

The Science of Technology and Human Behavior: Standards, Old and New

Malte Elson

Ruhr University Bochum

Andrew K. Przybylski

University of Oxford

The Present

Technology and Human Behavior – a title that could not be more generic for a Special Issue. On its face, the phenomena examined in this issue are not distinct from others published in the Journal of Media Psychology (*JMP*): Can immersive technology be used to promote pro-environmental behaviors? (Soliman, Peetz, & Davydenko); How do subtitles and complexity of MOOC-type videos impact learning outcomes? (van der Zee, Admiraal, Paas, Saab, & Giesbers); Does cooperative video game play foster prosocial behavior? (Breuer, Velez, Bowman, Wulf, & Bente); What is the role of video game use in the unique risk environment of college students? (Holz Ivory, Ivory, & Lanier); Do interactive narratives have the potential to advocate social change? (Steinemann, Iten, Opwis, Forde, Frasseck, & Mekler).

What makes this issue *special* is not a thematic focus, but the nature of the scientific approach to hypothesis testing: It is explicitly confirmatory. All five studies are registered reports which are reviewed in two phases: First, the theoretical background, hypotheses, methods, and analysis plans of a study are peer-reviewed before the data are collected. If they are evaluated as sound, the study receives an “in-principle” acceptance, and researchers proceed to conduct it (taking potential changes or additions suggested by the reviewers into consideration). Consequently, the data collected can be used as a true (dis-)confirmatory hypothesis test. In a second step, the soundness of the analyses and discussion section are reviewed, but the publication decision is not contingent on the outcome of the study (see our call for papers; Elson, Przybylski, and Krämer, 2015). All additional, nonpreregistered analyses conducted are clearly labelled as exploratory and serve to discover alternative explanations or generate new hypotheses.

Further, the authors were required to provide a sampling plan designed to achieve at least 80% statistical power (or comparable criterion for Bayesian analysis strategies) for all of their confirmatory hypothesis tests, and to make all materials, data, and analysis scripts freely available on the Open Science Framework (OSF) at <https://osf.io/5cvkr/>. We believe that making these materials available to anyone increases the value of the research

as it allows others to reproduce analyses, replicate the studies, or build on and extend the empirical foundation. As such, the five studies represent the first in *JMP* that employ these new practices. It is our hope that these contributions will serve as an inspiration and model for other media researchers, and encourage scientists studying media to preregister designs and share their data and materials openly.

All research proposals were reviewed by content experts from within the field and additional outside experts in methodology and statistics. Their reviews, too, are available on the OSF, and we deeply appreciate their contributions to meliorate each individual research report and their commitment to open and reproducible science: Marko Bachl, Chris Chambers, Julia Erdmann, Pete Etchells, Alexander Etz, Karin Fikkers, Jesse Fox, Chris Hartgerink, Moritz Heene, Joe Hilgard, Markus Huff, Rey Junco, Daniël Lakens, Benny Liebold, Patrick Markey, Jörg Matthes, Candice Morey, Richard Morey, Michèle Nuijten, Elizabeth Page-Gould, Daniel Pietschmann, Michael Scharkow, Felix Schönbrodt, Cary Stothart, Morgan Tear, Netta Weinstein, and additional reviewers who would like to remain anonymous.

Finally, we would like to extend our sincerest gratitude to *JMP*'s editor-in-chief Nicole Krämer and editorial assistant German Neubaum for their support and guidance from the conception to the publication of this issue.

The Past

Concerns have been raised about the integrity of the empirical foundation of psychological science, such as the average statistical power and publication bias (Schimmack, 2012), availability of data (Wicherts, Borsboom, Kats, & Molenaar, 2006), and the rate of statistical reporting errors (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2015). Currently, there is little information to which extent these issues also exist within the media psychology literature. Therefore, to provide a first prevalence estimate, and to illustrate how some of the practices adopted for this special issue can help reducing these problems, we surveyed research designs, availability of data, errors in the reporting of statistical analyses, and statistical power of studies published in the traditional format in *JMP*. We analyzed the research published in *JMP* between volume 20/1, when it became an English-language publication, and volume 28/2 (the most recent issue when this analysis was planned). The raw data, analysis code, and code book are freely available at <https://osf.io/5cvkr/>.

Sample of Publications

Publications in *JMP* represent a rich range of empirical approaches. Of the $N = 146$ original research articles¹ identified, $n_{\text{emp}} = 131$ (89.7%) report data from at least one empirical study (147 studies in total²). Of those, more than half are experiments (54.4%) or quasi-experiments (8.8%), followed by cross-sectional surveys (23.8%) and longitudinal

¹Editorials, calls for papers, volume information tables, meeting calendars, and other announcements were excluded.

²Pilot studies were excluded.

studies (7.5%). The rest are content analyses, observational studies, or interview studies (5.4%).

Availability of Data and Materials

Recently, a number of open science initiatives including the Transparency and Openness Promotion Guidelines, the Peer Reviewers' Openness Initiative, and the Commitment to Research Transparency³ have been successful in raising awareness of the benefits of open science and increasing the rate of public sharing of datasets (Kidwell et al., 2016). Historically the availability of research data in psychology has been poor (Wicherts et al., 2006). Our sample of *JMP* publications suggests that media psychology is no exception to this, as we were not able to identify a single publication reporting a link to research data in a public repository or the journal's supplementary materials.⁴

Statistical Reporting Errors

Most conclusions in empirical media psychology, and psychology overall, are based on Null Hypothesis Significance Tests (NHSTs). Therefore, it is important that all statistical parameters and NHST results be reported accurately. However, a recent study by Nuijten et al. (2015) indicates a high rate of reporting errors in psychological research reports. The consequences of such inconsistencies are potentially serious as the analyses reported and conclusions drawn may not be supported by the data. Similar concerns have been voiced for published empirical studies in communication research (Vermeulen et al., 2015).

To make sure such inconsistencies were avoided for our special issue, we validated all accepted research reports with *statcheck* (version 1.2.2; Epskamp and Nuijten, 2015), a package for the statistical programming language R (R Core Team, 2016) that works like a spellchecker for NHSTs by automatically extracting reported statistics from documents and recomputing⁵ *p*-values.

For our own analyses, we downloaded all $n_{\text{emp}} = 131$ *JMP* publications as HTML files⁶ and scanned them with *statcheck* to obtain an estimate for the reporting error rate in *JMP*. *Statcheck* extracted a total of $K = 1036$ NHSTs⁷ reported in $n_{\text{nhst}} = 98$ articles. Initially, 134 tests were flagged as *inconsistent* (i.e., reported test statistics and degrees of freedom do not match reported *p*-values), of which 27 were *grossly inconsistent* (the reported *p*-value is $< .05$ while the recomputed *p*-value is $> .05$, or vice-versa). For one paper, a correction had

³<https://cos.io/top/>; <https://opennessinitiative.org>; <http://www.researchtransparency.org>

⁴It is, of course, entirely possible that some authors have made their data publicly available without clarifying this in the publication.

⁵*p*-values are recomputed from the reported test statistics and degrees of freedom. Thus, for the purpose of recomputation, it is assumed that test statistics and degrees of freedom are correctly reported, and that any inconsistency is caused by errors in the reporting of *p*-values. The actual inconsistencies, however, can just as well be caused by errors in the reporting of test statistics and/or degrees of freedom.

⁶Due to copyright restrictions, these are not available at <https://osf.io/5cvkr/>, but will be shared on request.

⁷Note that *statcheck* might not extract NHSTs from figures, tables, supplementary materials, or when their reporting style deviates from the APA guidelines. For further details on the extraction method see Nuijten et al. (2015).

been published, now reporting a consistent p -value. A number of inconsistent tests were marked as being consistent with one-tailed testing. Therefore, we manually checked those papers for any indication that one-tailed tests instead of two-tailed tests were conducted. Four tests were explicitly one-tailed in the corresponding publications, reducing the number to 129 inconsistent NHSTs (12.5% of K), of which 23 (2.2% of K) were grossly inconsistent. Forty-one publications (41.8% of n_{nhst}) reported at least one inconsistent NHST (range 1 to 21), and 16 publications (16.3% of n_{nhst}) reported at least one grossly inconsistent NHST (range 1 to 4) (see Figure 1). Thus, a substantial proportion of publications in *JMP* seem to contain inaccurately reported statistical analyses, of which some might affect the conclusions drawn from them.

Types of Errors. Many of the inconsistencies are probably clerical errors that do not alter the inferences or conclusions in any way. For example, in 20 cases the authors reported $p = .000$, which is mathematically impossible (for each of these $p_{\text{recomputed}} < .001$). Other inconsistencies might be explained by authors not declaring that their tests were one-tailed (which is relevant for their interpretation). Of course, in many cases we could not determine the source of errors without being able to access the study data or analysis scripts.

Although nearly one in six of the papers with NHSTs contain gross inconsistencies potentially affecting reported conclusions, caution is advised when speculating about the causes. As with other inconsistencies, random human error certainly plays an important part. However, with some concern, we observe it is unlikely to be the only cause, as in 19 out of 23 cases, the reported p -values were equal to or smaller than .05 while the recomputed p -values were larger than .05, whereas the opposite pattern was observed in only four cases. Indeed, if incorrectly reported p -values resulted merely from clerical errors, we would expect inconsistencies in both directions to occur at approximately equal frequencies.

We acknowledge that before the development of valuable tools like *statcheck*, there was little awareness of the high prevalence of reporting errors in psychology generally (including media psychology). All of these inconsistencies can easily be detected using the freely available R package *statcheck* or via www.statcheck.io for those who do not use R. *JMP* will adopt this practice for all forthcoming papers prior to publication, and we recommend researchers use *statcheck* for their own manuscripts and for the works of others in their role as reviewers.

Sample Sizes and Statistical Power

High statistical power is paramount in order to reliably detect true effects in a sample and, thus, to correctly reject the null hypothesis when it is false. Further, low power reduces the confidence that a statistically significant result actually reflects a true effect (Button et al., 2013; Schimmack, 2012). A generally low-powered field is more likely to yield unreliable estimates of effect sizes and low reproducibility of results.

We are not aware of any previous attempts to estimate average power in media psychology. One obvious strategy for estimating average statistical power is to examine the reported power analyses in empirical research articles. For publications in *JMP*, however,

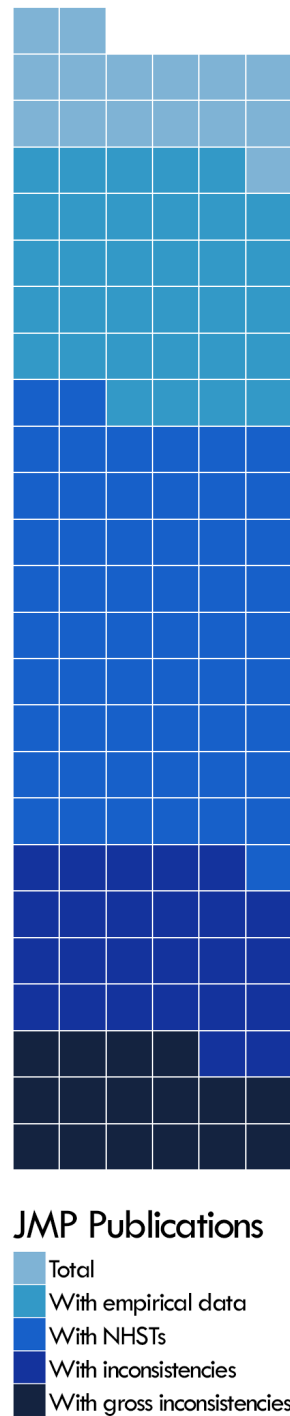


Figure 1.

Nested chart of publication characteristics in JMP volumes 20/1 to 28/2. Each square represents one paper. Figure created with the R package *waffle* (version 0.6.0; Rudis and Gandy, 2016).

Table 1

Sample sizes by research design in empirical studies published in JMP (volumes 20/1 through 28/2).

Research design	n	MD_I	M_I	SD_I	Min_I	Max_I
(Quasi-)experiments	93	107	139.71	130.53	18	748
Between-subjects	71	119	149.03	117.34	29	748
Mixed	12	54	73.58	44.24	29	176
Within-subjects	10	40	152.90	240.77	18	666
Cross-sectional surveys	35	327	500.51	531.07	58	2582
Longitudinal studies	8	378.50	595.00	489.70	97	1536
Total	136	139	259.35	355.94	18	2582

n = Number of published studies; MD_I = Median sample size; M_I = Mean sample size; SD_I = Standard deviation of M_I ; Min_I/Max_I = Smallest/largest reported sample size.

this is difficult, as searching all papers for the word “power” yielded only a single article reporting an a priori determined sample size. This is not to say media psychologists are generally unaware of the concept of power. In 19 further articles, power is indeed mentioned, in many cases to either demonstrate *observed* or *post-hoc* power (which is redundant with reported NHSTs; see, e.g., Lakens, 2014), to suggest larger samples should be used in future research, or to explain why an observed nonsignificant “trend” would in fact be significant had the statistical power been higher.

Another strategy is to examine the power for different effect sizes, e.g. using Cohen’s 1988 rule of thumb⁸, given the average sample size (I) found in the literature. The median sample size in *JMP* is 139 with a considerable range across all experiments and surveys (see Table 1). As in other fields, surveys tend to have healthy sample sizes apt to reliably detect medium to large relationships between variables. The median sample size for survey studies is 327, allowing researchers to detect small bivariate correlations of $r = .1$ at 44% power ($rs = .3$ and $.5$ both $> 99\%$).⁹ Longitudinal research exhibits similar characteristics, with a median sample size of 378.50, allowing researchers to detect $r = .1$ at 49% power ($rs = .3$ and $.5$ at $> 99\%$).

For experiments (including quasi-experiments), the outlook is a bit different, with a median sample size of 107. To determine average power in experimental designs, two further parameters must be considered: a) the study design (between-subjects or within-subjects), and b) the number of cells (or conditions) realized. Across all types of designs (see Table 2), the median cell size is 30.67. Thus, the average power of experiments published in *JMP* to detect small differences between conditions ($d = .20$) is 12%, 49% for medium effects ($d = .50$), and 87% for large effects ($d = .80$).

⁸Small: $r = .10/d = .20$; Medium: $r = .30/d = .50$; Large: $r = .50/d = .80$; α fixed at .05, all tests two-tailed. Power analyses were conducted with the R package *pwr* (version 1.20; Champely, 2016).

⁹Naturally, when anticipating more complex relationships between multiple variables, those numbers are dramatically different.

Table 2

Average cell sizes and power of (quasi-)experiments published in JMP (volumes 20/1 through 28/2) for different effect sizes.

Design	n	$MD_{i/cell}$	$M_{i/cell}$	$SD_{i/cell}$	$Min_{i/cell}$	$Max_{i/cell}$	$1-\beta_{d=.2}$	$1-\beta_{d=.5}$	$1-\beta_{d=.8}$
Between	71	30	43.57	46.87	14.50	374	12%	48%	86%
Mixed	12	26	34.25	21.71	14.50	88	11%	42%	81%
Within	10	40	152.90	240.77	18	666	23%	87%	99%
Total	93	30.67	54.34	93.18	14.50	666	12%	49%	87%

n = Number of published studies; $MD_{i/cell}$ = Median cell size; $M_{i/cell}$ = Mean cell size; $SD_{i/cell}$ = Standard deviation of $M_{i/cell}$; $Min_{i/cell}/Max_{i/cell}$ = Smallest/largest reported cell size; $1-\beta_{d=.2}/1-\beta_{d=.5}/1-\beta_{d=.8}$ = Power to detect small/medium/large differences between cells.

For between-subjects, mixed designs, and total we assumed independent t -tests. For within-subjects designs we assumed dependent t -tests.

Again, we currently do not have reliable estimates of the average true, expected, or even observed effect size in media psychology. But even when assuming that the effects examined in the media psychological literature could be as large as those in social psychology (average $d = .43$ according to Richard, Bond, and Stokes-Zoota, 2003), our results indicate that the chance that an experiment published in JMP will detect them is worse (at 38%) than flipping a coin – an operation that would also be considerably less expensive. We do not think this is a sustainable way of accumulating scientific knowledge and spending (public) resources.

The Future

Psychological and communication scientists use a wide range of methodologies to enhance our understanding of the role of media in human behavior. Unfortunately, as in other fields of social science (Pashler & Harris, 2012), much of what we think we know may be based on a tenuous empirical foundation. As a first estimation in this field, our analysis of *JMP* publications indicates that materials and data of few, if any, media psychology reports are openly available; many lack the statistical power required to reliably detect the effects they were set out to detect; and a substantial number contain statistical errors of which some might alter the conclusions the research draws. Although these observations are deeply worrying, they also provide some clear guiding points on how to improve our field.

Self-Reflection

Our observations could lead readers to believe that we are concerned about the quality of publications in *JMP* in particular. If anything, the opposite is true, as this journal recently committed itself to a number of changes in its publishing practices to promote open, reproducible, high-quality research. The space provided by the editor-in-chief for our analysis is simply another step in this phase of sincere self-reflection. Similar analyses in

other fields suggest that the issues we discuss here go far beyond media psychology (or the *JMP*): About half of all articles in major psychology journals report at least one inconsistent NHST (Nuijten et al., 2015) and at least one mean value that is inconsistent with the sample size and integer data (Brown & Heathers, 2016). Estimates of average statistical power in social psychology are similar to those in *JMP* (Fraley & Vazire, 2014), but as low as 18% in neuroscience (Button et al., 2013).

Thus, we would like these findings, troubling as they are, to be taken not as a verdict, but as an opportunity for researchers, journals, and organizations to reflect similarly on their own practices and hence improve the field as a whole.

Construction and Testing of Theories

One key area which could be improved in response to these challenges is how researchers create, test, and refine psychological theories used to study media. Like other psychology subfields, media psychology is characterized by frequent emergence of new theories which purport to explain phenomena of interest (Anderson, 2016). This generativity may, in part, be a consequence of the fuzzy boundaries between exploratory and confirmatory modes of social sciences research (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012).

Exploratory work – research meant to introduce new ideas and generate hypotheses – often involves taking a look at the results of a study and flexibly listening to “what the data have to say”. This mode is fundamental to social science work in general and also creatively informs media psychology. However, conclusions drawn from this mode are of limited strength, as they may reflect chance variations in the data. Confirmatory work, by contrast, involves theory testing, a process that requires research questions and hypotheses to be clearly stated in advance of data collection. This mode allows researchers and audiences to trust the results of a study that rigorously tests a specific prediction. A problem confronting theories in media psychology is that the boundaries between exploratory and confirmatory work are often blurred. This means that young scholars, parents, and policy-makers cannot know which studies present tentative results and which ones are worthy of investing limited resources to build on or implement in the real world.

The articles in this special issue provide a clear example of how exploratory and confirmatory modes of research can coexist and how science can thrive as a result. Most present both exploratory and confirmatory elements, each clearly labeled. As a result, it is easier to see and understand what the researchers were expecting based on knowledge of the relevant literature, and what they eventually found in their studies. It allows readers to build a clearer idea of the research process and the elements of the studies that came as inspiration after the reviews and the data collection were completed.

Both modes of research – studying previously observed phenomena and exploring uncharted territory – benefit from preregistration. Drawing this distinction helps the reader determine which hypotheses carefully test ideas derived from theory and previous empirical studies, and it liberates exploratory research from the pressure to present an artificial

hypothesis-testing narrative. The Registered Reports model is also an effective counter-measure against psychology’s aversion to statistically nonsignificant or “null” results, as study protocols are reviewed and accepted before results are known. Further, it ensures that p -values can be meaningfully interpreted, given these lose their meaning in data exploration as the Type I error inflation is unknown (Wagenmakers et al., 2012). If adopted by media psychology researchers, this approach could allow us to rigorously test and extend promising theories, and to retire theories that do not reliably account for observed data.

Increasing the Value of Media Psychology with Open Science Tools

As technology experts, media psychology researchers are well positioned to use and study new tools that shape our science. A range of new Internet-based platforms have been built by scientists and engineers at the Center for Open Science, including their flagship, the OSF (<http://www.osf.io>), and preprint services like PsyArXiv (<http://www.psycharxiv.com>) and SocArXiv (<http://www.socarxiv.com>). Designed to work with scientists’ existing research flows, these tools can help prevent data loss due to hardware malfunctions, misplacement, or relocations of researchers, while enabling scientists to claim more credit by allowing others to use and cite their materials, protocols, and data.¹⁰

High Stakes Facing Media Psychology

Like psychological science as a whole, media psychology faces a pressing credibility gap. Unlike some other areas of psychological inquiry, however, media research — whether concerning the Internet, video games, or film — speaks directly to everyday life in the modern world. It affects how the public forms their perceptions of media effects (Przybylski & Weinstein, 2016), and how professional groups and governmental bodies make policies and recommendations (Council on Communications and Media, 2016). In part because it is key to professional policy, empirical findings disseminated to caregivers, practitioners, and educators should be built on an empirical foundation with sufficient rigor. If policy makers and the public are to value our views as experts, we must take steps to demonstrate this trust is warranted. Such challenges and high stakes are by no means unique to media psychology.

Indeed, in medical and drug research, the study registration movement (Goldacre & Gray, 2016) has lead the way with publicly accessible registries that include all the studies, published or not, conducted in a research area. To build good faith for the general public, industry collaborators, and policy makers, we propose the creation of a registry (<https://osf.io/registries/>) for confirmatory media psychology research. Creative exploratory research (i.e. theory building) would continue as it does now, but confirmatory work (i.e. theory testing) could be registered in a central repository so that its results — positive, null, or negative — would be available for scrutiny. This would also allow media psychologists to educate the general public and journalists about the distinction between exploratory and confirmatory research. Through such a public registry, researchers

¹⁰A public repository for media psychology research materials is already in place at <https://osf.io/wzb32/>

and policy makers could quickly determine which evidence is promising (though tentative), and which conclusions are suitable as the basis for interventions, policy decision making, caregiver guidance, or new products.

Closing Thoughts

We are, on balance, optimistic that media psychologists can meet these challenges and lead the way for psychologists in other areas. This special issue and the registered reports submission track present an important step in this direction and we thank the *JMP* editorial board, our expert reviewers, and of course, the dedicated researchers who devoted their limited resources to this effort. The promise of building an empirically-based understanding of how we use, shape, and are shaped by technology is an alluring one. We firmly believe that incremental steps taken towards scientific transparency and empirical rigor will help us realize this potential.

Acknowledgments

We sincerely thank Charleen Brand and Hannah Borgmann for their invaluable assistance with the data preparation for this editorial.

References

- Anderson, J. A. (2016). Communication descending. *International Communication Gazette*, 78(7), 612–620. doi:10.1177/1748048516655708
- Brown, N. J. L. & Heathers, J. A. J. (2016). The GRIM Test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science*. doi:10.1177/1948550616673876
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. doi:10.1038/nrn3475
- Champely, S. (2016). pwr: Basic functions for power analysis. Retrieved from <http://cran.r-project.org/package=pwr>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Council on Communications and Media. (2016). Media and young minds. *Pediatrics*, 138(5), e20162591–e20162591. doi:10.1542/peds.2016-2591
- Elson, M., Przybylski, A. K., & Krämer, N. C. (2015). Technology and human behavior: A preregistered special issue of the Journal of Media Psychology. *Journal of Media Psychology*, 27(4), 203–204. doi:10.1027/1864-1105/a000170
- Epskamp, S. & Nuijten, M. B. (2015). statcheck: Extract statistics from articles and recompute p values. doi:<http://CRAN.R-project.org/package=statcheck>
- Fraley, R. C. & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLOS ONE*, 9(10), e109019. doi:10.1371/journal.pone.0109019

- Goldacre, B. & Gray, J. (2016). OpenTrials: Towards a collaborative open database of all available information on all clinical trials. *Trials*, 17(1), 164. doi:10.1186/s13063-016-1290-8
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., ... Nosek, B. A. (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
- Lakens, D. (2014). Observed power, and what to do if your editor asks for post-hoc power analyses. Retrieved from <http://daniellakens.blogspot.de/2014/12/observed-power-and-what-to-do-if-your.html>
- Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2015). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*. doi:10.3758/s13428-015-0664-2
- Pashler, H. & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531–536. doi:10.1177/1745691612463401
- Przybylski, A. K. & Weinstein, N. (2016). How we see electronic games. *PeerJ*, 4, e1931. doi:10.7717/peerj.1931
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org>
- Richard, F. D., Bond, C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7(4), 331–363. doi:10.1037/1089-2680.7.4.331
- Rudis, B. & Gandy, D. (2016). waffle: Create waffle chart visualizations in R. Retrieved from <https://cran.r-project.org/package=waffle>
- Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, 17(4), 551–566. doi:10.1037/a0029487
- Vermeulen, I., Beukeboom, C. J., Batenburg, A., Avramiea, A., Stoyanov, D., van de Velde, B., & Oegema, D. (2015). Blinded by the light: How a focus on statistical "significance" may cause p-value misreporting and an excess of p-values just below .05 in communication science. *Communication Methods and Measures*, 9(4), 253–279. doi:10.1080/19312458.2015.1096333
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638. doi:10.1177/1745691612463078
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728. doi:10.1037/0003-066X.61.7.726