article [...] in a scientific publication is **not** the scholarship itself, it is merely **advertising** of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures" (Buckheit and Donoho, 1995, p. 59, emphasis in the original). There are many steps between collecting data and generating statistics and figures reported in a publication. For research to be truly reproducible, researchers must open that entire process to scrutiny. Currently, this is not possible for most publications because the relevant information is not readily accessible to other scientists. For example, in a survey of 441 biomedical articles from 2000 to 2014, only one was fully reproducible (Iqbal et al., 2016). When data and the code generating analyses are unavailable, this prevents truly reproducible research.

2. UNIQUE CHALLENGES FOR COMPARATIVE PSYCHOLOGY

In addition to the general factors contributing to the replication crisis across science, working with non-human animals poses unique challenges for comparative psychology.

- *Small sample sizes*—With over 7 billion humans on the planet, many areas of psychology have a large population to draw from for research participants. Indeed, given that undergraduate students comprise most of the psychology study participants (Arnett, 2008) and that most colleges and universities have hundreds to tens of thousands of students, recruiting psychology study subjects is relatively easy. Yet, for comparative psychologists, large sample sizes can prove more challenging to acquire due to low numbers of individual animals in captivity, regulations limiting research, and the expense of maintaining colonies of animals. These small sample sizes can prove problematic, potentially resulting in spurious results (Agrillo and Petrazzini, 2012).
- *Repeated testing*—For researchers studying human psychology, colleges and universities refresh the subject pool with a cohort of new students every year. The effort, expense, and logistics of acquiring new animals for a comparative psychology lab, however, can be prohibitive. Therefore, for many labs working with long-lived species, such as parrots, corvids, and primates, researchers test the same individuals repeatedly. Repeated testing can result in previous experimental histories influencing behavioral performance, which can impact the ability of other researchers to replicate results based on these individuals.
- *Exploratory data analysis*—Having few individuals may also drive researchers to extract as much data as possible from subjects. Collecting large amounts of data and conducting extensive analysis on those data is not problematic itself. However, if these analyses are conducted without sufficient forethought and a priori predictions (Anderson et al., 2001; Wagenmakers et al., 2012), exploratory analyses can result in "data-fishing expeditions" (Bem, 2000; Wagenmakers et al., 2011) that produce false positives.
- *Species coverage*—Historically, comparative psychology has focused on a few species, primarily pigeons, mice, and rats (Beach, 1950). Though these species are still the workhorses of the discipline, comparative psychology has enjoyed a robust broadening of the pool of species studied (Shettleworth, 2009, **Figure 1**). Beach (1950) showed that, in *Journal of Comparative Psychology*, researchers studied about 14 different species a year from 1921 to 1946. From 1983 to 1987, the same journal published studies on 102 different species and, from 2010 to 2015, 144 different species[1] (**Figure 1**).

  The increase in species studied is clearly advantageous to the field because it expands the scope of our understanding across a wide range of taxa. But it also has the disadvantage of reducing the depth of coverage for each species, and depth is required for replication. In *Journal of Comparative Psychology* from 1983 to 1987, researchers published 2.3 articles per species. By 2010-2015, that number dropped to 1.8 articles per species. From 1983 to 1987, 62 species were studied in a single article during that 5-year span, whereas from 2010 to 2015,

[1]I downloaded from Web of Science citation information from all articles published in *Journal of Comparative Psychology* (*JCP*) from 1983 (when *Journal of Comparative and Physiological Psychology* separated into *JCP* and *Behavioral Neuroscience*) to 1987 ($N = 235$) and from 2010 to 2015 ($N = 254$). Based on the title and abstract, I coded the species tested in all empirical papers. Data and the R script used to analyze the data are available as supplementary materials.

Discussion Publishers edition DPR-2026-1225 https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92

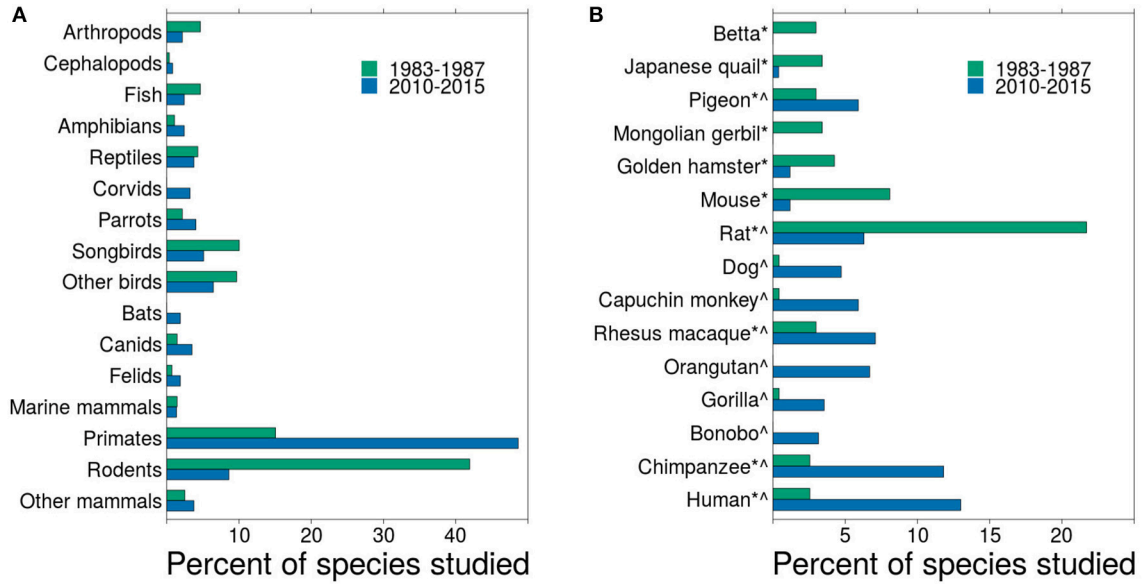


FIGURE 1 | Changes in species studied in *Journal of Comparative Psychology* from 1983 to 1987 and from 2010 to 2015. (A) shows the frequency of taxonomic groups included in empirical articles for both time periods. Because the time periods differed in the number of articles published, the percent of species included in all articles is presented. **(B)** shows a subset of the data that includes the top 10 most frequently studied species for each time period. *Top 10 most frequent from 1983 to 1987. ^Top 10 most frequent from 2010 to 2015.

103 species were studied only in a single article. Some of the species tested are quite rare, which can limit access to them. For example, Gartner et al. (2014) explored personality in clouded leopards, which have fewer than 10,000 individuals in the wild (Grassman et al., 2016) and fewer than 200 in captivity (Wildlife Institute of India, 2014). Expanding comparative psychology to a wide range of species spreads out resources, making replication less likely.

- *Substituting species*—When attempting to replicate or challenge another study's findings, comparative psychologists sometimes turn to the most convenient species to test rather than testing the species used in the original study. This is problematic because substituting a different species is not a direct replication (Schmidt, 2009; Makel et al., 2012), and it is not clear what a failure to replicate or an alternative outcome means across species. Even within a species, strains of mice and rats, for instance, vary greatly in their behavior. Thus, for comparative psychology, a direct replication requires testing the same species and/or strain to match the original study as closely as possible.

3. RESOLVING THE CRISIS

The replicability crisis in psychology has spawned a number of solutions to the problems (Wagenmakers, 2007; Frank and Saxe, 2012; Koole and Lakens, 2012; Nosek et al., 2012; Wagenmakers et al., 2012; Asendorpf et al., 2013). In addition to encouraging more direct replications, these solutions address the problems of null hypothesis significance testing, p-hacking, and reproducing analyses.

3.1. Null Hypothesis Significance Testing

- *Effect sizes*—Despite decades of warnings about the perils of null hypothesis significance testing (Rozeboom, 1960; Gigerenzer, 1998; Marewski and Olsson, 2009), psychology has been slow to move away from this tradition. However, a number of publishers in psychology have begun requiring or strongly urging authors to include effect sizes in their statistical analyses. This diverts focus from the binary notion of "significant" or "not significant" to a description of the strength of effects.
- *Bayesian inference*—Another recent trend is to abandon significance testing altogether and switch to Bayesian statistics. While significance testing yields the probability of the data given a hypothesis (Cohen, 1990), a Bayesian approach provides the probability of a hypothesis given the data, which is what researchers typically seek (Wagenmakers, 2007). A key advantage of this approach is that it offers the strength of evidence favoring one hypothesis over another.
- *Multiple hypotheses*—Rather than testing a single hypothesis against a null, researchers can make stronger inferences by developing and testing multiple hypotheses (Chamberlin, 1890; Platt, 1964). Information-theoretic approaches (Burnham and Anderson, 2010) and Bayesian inference (Wagenmakers, 2007) allow researchers to test the strength of evidence among these hypotheses.

3.2. P-Hacking

- *Labeling confirmatory and exploratory analyses*—Confirmatory data analysis tests a priori hypotheses. Analyzing data after observing the data, however, is exploratory analysis. Though exploratory analysis is not

Discussion Publishers edition DPR-2026-1225 https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92

inherently 'bad', it is disingenuous and statistically invalid to treat exploratory analyses as confirmatory analyses (Wagenmakers et al., 2011, 2012). To clarify between these types of analyses, researchers should clearly label confirmatory and exploratory analyses. Also, researchers can convert exploratory analyses to confirmatory analyses by collecting follow-up data to replicate the exploratory effects.

- *Pre-registration*—A more rigorous way for researchers to avoid p-hacking and data-fishing expeditions is to commit to specific data analyses before collecting data. Researchers can pre-register their studies at pre-registration websites (e.g., https://aspredicted.org/) by specifying in advance the research questions, variables, experimental methods, and planned analyses (Wagenmakers et al., 2012). Registered reports take this a step forward by subjecting the pre-registration to peer review. Journals that allow registered reports agree that "manuscripts that survive pre-study peer review receive an in-principle acceptance that will not be revoked based on the outcomes", though they may be rejected for other reasons (Center for Open Science, 2016).

3.3. Reproducing Analyses

- *Archiving data and analyses*—A first step toward reproducing data analysis is to archive the data (and a description of it) publicly, which allows other researchers to access the data for their own analyses. An important second step is to archive a record of the analysis itself. Many software packages allow researchers to output the scripts that generate the analyses. The statistical software package R (R Core Team, 2017) is free, publicly available software that allows researchers to save scripts of the statistical analysis. Archiving the data and R scripts makes the complete data analysis reproducible by anyone without requiring costly software licenses. Data repositories, such as Dryad (http://datadryad.org/) and Open Science Framework (https://osf.io/) archive these files.
- *Publishing workflows*—The process from developing a research question to submitting a manuscript for publication takes many steps and long periods of time, usually on the order of years. A perfectly reproducible scientific workflow would track each step of this process and make them available for others to access (Nosek et al., 2012). Websites, such as the Open Science Framework (https://osf.io/) can manage scientific workflows for projects by providing researchers a place to store literature, IRB materials, experimental materials (stimuli, software scripts), data, analysis scripts, presentations, and manuscripts. This workflow management system allows researchers to collaborate remotely and make the materials publicly available for other researchers to access.

3.4. Replicability in Comparative Psychology

Comparative psychologists can improve the rigor and replicability by following these general recommendations. However, a number of practices specific to the field will improve our scientific rigor.

- *Multi-species studies*—Though many comparative psychology studies have smaller samples sizes than is ideal, testing multiple species in a study can boost sample size. *Journal of Comparative Psychology* showed an increase in the number of species tested per article from 1.2 in 1983–1987 to 1.5 in 2010–2015. Many of these studies explore species differences in learning and cognition, but they can also act as replications across species.
- *Multi-lab collaborations*—To investigate the replication problem in human psychology, researchers replicate studies across different labs (Klein et al., 2014; Open Science Collaboration, 2015). Multi-lab collaborations are more challenging for comparative psychologists because there is a limited number of other facilities with access to the species under investigation. Nevertheless, comparative psychologists do engage in collaborations testing the same species in different facilities (e.g., Addessi et al., 2013). A more recent research strategy is to conduct the same experimental method across a broad range of species in different facilities (e.g., MacLean et al., 2014). Again, though these studies are often investigating species differences, showing similarities across species acts as a replication.
- *Accessible species*—Captive animal research facilities are increasingly under pressure due to increased costs and changing regulations and funding priorities. In the last decade, many animal research facilities have closed, and researchers have turned to more accessible species, especially dogs. Because researchers do not have to house people's pets, the costs of conducting research on dogs is much lower than maintaining an animal facility. A key advantage of dog research is that they are abundant, with about 500 million individuals worldwide (Coren, 2012). This allows ample opportunities for large sample sizes. Moreover, researchers are opening dog cognition labs all over the world, which provides the possibility of multi-lab collaborations and replications.
- *Accessible facilities*—Another alternative to maintaining animal research facilities is to leverage existing animal colonies. Zoos provide a wide variety of species to study and are available in many metropolitan areas. Though the sample sizes per species may be low, collaboration across zoos is possible. Animal sanctuaries provide another avenue for studying more exotic species, potentially with large sample sizes.

In summary, like all of psychology and science in general, comparative psychology can improve its scientific rigor by rewarding and facilitating replication, strengthening statistical methods, avoiding p-hacking, and ensuring that our methods, data, and analyses are reproducible. In addition, comparative psychologists can use field-specific strategies, such as testing multiple species, collaborating across labs, and using accessible species and facilities to improve replicability.

Discussion Publishers edition DPR-2026-1225 https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92

Frontiers in Comparative Psychology will continue to publish high-quality research exploring the psychological mechanisms underlying animal behavior (Stevens, 2010). To help meet the grand challenge of replicability and reproducibility in comparative psychology, I highly encourage authors to (1) conduct and submit replications of their own or other researchers' studies, (2) participate in cross-lab collaborations, (3) pre-register methods and data analysis, (4) use robust statistical methods, (5) clearly delineate confirmatory and exploratory analyses/results, and (6) publish data and statistical scripts with their research articles and on data repositories. Combining these solutions can ensure the validity of our science and the importance of animal research for the future.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approves it for publication.

ACKNOWLEDGMENTS

The author thanks Michael Beran for comments on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00862/full#supplementary-material

REFERENCES

Addessi, E., Paglieri, F., Beran, M. J., Evans, T. A., Macchitella, L., De Petrillo, F., et al. (2013). Delay choice versus delay maintenance: different measures of delayed gratification in capuchin monkeys (*Cebus apella*). *J. Comp. Psychol.* 127, 392–398. doi: 10.1037/a0031869

Agrillo, C., and Petrazzini, M. E. M. (2012). The importance of replication in comparative psychology: the lesson of elephant quantity judgments. *Front. Psychol.* 3:181. doi: 10.3389/fpsyg.2012.00181

Anderson, D. R., Burnham, K. P., Gould, W. R., and Cherry, S. (2001). Concerns about finding effects that are actually spurious. *Wildl. Soc. Bull.* 29, 311–316. Available online at: http://www.jstor.org/stable/3784014

Arnett, J. J. (2008). The neglected 95%: why American psychology needs to become less American. *Am. Psychol.* 63, 602–614. doi: 10.1037/0003-066X.63.7.602

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., et al. (2013). Recommendations for increasing replicability in psychology. *Eur. J. Personal.* 27, 108–119. doi: 10.1002/per.1919

Association for Psychological Science (2013). *Leading Psychological Science Journal Launches Initiative on Research Replication*. Available online at: https://www.psychologicalscience.org/news/releases/initiative-on-research-replication.html (Accessed Dec 29, 2016).

Bartlett, T. (2014). *Replication Crisis in Psychology Research Turns Ugly and Odd. The Chronicle of Higher Education*. Available online at: http://www.chronicle.com/article/Replication-Crisis-in/147301/(Accessed Dec 29, 2016).

Beach, F. A. (1950). The Snark was a Boojum. *Am. Psychol.* 5, 115–124. doi: 10.1037/h0056510

Bem, D. J. (2000). "Writing an empirical article," in *Guide to Publishing in Psychology Journals*, ed R. J. Sternberg (Cambridge: Cambridge University Press), 3–16.

Bohannon, J. (2014). Replication effort provokes praise—and "bullying" charges. *Science* 344, 788–789. doi: 10.1126/science.344.6186.788

Buckheit, J. and Donoho, D. L. (1995). "Wavelab and reproducible research," in *Wavelets and Statistics*, eds A. Antoniadis and G. Oppenheim (New York, NY: Springer-Verlag), 55–81.

Burnham, K. P. and Anderson, D. R. (2010). *Model Selection and Multimodel Inference*. New York, NY: Springer.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science* 351, 1433–1436. doi: 10.1126/science.aaf0918

Center for Open Science (2016). *Registered Reports*. Available online at: https://cos.io/rr/ (Accessed Dec 29, 2016).

Chamberlin, T. C. (1890). The method of multiple working hypotheses. *Science* 15, 92–96.

Cohen, J. (1990). Things I have learned (so far). *Am. Psychol.* 45, 1304–1312. doi: 10.1037/0003-066X.45.12.1304

Collins, F. S., and Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature* 505:612. doi: 10.1038/505612a

Coren, S. (2012). How many dogs are there in the world? *Psychology Today*. Available online at: http://www.psychologytoday.com/blog/canine-corner/201209/how-many-dogs-are-there-in-the-world (Accessed Dec 29, 2016).

Fiske, S. T. (2016). *A Call to Change Science's Culture of Shaming. APS Observer*. Available online at: https://www.psychologicalscience.org/observer/a–call–to–change–sciences–culture–of–shaming.

Frank, M. C., and Saxe, R. (2012). Teaching replication. *Perspect. Psychol. Sci.* 7, 600–604. doi: 10.1177/1745691612460686

Gartner, M. C., Powell, D. M., and Weiss, A. (2014). Personality structure in the domestic cat (*Felis silvestris catus*), Scottish wildcat (*Felis silvestris grampia*), clouded leopard (*Neofelis nebulosa*), snow leopard (*Panthera uncia*), and African lion (*Panthera leo*): a comparative study. *J. Comp. Psychol.* 128, 414–426. doi: 10.1037/a0037104

Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behav. Brain Sci.* 21, 199–200. doi: 10.1017/S0140525X98281167

Grassman, L., Lynam, A., Mohamad, S., Duckworth, J., Bora, J., Wilcox, D., et al. (2016). *Neofelis nebulosa*. The IUCN Red List of Threatened Species 2016, e.T14519A97215090. doi: 10.2305/IUCN.UK.2016-1.RLTS.T14519A97215090.en

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., and Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biol.* 13:e1002106. doi: 10.1371/journal.pbio.1002106

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., and Ioannidis, J. P. (2016). Reproducible research practices and transparency across the biomedical literature. *PLoS Biol.* 14:e1002333. doi: 10.1371/journal.pbio.1002333

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, V., Bernstein, M. J., et al. (2014). Investigating variation in replicability. *Soc. Psychol.* 45, 142–152. doi: 10.1027/1864-9335/a000178

Koole, S. L. and Lakens, D. (2012). Rewarding replications: a sure and simple way to improve psychological science. *Perspect. Psychol. Sci.* 7, 608–614. doi: 10.1177/1745691612462586

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., et al. (2014). The evolution of self-control. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2140–E2148. doi: 10.1073/pnas.1323533111

Makel, M. C., Plucker, J. A., and Hegarty, B. (2012). Replications in psychology research: how often do they really occur? *Perspect. Psychol. Sci.* 7, 537–542. doi: 10.1177/1745691612460688

Discussion Publishers edition DPR-2026-1225 https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92

Marewski, J. N. and Olsson, H. (2009). Beyond the null ritual: formal modeling of psychological processes. *J. Psychol.* 217, 49–60. doi: 10.1027/0044-3409.217.1.49

Marsh, D. M., and Hanlon, T. J. (2007). Seeing what we want to see: confirmation bias in animal behavior research. *Ethology* 113, 1089–1098. doi: 10.1111/j.1439-0310.2007.01406.x

Masicampo, E. J., and Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Q. J. Exp. Psychol.* 65, 2271–2279. doi: 10.1080/17470218.2012.711335

McNutt, M. (2014). Journals unite for reproducibility. *Science* 346, 679–679. doi: 10.1126/science.aaa1724

Munafò, M. R. (2009). Reliability and replicability of genetic association studies. *Addiction* 104, 1439–1440. doi: 10.1111/j.1360-0443.2009.02662.x

Neuliep, J. W., and Crandall, R. (1990). Editorial bias against replication research. *J. Soc. Behav. Personal.* 5, 85–90.

Neuliep, J. W., and Crandall, R. (1993). Reviewer bias against replication research. *J. Soc. Behav. Personal.* 8, 21–29.

Nickerson, R. S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* 2, 175–220. doi: 10.1037/1089-2680.2.2.175

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychol. Methods* 5, 241–301. doi: 10.1037/1082-989X.5.2.241

Nosek, B. A., Spies, J. R., and Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* 7, 615–631. doi: 10.1177/1745691612459058

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

Patil, P., Peng, R. D., and Leek, J. (2016). A statistical definition for reproducibility and replicability. *bioRxiv*. doi: 10.1101/066803

Platt, J. R. (1964). Strong inference. *Science* 146, 347–353. doi: 10.1126/science.146.3642.347

Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nat. Rev. Drug Discov.* 10, 712–712. doi: 10.1038/nrd3439-c1

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

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

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychol. Bull.* 57, 416–428. doi: 10.1037/h0042040

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Rev. Gen. Psychol.* 13, 90–100. doi: 10.1037/a0015108

Shettleworth, S. J. (2009). The evolution of comparative cognition: is the snark still a boojum? *Behav. Proc.* 80, 210–217. doi: 10.1016/j.beproc.2008.09.001

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-curve: a key to the file-drawer. *J. Exp. Psychol. Gen.* 143, 534–547. doi: 10.1037/a0033242

Sohn, D. (1993). Psychology of the scientist: LXVI. The idiot savants have taken over the psychology labs! or why in science using the rejection of the null hypothesis as the basis for affirming the research hypothesis is unwarranted. *Psychol. Rep.* 73, 1167–1175. doi: 10.2466/pr0.1993.73.3f.1167

Stevens, J. R. (2010). The challenges of understanding animal minds. *Front. Psychol.* 1:203. doi: 10.3389/fpsyg.2010.00203

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., and van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *J. Personal. Soc. Psychol.* 100, 426–432. doi: 10.1037/a0022790

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., and Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspect. Psychol. Sci.* 7, 632–638. doi: 10.1177/1745691612463078

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Q. J. Exp. Psychol.* 12, 129–140. doi: 10.1080/17470216008416717

Wildlife Institute of India (2014). *National Studbook Clouded Leopard (Neofelis nebulosa)*. New Delhi: Wildlife Institute of India, Dehradun and Central Zoo Authority.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Stevens. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Discussion Publishers edition DPR-2026-1225 https://discussionpublishers.com/research/141135b3-19ce-405a-860c-7c50f0829e92