The Downtown Review

Volume 3 | Issue 2 Article 3

December 2016

Unreplicable: The Unscientific Nature of Science Journals

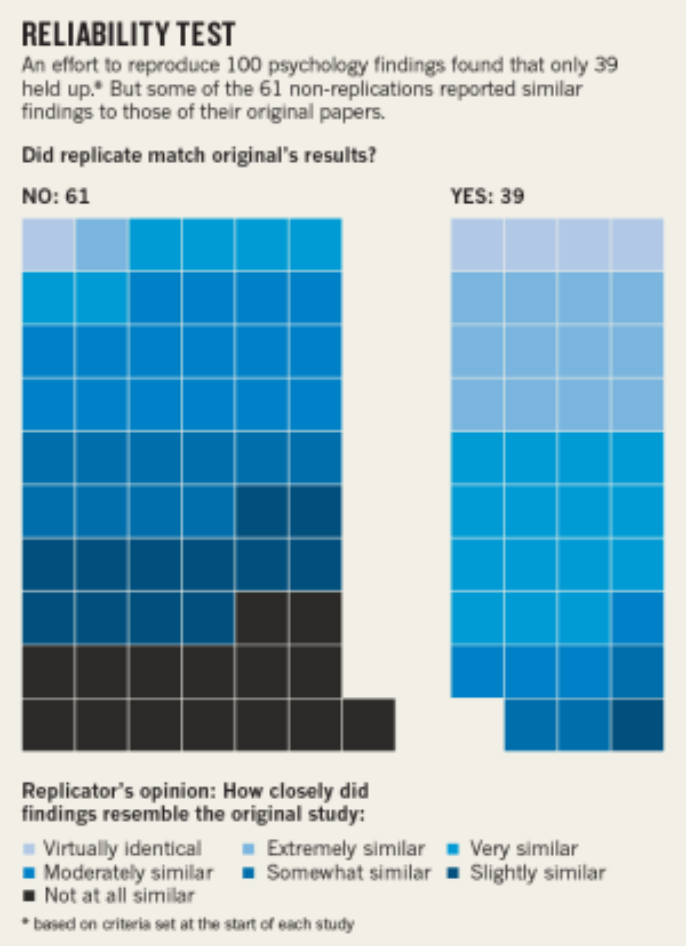
Ernest M. Oleksy
Cleveland State University

In science, as in law, there is the spirit of the subject and then the letter of the subject. In an ideal world, all scientific research would be ethical and thorough, innovative and veracious, well-run and well-replicated. However, science, a creation of the Anthropocene, cannot afford to remain pure; it must engage with and abide by the strictures of the human effect. Scientific journals are the letter of science, which play a vital role in the sustainment and development of the discipline by providing a niche for breakthroughs within college professor and R&D offices. Though these professional roles are amongst society's thought-leaders and rationalists, dominant trends uncover that they are susceptible to "the tyranny of the new" (Caruso, 1998, p. 24). The value of scientific progress is undebatable, but meta-research suggests that more emphasis needs to be placed on evaluating past publications.

Nosek's Research

In 2008, a research team led by Brian Nosek strove to recreate the significantly positive findings of numerous psychological experiments. Nosek's study failed to replicate the results from most of the 2008 studies that he and his team attempted to reproduce. In the discussion of his article, Nosek confesses that none of the indicators that his study used to entail reproducibility success did an adequate job. The startling illumination that his analysis made, however, was that his research led him to believe that most replications lead to less significant positive results than the original trials, as ascertained by metrics like P-values. These replications were run with materials from the original trials and overseen by the original authors to ensure high fidelity (Nosek, 2015, p. 1), yet the replicators only had a success rate of 36% in attaining

significant outcomes, compared to the 97% rate of the original runs. Nosek's graphic representation of these findings is as follows:



These rates were determined from a sample size of 100 for the replications and of 98 for the original runs, so the small sample size bias is not attributable to these measures (Baker, 2015, p.1).

The Nature and Consequences of Unpublished Findings

Perhaps the most statistically-evident reason as to why replications are valuable is the disparity between the rate of significant results quantified in the original, published studies

(97%) and the rate of significant results quantified in well-run replications (36%) (Nosek, 2015, p.1). This comparison implies that the ubiquity of successfully-run experiments published in academic journals is not generalizable to the population of trial runs of the greater, unpublished literature. There are outlets for unpublished work to make its way onto the internet, like PsychFileDrawer, but the lack of prestige and career enhancement often makes academics wary of utilizing such online-compendiums (Yong, 2012, p. 1).

Though the impetus for journals to publish novel, fascinating results may just be an axiom of consumerism, this endemic of proliferating significant results seems to have been exacerbated in recent history. University of Edinburgh social scientist Daniele Fanelli ran a study that discovered that published, positive results have increased over 22% from 1990 to 2007 (Fanelli, 2012, p. 902). It is unlikely that this is due to researchers becoming better at experimental design, but rather that novelty has become more coveted by journals.

The premium that academia places on novelty also results in a file-drawer dilemma, where unpublished replications or research that fails to attain positive results gets shoved into professors' offices, never to be released to the general public (Yong, 2012, p.1). The obvious issue that arises is the existence of a bevy of information that may never be known to the greater scientific community. Even though this data failed to demonstrate significant correlation and causality, the failed runs provide insight into confounding variables, extraneous factors, or experimental design that can be amended and refined in replications that may lead to ground-breaking findings. However, studies cannot be replicated if they are never released, thus the file-drawer dilemma eliminates the immense potential for progress that replications hold.

Furthermore, replications can validate and falsify past research. Because of a lack of interest in replications, researchers like Diederik Stapel of Tillburg University was able to get

away with forging studies that were never actually run. His research, which correlated orderliness with degrees of racism, had major, erroneous implications for how academia and the laity rationalized racial tensions (Stapel & Lindenberg, 2011, Abstract). Due to Stapel's established prestige and a general disinterest in replications, his fallacious claims were allowed to stand until 2011 (Yong, 2012, p. 1). Replications can also affirm or debunk fantastic findings that only resulted due to a combination of luck and shrewd design (Baker, 2015, p.1) – for instance, a satirical study that showed listening to the Beatles' song *When I'm Sixty-Four* caused listeners to become 1.5 years younger (Simmons & Simonsohn, 2011, pp. 1360-1361).

The Role of the Consumer

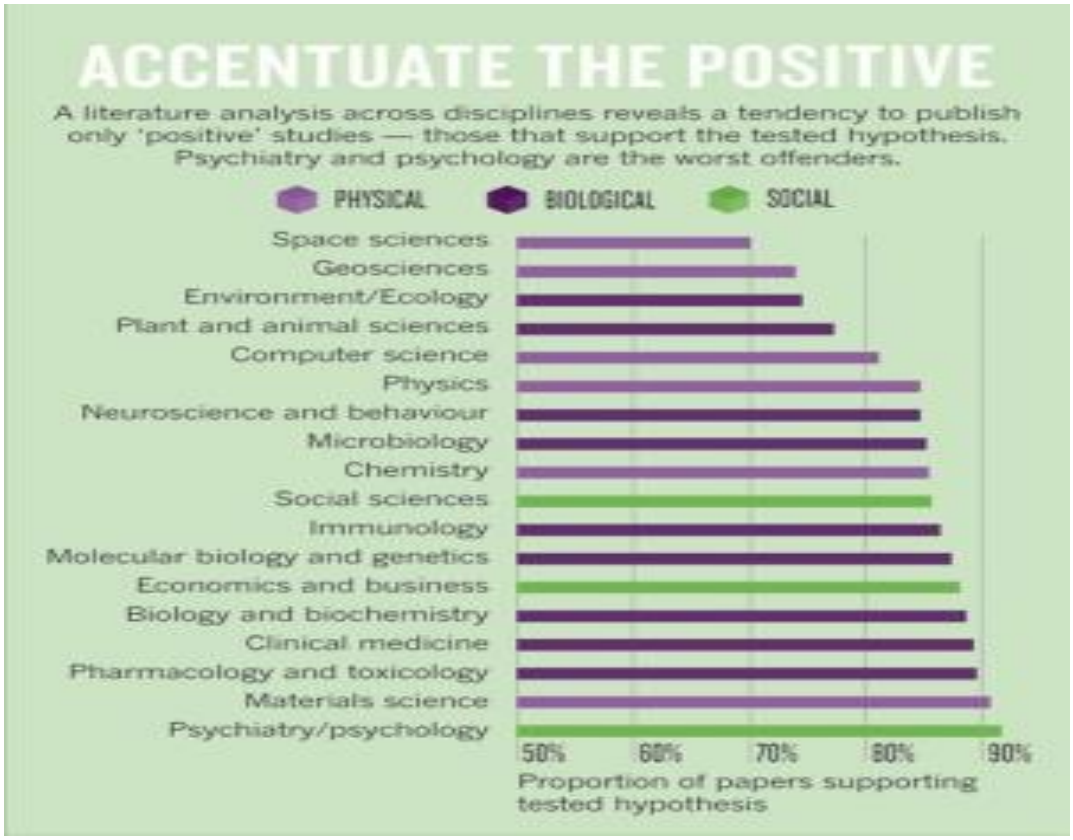
Though journals have been much maligned throughout this paper, it is important to note that their hesitance towards publishing replications is largely due to giving their consumers what they want. If people were genuinely interested in reading about replications of old studies and were willing to pay money for them, journals would respond to the demand with a ratcheted supply. However, what academia wants is sexy results and up-to-date information on the newest innovations, so publication bias favors novelty (Easterbrook et al., 1991, p. 868).

People love new discoveries, and when these new discoveries are fortified by statistical analysis, it is only natural that people will rationalize them to be true *a posteriori*. As with the case of forgers like Stapel, Ruggiero, and Hauser, research that seems to make common sense may be more readily accepted (Bornstein, 1989, p. 270-272). This calls to mind human errors, like the hindsight bias, as reasons as to why people believe duplicitous data and pseudo-sciences (Roese, 2012, p. 1). For example, if the folk rationalizes that people who are messy are

inherently obtuse and lack consideration for others, and that racists are also obtuse and lacking in consideration, individuals are capable of believing a syllogistic fallacy that disorderliness and racism are, at the very least, positively correlated.

The Role of the Experts

Practitioners are often guilty of wanting to be up to date with the newest findings in their fields of expertise in order to maintain a knowledgeable *ethos*. Misguided, this preference for novelty leads to the normalization of pseudo-scientific practices like humor-balancing being adopted as official techniques. This can become potentially deleterious to modern-day patients when observing that pharmacology has the third most positive-resulting research, as shown by Fanelli in her graph:



This finding implies that there may be a recklessly loose standard in the DEA for legitimizing drugs that doctors can prescribe for patients, or that drugs might be speciously authorized for niche conditions, like testing amphetamines on schizophrenics (Nolte et al., 2004, p. 6-11). It should be noted that this discussion on pharmacology was largely hypothesis and was not irrefutably proven within any sources used in this research.

Journals publish not only for pure scientists, but also for these applied scientists. As such, the publication bias leans towards releasing research that may be rooted in pseudo-science, like phrenology in the 19th century or Bem's psi in the 20th century (Gross, 1990, p. 1645-1646). It should also be noted that when social psychologists challenged Bem's findings with numerous negative-resulting replications, large-scale academic journals like *JPSP*, *Science*, and *Psychological Science* rejected them just because they were replications and not desired by their readership (Yong, 2012, p. 1).

Bem's Research and its Circumstances

As formerly mentioned, Bem's moonshot conclusions caused overwhelming skepticism in the psychological community. Three research teams rose to the challenge of conducting replications of Bem's claims of trans-temporal, morphic field communication, each incapable of producing the original results (Yong, 2012, p. 1). This meant that out of the four known attempts to demonstrate the principles of psi, only Bem's study resulted in positive results, possibly due to how Bem designed his experiment or analyzed his data (Bem, 2011, p. 411-416). The replication teams reached out to major journals in hopes that their research would be published and consumed by the scientific community, effectively ending the popular mania for psi.

Unfortunately, though major journals were willing to expeditiously publish sexy, unfathomable findings, they all had policies against straight replications. It should be noted that Bem was a reviewer for these journals (Yong, 2012, p. 1).

The psi replications ultimately found refuge in the journal *PLoS ONE*, a journal that accepts all types of properly-conducted research, though they did not make anywhere near the splash that Bem did. Bem would go on to claim that he does not take issue with all replications, just poorly conducted ones that endanger important findings of original research (Yong, 2012, p. 1). Yong also points out that negative-resulting replications of popular studies struggling to get published is not unprecedented. Brussel's cognitive psychologist Stephane Dargh also had to publish in *PLoS ONE* when he replicated a priming study concerning age-related words and walking speed.

Conclusions and Discussion

The scientific community, particularly psychologists, must be candid on the pervasiveness of irreproducibility. Though the goal of any experiment is to eliminate as much human interference as possible, forces like supply-and-demand, publication prestige, and the popularity of the unique and the niche may mar research. The ideal of any scientific theory is to be well-replicated so its generalizability is obvious. In order to make this ideal a reality, academia must be intellectually honest about this problem and work towards resolution. Ironically, unreplicability's greatest culprit, psychology, is the discipline most suited to work towards its amelioration. Perhaps motivation psychologists can conduct research on how to convince researchers to produce more replications. Solutions that mandate reproductions of

researchers should be avoided since the potential fallout and bad publicity would cause more problems than were solved. Nonetheless, expediently determined conclusions lead to the adoption of unproven theories, like in the cases of Stapel and Bem. All development begins with research, and if measures are not taken to maximize the fidelity and veracity of experiments, science will stagnate.

References

Baker, M. (2015). "Over Half of Psychology Studies Fail Reproducibility Test." *Nature*. doi:10.1038/nature.2015.18248.

Bem, D. J., & Sheehan, D. P. (2011). "Experimental Evidence for Anomalous Retroactive Influences on Human Cognition and Affect." *Journal of Personality and Social Psychology*, 100(3), 407-425. doi:10.1063/1.3663724

Bornstein, R. F. (1989). "Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987." *Psychological Bulletin*, 106(2), 265-289. doi:10.1037//0033-2909.106.2.265

Caruso, A. (1998). "The Tyranny of the New." *Blueprint*, 150, 24-25. Retrieved May 4, 2017, from http://www.carusostjohn.com/text/the-tyranny-of-the-new/

Easterbrook, P. J., Gopalan, R., Berlin, J. A., & Matthews, D. R. (1991). "Publication Bias in Clinical Research." *The Lancet*, 337(8746), 867-872. doi:10.1016/0140-6736(91)90201-y

Fanelli, D. (2011). "Negative Results are Disappearing from Most Disciplines and Countries." *Scientometrics*, 90(3), 891-904. doi: 10.1007/s11192-011-0494-7

Gross, C. G. (1990). "Phrenology." Princeton University, 1645-1646. Retrieved May 4, 2017, from https://www.princeton.edu/~cggross/ency_neuro_phren.pdf.

Nolte, S., Wong, D., Latchford, G., Boyle, O., & Anaenugwu, A. (2004). "Amphetamines for Schizophrenia." *Cochrane Database of Systematic Reviews*, (3), 1-46. doi:10.1002/14651858.cd004964

Nosek, B. (2015). "Estimating the Reproducibility of Psychological Science." *Science*, 349(6251), 943-Aac4716-8.

Roese, N. J. (2012). "'I Knew It All Along…Didn't I?' – Understanding Hindsight Bias." *Association for Psychological Science*. Retrieved May 4, 2017, from https://www.psychologicalscience.org/news/releases/i-knew-it-all-along-didnt-i-understanding-hindsight-bias.html#.WQzsf1XyuM8

Simmons, J., Nelson, L., & Simonsohn, U. (2011). "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *PsycEXTRA Dataset*, 22(11), 1359-1366. doi: 10.1037/e519702015-014

Stapel, D. A., & Lindenberg, S. (2011). "Coping with Chaos: How Disordered Contexts Promote Stereotyping and Discrimination." *Science*, 332(6026), 251-253. doi:10.1126/science.1201068

Yong, E. (2012). "Replication Studies: Bad Copy." *Nature*, 485, 298-300. doi: 10.1038/485298a