Oller, D. K., & Griebel, U. (Eds.). (2004). *Evolution of communication systems: A comparative approach.* Cambridge, MA: Bradford.

Schütze, C. T. (2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. Berlin, Germany: Language Science Press. Retrieved from http://www.oapen.org/search?identifier=603356

Smith, E. A. (2010). Communication and collective action: Language and the evolution of human cooperation. *Evolution and Human Behavior, 31,* 231–245. Retrieved from https://doi.org/10.1016/j.evolhumbehav.2010.03.001

Thompson, B., Kirby, S., & Smith, K. (2016). Culture shapes the evolution of cognition. *Proceedings of the National Academy of Sciences, 113,* 4530–4535. Retrieved from https://doi.org/10.1073/pnas.1523631113

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

HOW NOT TO PLAY THE GAME OF PSYCHOLOGICAL INQUIRY

The Seven Deadly Sins of Psychology: A Manifesto for Reforming the Culture of Scientific Practice

By Chris Chambers. Princeton, NJ: Princeton University Press, 2017. 288 pp. Hardcover, $29.95.

In another lifetime, when I was a fresh assistant professor at the University of Wisconsin, a colleague and I were discussing his research. I asked him, "Yes, but how important is this finding really?" He replied immediately, "$p < .05$." I was a little taken aback to hear that this statistical value was a measure of importance. I thought I was somewhat of a lone voice because our reputable journals were requesting even more inferential statistics to perhaps justify a "science" publication. Now, many decades later, Chris Chambers offers a manifesto that rightly denigrates inferential statistics as part of his list of seven sins of psychological inquiry.

The Deadly Sins

So what are the seven deadly sins and the concomitant commandments that should be followed to lead a pure scientific life? First on the list is our intrinsic original sin of bias, most notably confirmation bias. Psychologists and behavioral scientists are not immune to seeking and favoring evidence that support their beliefs and ignoring or denigrating results that somehow disagree with these beliefs. It is not necessary to sermonize readers of this journal about this persistent bias. Confirmation bias was very apparent in the 2016 election and its aftermath, but it is not limited to politics. Mercier and Sperber (2017) provide a rationalization for confirmation bias that fits well in much of psychological inquiry: Winning arguments takes precedence over truth.

Chambers shows that, even in the context of the respected hypothetico-deductive model of the scientific method, researchers have evolved various techniques to instantiate confirmation bias. Thus this venerable method does not ensure that confirmation bias does not enter the everyday life of scientific inquiry. Our science rewards novel and positive results, not negative findings or replications of previous results in the literature. This payoff system encourages investigators to game the system. Thus, the literature tends to archive only positive findings; negative findings are demoted to the "file drawer" in good faith or even trashed by scientists with less of a conscience.

Seeking positive results can seamlessly convert researchers to Harking (Hypothesizing After Results Are Known) and other ritualistic strategies to guarantee success. One strategy is to change the investigator's initial hypothesis to one compatible with the outcome of the research. Another colleague confided to me that once the results were in, he realized how his initial predictions from Freudian theory were misguided. Thus, his opinion and resulting publication postulated that Freudian theory was able to survive yet another critical challenge.

The second sin is to exploit the hidden flexibility we have as researchers to maintain our good standing in the club. Inferential statistics provide many ways to lie. If an investigator has several possible dependent measures to draw from, then the odds of one of them being significant are greatly increased. Another strategy is to test additional participants until the magical p value is obtained. Another gift of flexibility is that we are free to double check the results when they disagree with our wishes, but of course there is no need to double check the results when they support them.

Chambers captures a variety of flaws encompassed in the third sin of unreliability. First on the list is our field's reluctance to replicate ("Replication Crisis," 2017), and when it does occur with different outcomes, it is too easily palmed off as not a true replication. Handicapping replication research are the typical ills of inquiry, including lack of power, and statistical fallacies, as well as the societal ills of disclosing important details of the study and reluctance to admit being wrong.

Although it occurs less than it should in our discipline, replication research also promotes positive

© Copyright 2018 by the Board of Trustees of the University of Illinois. No part of this article may be reproduced, photocopied, posted online, or distributed through any means without the permission of the University of Illinois Press.

over negative results. Replication of an experiment is seamlessly relegated to conceptual replication, so that a positive outcome can be interpreted as supporting the result and interpretation being replicated. A negative result, on the other hand, can always be pawned off as failing to exactly replicate the experiment in question.

The fourth sin is data hoarding. Certainly, we see through the egregious belief that "My data are my data and are meant only to raise my boat in the water." I have had mixed experiences in requesting raw data from colleagues. In the majority of cases, the investigators have willingly fulfilled the request. Data sharing is important because it is critical that other researchers, particularly those with different theoretical leanings, produce results that can be tested within your metatheoretical framework. On a personal note, our Fuzzy Logical Model of Perception (FLMP) has gained credence because of its descriptions of a variety of results from different laboratories and from different investigators with no allegiance to the model (Movellan & McClelland, 2001).

In this reviewer's mind, perhaps Chambers has unearthed only four mortal sins. Although I am the first to accept fuzzy boundaries, the next three might be considered venial sins in the sense that they build on the mortal sins already revealed or deal more directly with the sociology and business of our discipline.

The fifth sin is corruptibility. Of course, there many reasons that researchers might corrupt their research endeavors. There is no shortage of justifications for a little fraud: achieving tenure to support a growing family, helping a graduate enter the marketplace, convincing yourself that everyone does it, being confident that additional research would surely come out that way, and many others. Some of the corrupt interventions hark back to the earlier mortal sins. Chambers envisions the change from good scientific practice to fraudulent misconduct as being separated by a thin gray line. As with most dimensions, we can envision a continuum between bona fide research and outright intentional fraud.

Chapter 6 explicates the sixth sin of internment. A long-lasting barrier to scientific inquiry has been publishing. How much has to be compromised to make your results public in a respected venue? We valued refereed journals because putatively peer reviewers have vetted the research published in these journals. However, history has taught us that biases extend their reach well beyond the individual investigator. An old case in point is the excommunication of researchers who questioned the validity of the accepted belief that bees have language (Massaro, 1992; Wenner & Wells, 1990). More recent examples come from a Nobel laureate in chemistry (Massaro, 2012). There is obviously a tension between peer review validation and open access publishing. We want to make our results publicly available, but it is close to essential to have them bear the gold star of peer review. In this chapter we learn about the debate between open access and peer-reviewed publication. Chambers hopes to convert his readers to open access, which would democratize psychological inquiry without sacrificing quality.

The seventh deadly sin is bean counting, a measurement of research prowess that is central to promoting young scientists to elite organizations such as prestigious universities and businesses. The author uses this soapbox to expose the misleading attractiveness of journal impact factor, the priority of obtaining research grants over motivated empirical inquiry, and the overly emphasized concern with the order of authorship on published research. Like it or not, psychological science is also a social endeavor and a business. These dimensions retard idealized inquiry. In this burgeoning age of artificial intelligence, we might ask whether robots could be programmed to better direct scientific inquiry in a purely unbiased manner.

The Commandments

Although Chambers offers solutions to each of the sins as they are presented, he devotes a final chapter to the strategies that can mitigate them. To guard against confirmation bias, the author prescribes publicly registering our research intentions before we initiate the empirical inquiry. With this recorded formulation, the investigator is accountable for this inquiry with very little wiggle room. The author has already advocated and has had some success in finding journals and other forums to implement this form of research registration. It remains a question whether this constraint will mitigate results contaminated by confirmation bias.

He shares his early personal experience of having a paper rejected because "the methods are solid but the findings are not very important" (pp. 174–175). Then, about 13 years later, he had a paper rejected because it reported a nonsignificant outcome. This revealed to him a bias for positive findings and therefore an increased likelihood of false positives (significant results that are not really significant). This led him to undertake, with others, the promotion of

© Copyright 2018 by the Board of Trustees of the University of Illinois. No part of this article may be reproduced, photocopied, posted online, or distributed through any means without the permission of the University of Illinois Press.

Registered Reports ("Registered Replication Reports," 2017). Within this framework, publication is a two-stage process. First, authors submit an introduction, design, and data analysis before any results are collected. This submission is evaluated on our tried-and-true criteria that have been honed over the past century. If it passes muster, then it achieves an "in-principle acceptance." The investigators then implement the study, add the results and discussion, and submit the paper for publication. The paper is published if it adheres to the established criteria, such as not including any new embellishments not anticipated in the original submission.

This proposal generated a flood of interest and commentary, and as expected, several objections permeated the Web. An obvious complaint is that researchers can game the system by having the data in hand before they submit the stage 1 report. The check against this possibility is that the authors are required to time-stamp the results and certify that they were collected on these dates after the stage 1 report was accepted. The investigators are still free to accumulate lots of pilot data before the stage 1 submission to hone their stage 1 introduction and design. Although this might still be considered gaming the system, the effort invested in pilot studies seems to have more positive than negative outcomes. Although the author reasonably counters several possible objections, I worry that given the extant milieu of excessive litigation, we are increasing its role in research practice.

Preregistration of experimental research can mitigate against the second sin of flexibility because it should include sufficient detail about the method and the proposed data analysis.

To preclude the third sin of unreliability, the author suggests developing a reproducibility index. However, there are currently too many options, and settling on a single index that will be accepted and effective would involve too much effort and expense. Such an index might supplement meta-analyses in which the previous datasets are analyzed to achieve greater power. The proposed safeguards to reduce unreliability include attention to power analyses, Bayesian statistics rather than null hypothesis statistical tests, and full disclosure of the details of the experimental design, method, and data analyses. At least one journal now requires a checklist to ensure full transparency of the research details.

Embedded in the third-sin chapter is the author's depiction of a replication strategy within the hypothetico-deductive model of the scientific method. This research strategy can be described as following John Platt's (1964) strong inference, which embellishes Karl Popper's (1959) idea of a falsification research strategy. Ideally, competing hypotheses are contrasted against a dataset from an experiment designed to falsify at least one of the hypotheses. Bonett (2012) formalizes a valuable replication–extension paradigm that considers replication essential to extensions of previous research.

In his manifesto, Chambers does not acknowledge the additional safeguards that can be achieved by formalizing hypotheses in precise quantitative models. The sins of statistical testing can often be bypassed by the approach of mathematical psychology. Quantitative analysis and mathematical models are central to inquiry in which the assumptions being made can be quantified and tested exactly. We have demonstrated this in our various tests of models of processing multiple sources of information (Massaro, 1998). Within this framework, it is possible to make theoretical distinctions between single-channel models when only one source of information is used during any given test event and integration models that combine or use multiple sources of information on a given trial. This approach has been highly successful at distinguishing between various models of how multiple sources of information are processed in pattern recognition and memory. This approach also uses a benchmark describing the goodness of fit of a model. The benchmark provides a measurement standard that assesses how well the model does relative to the best possible model given the variability in the data. Another central component is to carry out individual subject analysis because we know that the average across a group of subjects may not reflect any of the single subjects that make up the average.

As a solution to the fourth sin of data hoarding is archiving results to make them available to the scientific community at large. Chambers has been part of an initiative called the Peer Reviewers' Openness Initiative (2017), which asks authors of submitted manuscripts to place their data in a public archive or provide a reason for not sharing. Given the proven usefulness of big data, one can only hope that we will see data sharing as typical practice rather than a rarity in the field. Several depositories of databases have provided good examples of the value of sharing data (databases, 2017).

Corruptibility could be counteracted by various monitoring and profiling activities but most importantly in my view by emphasizing replication across various laboratories. Chambers lobbies for data shar-

© Copyright 2018 by the Board of Trustees of the University of Illinois. No part of this article may be reproduced, photocopied, posted online, or distributed through any means without the permission of the University of Illinois Press.

ing for many reasons, and one of these is to facilitate meta-analyses. He acknowledges that meta-analysis is a valuable endeavor for psychologists, but he does not consider the possibility that the abundance of meta-analyses must mean that replication is indeed occurring.

Internment, the limited access now available to researchers, might be reduced by having open access forums (becoming more common with the increasing popularity of Internet dialogs). A good example of such a forum in speech science is *Talking Brains* (2017), organized by Greg Hickok and David Poeppel.

Finally, to reduce evaluation by bean counting, more emphasis could be placed on methodological rigor and theoretical rigor. Chambers ends the chapter and the book with concrete steps for reform aimed at junior and senior researchers; journals, grant funders, professional societies, and universities; and journalists and citizens.

Potential Embellishments

Chambers has contributed significantly to our discipline both by this book and by his active advocacy for improving how our science is practiced. I end this review by embellishing various commandments and advocating a perspective even more formal than is extant. Consider a result that is now being replicated in a registered replication report (Mazar, Amir, & Ariely, 2008; "Registered Replication Reports," 2017), which concluded that a moral reminder significantly reduces cheating. Mazar et al. created a problem-solving task and gave their participants an incentive to perform well. In one condition, participants were able to report solving a greater number of problems than they actually did, with no risk of being caught. However, when participants were given moral reminder (recall the Ten Commandments) before the task, they reported fewer solved problems than those given a neutral reminder (i.e., recall 10 books they read in high school).

Inferential statistics indicated that subjects cheated when given the opportunity to do so but not when they were given the moral prime before the task. Thus, there is evidence that dishonesty can be abated when attention is drawn to honesty standards. Should we be surprised by this result, and should a scientist expect to replicate it with ease? In terms of surprise value, religions and families have traditionally used moral reminders to maintain honesty (of course, they might not be effective, but we believe they are). Mazar et al. (2008) also reported that cheating across all conditions was much less than possible, occurring only 6.7% of the maximum. From this small number, we might speculate that the authors were close to a floor effect in which their dependent variable might be insensitive to their manipulation of interest. Thus, we should not be surprised if the results were not easily replicated.

Failing to replicate this finding turned the issue into a controversy, but it is not surprising that the effectiveness of something like a moral reminder would depend on many different conditions and other sources of influence effective in the experimental situation. Initially the scientist could attend to the ecological domain of moral reminders and the other possible influences that could be expected to modulate cheating behavior. Then, following my idea of an expanded factorial design (Massaro, 1998), the scientist could study many obvious variables. The participants in the task would be given multiple trials across all of the experimental conditions in order to make individual subject analyses feasible. The outcome for each participant could be tested across a range of quantitative models. In this way, goodness of fit of the models is a deciding factor in terms of the best description of how people use multiple sources of information to influence potential cheating behavior.

A good example of this approach was carried out by Norman Anderson and his students (Leon, 1980; Anderson, 2012). In his analysis they speculated that the blame for an act would be a function of Responsibility (Intent) and Consequences (Harm). Evidence for this formalization was determined from a factorial design experiment in which participants judged the naughtiness of hypothetical acts that were described by the actor's intent and the consequences of the act. The intent could be an act of malice, displacement, or accident. The consequences were four levels of severity. The results followed Anderson's parallelism law in which intent and consequences made independent and additive contributions to judgments of naughtiness. This is a strong formalization because it basically relegates all other influences as playing a different role.

Within our competitive research environment, another investigator could very well manipulate another source of information, but it should be in the context of Leon's original factorial experiment. This strategy would then implement replication and extension in the sense that the investigator would expect to replicate the results but also measure some influence of yet another variable. For example, an apology might

© Copyright 2018 by the Board of Trustees of the University of Illinois. No part of this article may be reproduced, photocopied, posted online, or distributed through any means without the permission of the University of Illinois Press.

reduce the judgment of naughtiness. On the other hand, if this new manipulated variable somehow negated Leon's original results, then we have a whole new ballgame. And of course this is what we expect during a scientific enterprise: to successively puzzle through manipulations and results while keeping ecological validity in mind.

We want our research findings to motivate positive behavioral interventions. We can look at some positive interventions in decision making for examples to see what kind of experiments led to these interventions that were proven to be effective. One example is whether to opt out or opt in when considering whether a portion of a person's earnings should be set aside for retirement or a college fund for their children. Employees are more likely to enroll in the savings program if this is the default setting and thus they do not have to actively opt into the program (NBER, 2017). If they have to opt in by checking a box, then they are less likely to set earnings aside for savings. This discussion leads to an insight that perhaps a successful implementation is more convincing than replications in the research literature. If so, this is a call to promoting a much closer relationship between research and application than is currently the case.

Individual Differences

Given the persistent variation across individuals in all domains of study, psychological inquiry has to face up to how it is going to handle this variation. Clearly, inferential statistics is not the answer. Perhaps effect size could be valuable, but effect size is necessarily measured in a specific experiment in which the effect size may be much larger than what would be found in a natural setting. This is particularly true in single-factor experimental designs in which all influences are made as neutral as possible and the influence of interest is manipulated across a wide range. This would necessarily give a larger measure of the influence of interest than would be expected from a more complex design in which many different sources of information are informative (limiting the ecological validity of the experiment).

A personal anecdote comes to mind. We had a colloquium from a visiting IBM engineer who criticized the scroll mouse (IBM, scroll mouse, this was many years ago). Before he could complete his criticism, one of our computer scientists blurted out, "I love my scroll mouse." As Samuel Johnson said two and a half centuries ago, "What we have long used we naturally like" (1775/2002, p. 42). We are not going to debate the value of the scroll mouse but simply make the point that individual differences are pervasive enough for us to believe they make the world go around. Importantly, our experience in many ways influences our preferences. Obviously the computer scientist had mastered the scroll mouse with lots of time on task, and it served his purpose very well. Moreover, he might have overinterpreted the value of the mouse itself as opposed to his acquired expertise (which might have been as accomplished with many other input devices).

It is probably premature to attempt to legislate an appropriate scientific method without first considering what can and cannot be accomplished in behavioral science. It is probably the case that we will have to accept substantial individual differences superimposed on most findings of interest. There are cases in physical science in which variation is accounted for by one or more free parameters within a formal mathematical description. Dynamic systems have proven to be accurate in predicting a so-called attractor toward which a system tends to evolve, for a wide variety of starting conditions of the system. The end state of the system is a set of numerical values that get close enough to the attractor values even if the system is slightly disturbed. The mathematical description of the dynamics toward the attractor requires free parameters that are a function of the actual physical system being modeled.

Analogous to the attractor example from physical science, we have shown in our research that the FLMP, a general algorithm, describes how people integrate multiple sources of information. Individual differences are most prominent in terms of the information available to perceivers but much less so in how the information is processed. Theoretically, it is not feasible to attempt to account for the amount of information available from each source. Each person has a unique genetic makeup, sensory abilities, and life experiences that result in large differences in the information available in a given situation but perhaps not with respect to fundamental algorithms of information processing. Thus, there is necessarily a free parameter in the model for a given source in order to provide a good description of how the multiple sources of information are processed.

Retrospective

I applaud Chambers for advocating reform of our science and this book for encouraging me to rethink our discipline. This book should be required reading for all graduate students and, of course, their mentors.

© Copyright 2018 by the Board of Trustees of the University of Illinois. No part of this article may be reproduced, photocopied, posted online, or distributed through any means without the permission of the University of Illinois Press.

I am looking forward to seeing how this revisionist view plays out in practice.

Dom Massaro
Department of Psychology, Social Sciences II
University of California–Santa Cruz
Santa Cruz, CA 95064 USA
E-mail: massaro@ucsc.edu

REFERENCES

Anderson, N. H. (2012). *Moral science*. Retrieved from http://www.psychology.ucsd.edu/people/profiles/nanderson.html

Bonett, D. G. (2012). Replication-extension studies. *Current Directions in Psychological Science, 21,* 409–412.

Databases. (2017). http://wordbank.stanford.edu/; http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm; http://elexicon.wustl.edu/

Johnson, S. (2002). *A journey to the western islands of Scotland*. New York, NY: Alfred A. Knopf. (Original work published 1775)

Leon, M. (1980). Integration of intent and consequence information in children's moral judgments. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 71–97). Mahwah, NJ: Erlbaum.

Massaro, D. W. (1992). *Anatomy of a controversy: The question of a "language" among bees* by Adrian M. Wenner, Patrick H. Wells. *American Journal of Psychology, 105,* 653–659.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, MA: MIT Press.

Massaro, D. W. (2012). A quarter century of book reviews in *The American Journal of Psychology. American Journal of Psychology, 125,* 499–500. doi:10.5406/amerjpsyc.125.4.0499. Stable URL: http://www.jstor.org/stable/10.5406/amerjpsyc.125.4.0499

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research, 45*(6), 633–644.

Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.

Movellan, J., & McClelland, J. L. (2001). The Morton–Massaro law of information integration: Implications for models of perception. *Psychological Review, 108,* 113–148.

NBER. (2017). The effect of default options on retirement savings. Retrieved from http://www.nber.org/aginghealth/summer06/w12009.html

Peer Reviewers' Openness (PRO) Initiative. (2017). Retrieved from https://opennessinitiative.org/

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.

Registered replication reports. (2017). Retrieved from https://www.psychologicalscience.org/publications/observer/obsonline/two-priming-effects-to-be-examined-in-new-registered-replication-reports-with-combined-protocol.html#.WRXx2cm1uEA

Replication crisis. (2017). Retrieved from https://en.wikipedia.org/wiki/Replication_crisis

Talking brains. (2017). Retrieved from http://www.talkingbrains.org/

Wenner, A. M., & Wells, P. H. (1990). *Anatomy of a controversy: The question of a "language" among bees*. New York, NY: Columbia University Press.

THE PAYNE OF INEQUALITY

The Broken Ladder: How Inequality Affects the Way We Think, Live, and Die

By Keith Payne. New York, NY: Viking, 2017. 256 pp. Hardcover, $28.

As long as poverty, injustice and gross inequality persist in our world, none of us can truly rest.
—Nelson Mandela (n.d.)

Social psychology suffers from two recurring crises. One crisis is about the field's status as a science. Lack of a master theory, failures of replication, and the woes of weak statistical methods (Krueger & Heck, 2017), as well researchers' sloppiness or downright duplicitousness, cast one pall after another (Lilienfeld & Waldman, 2017; see Pratkanis, 2017, for an assertive response therein). The other crisis is the absence of a clear domain of application and hence a perceived lack of relevance (but see Steg, Buunk, & Rosengatter, 2008, for an effort to fix this). Yet we live in interesting times (in the Chinese sense of *interesting*) where challenges abound. How can social psychology not be relevant?

Perhaps the historically most prominent challenge taken up by social psychology is the problem of racism. The study of racism presents a dialectic that continues to frustrate many a researcher. On one hand, there is a social reality, which is structural and systemic. On the other hand, there are the psychological processes and mechanisms that psychologists must prioritize if they want to remain true to their field. The pendulum tends to swing more to the individual than to the social. Of late, the study of implicit bias has been particularly popular, to the point that everything that appears to be of consequence is located in the person's head, and beneath the threshold of awareness at that (see Mitchell & Tetlock, 2017, for a critical analysis). Sociologists since Durkheim

© Copyright 2018 by the Board of Trustees of the University of Illinois. No part of this article may be reproduced, photocopied, posted online, or distributed through any means without the permission of the University of Illinois Press.