EDITORIAL

M. Lynne Cooper
Editor, *Journal of Personality and Social Psychology: Personality Processes and Individual Differences*

Since the early 1990s, the field of personality psychology has matured as a science, growing in prominence, recognition, and respectability (perhaps more outside of psychology than within). Contemporary personality science—at its best—is characterized by an emphasis on explanation as opposed to description, even in trait/structural approaches that have long been disparaged for their failure to attend to explanatory mechanisms. It is also characterized by increasing (and increasingly successful) efforts to link personality across levels of analysis, from internal biological processes to macrolevel, sociocultural phenomena, and to connect to and draw on other fields as a way to understand personality, and its origins, dynamics, and consequences. Also on the rise are increasingly sophisticated efforts to understand and study the person as a dynamic whole, embedded in a larger, sociocultural context. Finally, concerted efforts have been undertaken to document and establish the relevance of personality to real-world processes and outcomes; thus, making a compelling case for personality's "place at the table." These developments, in many cases made possible by significant methodological and statistical advances, make personality science the exciting area of study it is today.

Nevertheless, a number of serious issues have been raised in recent years concerning the replicability and integrity of psychological research. Although, for the most part, these criticisms have not been aimed specifically at personality psychology, no area within psychology is immune (Roberts, 2011). As the premier outlet for empirical work in personality and individual differences, I believe that *Journal of Personality and Social Psychology: Personality Processes and Individual Differences (JPSP:PPID)* is not only in a unique position to take a lead role in addressing these issues but also has a special responsibility to do so.

Recent analyses establish in a very convincing manner that the proportion of positive results reported in the psychological literature (Bakker, van Dijk, & Wicherts, 2012) far exceeds what would be expected given the low power (Pashler & Harris, 2012) of the typical study published in psychology. Indeed, Pashler and Harris (2012) estimated that 56% of published findings are false positives, whereas Ioannidis (2012) argued that the proportion of false positives could reach as high as 95%, assuming even modest levels of publication bias. Regardless of where the actual percentage lies, such analyses are sobering and point to the need to reevaluate both how we define good science and how we weigh the strengths and weaknesses of articles submitted for publication. Although wholesale changes are neither prudent nor practical (after all, there is much to applaud about contemporary research on personality and individual differences), I describe below several changes in editorial policy and practice that will be undertaken during my term as editor, with the goal of broadening the definition of what we consider to be good science and, more long term, of increasing the "truth value" of research in the field of personality science.

Broad Policy Issues Regarding the Form, Style, and Substance of Research Articles

Greater Emphasis on Description

As Rozin (2001, 2009) has pointed out, psychology strongly privileges hypothesis testing, experiments, and sophisticated methodologies and statistical analyses, while simultaneously devaluing description of phenomena and assessment of their generality. In contrast, the natural sciences, especially the life sciences, pay less attention to models and hypotheses and more attention to what the data have to tell us about the nature of the phenomenon under study. Description, as Rozin pointed out, directs us to what is worth studying by carefully characterizing its properties and establishing its robustness outside the laboratory and across cultures. Description precedes and becomes the basis for theory and hypothesis testing in the natural sciences. Funder (2009) advanced a similar argument: "We need a map of the broader behavior terrain," and drawing this map,

Journal of Personality and Social Psychology, 2016, Vol. 110, No. 3, 431–434
© 2015 American Psychological Association 0022-3514/16/$12.00 http://dx.doi.org/10.1037/pspp0000033

he points out, will require that the field places a higher priority on descriptive, mostly correlational research that "measures interesting and consequential behaviors across a realistic range of situational variables" (p. 343). This position is not meant to eschew or denigrate the importance of explanation as an ultimate goal of science, but rather to acknowledge the foundational role that description can and should play in the building of a cumulative science.

Consistent with this line of thinking, I believe that we need to move away from the virtual requirement that every article published in *JPSP:PPID* must follow the standard deductive, theory-driven, hypothesis testing paradigm, and instead foster a broader appreciation of different, yet methodologically sound approaches to seeking and accumulating knowledge, including explicitly embracing the value of descriptive studies that aim to understand and characterize the natural ecology of important, psychologically meaningful behaviors and the factors that shape them. In short, I agree with Paul Rozin's (2009) analysis that, "The principal aim of academic psychology is to understand how humans and animals behave, think, and feel and how these events influence and are influenced by their material and social environment" (p. 435). Furthermore, it is this—not whether the contribution conforms to a particular model of doing science or a particular theoretical tenet—that should guide the evaluation of submitted articles. Moreover, a move in this direction would be highly responsive to concerns that confirmation bias is a logical consequence of the field's theory-centered research focus, and a key factor contributing to false-positive results in psychology (e.g., Ioannidis, 2012; LeBel & Peters, 2011; see also Greenwald, Pratkanis, Leippe, & Baumgardner, 1986).

Greater Methodological Heterogeneity

Although it is clear that articles published in *PPID* reflect a broader and more diverse array of methodological approaches than was true even a few years ago, promoting methodological heterogeneity nevertheless remains a high priority. As McGrath, Martin, and Kulka (1981; see also Campbell & Cook, 1979; Shadish, Cook, & Campbell, 2001) have argued, we can never know anything independent of the methods used to acquire the knowledge. Accordingly, whenever a field relies heavily on a single predominant method or paradigm, no matter the strength of that approach, content and method variance can no longer be easily separated, and as a result, the validity of the entire body of work is compromised. Under such circumstances, methods whose strengths and weaknesses complement rather than repeat what has gone before should be encouraged as the "next best step" in systematic efforts to explore a particular issue or topic (McGrath et al., 1981). Thus, although the traditional multistudy package of laboratory experiments will certainly continue to find a home at *PPID*, I will seek to achieve better balance between such studies and studies using more diverse methods and approaches.

Larger and More Diverse Samples

As a recent high-profile article made clear, the vast majority of studies (96% by the authors' estimate) published in top psychology journals, including *JPSP*, use WEIRD samples—that is, samples that are White; educated; and hail from industrialized, rich, and democratic countries (Henrich, Heine, & Norenzayan, 2010). In other words, psychology samples almost exclusively from a small, nonrepresentative, homogenous slice of humanity that is, in the words of Henrich and colleagues, "potentially peculiar." Although it is clear that obtaining non-WEIRD samples is difficult and costly and that the preponderance of research will (at least for the foreseeable future) continue to be conducted on samples from rich, industrialized, Western nations, we nevertheless can and should encourage authors to venture outside the undergraduate psychology pool to obtain their samples. Thus, manuscripts submitted to *JPSP:PPID* that use more diverse samples are strongly encouraged, as are those that select samples on the basis of appropriateness for the research question as opposed to convenience. In addition, although articles relying exclusively on convenience samples of college students will not be rejected solely on this basis, they will be closely scrutinized to ensure that their potential overall contribution is sufficient to offset weaknesses associated with the sample. Likewise, we will cut authors submitting studies that use rare or difficult-to-obtain samples some slack, particularly in terms of sample size, which is the second sample-related issue to which I now turn.

The average study in psychology is seriously underpowered, with an estimated power of only about 35% to detect a true, nonnull effect if one indeed exists in the population (Pashler & Harris, 2012). Although it is widely understood that low power can increase the probability of Type II error, recent analyses have shown that fields populated by underpowered studies actually experience greatly elevated rates of false-positive findings as well (Pashler & Harris, 2012).

Moreover, as emphasis on null hypothesis significance testing gives way to an emphasis on *accuracy* of estimation and the concomitant reporting of confidence intervals, researchers must become increasingly cognizant of the sample sizes needed to generate *accurate* (not just adequately powered) estimates of key effects (see Maxwell, Kelley, & Rausch, 2008, for a comprehensive treatment of this issue; see also Schönbrodt & Perugini, 2013). Because a host of complex issues (that vary as a function of both study design and goals) bear on the appropriateness of a given sample size for a given study, hard-and-fast rules about minimum sample sizes cannot be realistically generated, nor should they be. Nevertheless, authors are strongly encouraged to explicitly address the appropriateness of their sample size(s) in light of a study's design and goals, and articles whose samples are judged to be seriously underpowered and that fail to mount a convincing counterargument may be returned to the author without review. Of course, should an article be rejected for reasons of inadequate sample size, there is nothing to prevent the author from resubmitting the article after collecting additional data, and indeed this will be actively encouraged. Finally, authors are also discouraged from the common practice of conducting several small-*n* studies on a topic, when conducting one larger-*n* study would not only yield more accurate, adequately powered parameter estimates but also allow for more sophisticated and probing analyses (e.g., testing how X1 relates to Y in Study 1 and how X2 relates to Y in Study 2, as opposed to conducting a single, larger study in which the joint contribution of X1 and X2 to Y is tested).

Accepting Imperfection in the Data

Another factor contributing to false-positive findings in psychology concerns the unrealistic standards of perfection in one's findings that reviewers and editors often demand as a condition of publication (Giner-Sorolla, 2012; Simmons, Nelson, & Simonsohn, 2011). Whereas imperfect results seem to raise a red flag among reviewers, the exact opposite should be the case. Real data are messy, and imperfection (not perfection) should be expected. This position does, however, pose a potential conundrum for reviewers and editors. If not perfection, then what should the standard be? I would argue that the overall consistency, coherence, and strength of support must be weighed and should take precedence over an isolated result that fails to conform to the general pattern or to exceed the magical $p < .05$ cutoff. (Here I am reminded of Rosnow & Rosenthal's, 1989, contention that, "surely, God loves the .06 nearly as much as the .05," p. 1277). Meta-analytic techniques applied to a series of results obtained across multiple studies within a package (presumably testing a single hypothesis; see, e.g., Cooper et al., 2008) can also be helpful in summarizing and evaluating the overall strength and reliability of a set of findings (although Bem's, 2011, use of this approach in his recent, contested article on ESP underscores the fact that this strategy is not a panacea).

Promoting Exact or Close Replications

Given that replication forms the cornerstone of credible science (Ioannidis, 2012), it seems unquestionable that publishing competently executed replications of important, previously published findings would be good for the field, even though such studies would not necessarily satisfy the long-standing desideratum regarding novelty of contribution. Accordingly, *JPSP:PPID* encourages exact or close replications of previously published personality studies in several formats: (1) as a single stand-alone study, submitted in a short-report format; (2) as a second study in the initial submission of a multistudy package, or (3) as a first study in a multistudy "replication package" aimed at replicating and extending a previously published study. (Authors interested in submitting single replication studies are referred to recently published guidelines for submitting replication studies, which can be found at http://www.apa.org/pubs/journals/psp/?tab=4.)

A recent failure to replicate published in *PLOS ONE* provides a good example of the form "replication packages" might take. Doyen and colleagues (Doyen, Klein, Pichon, & Cleeremans, 2012) attempted an exact replication of a famous study conducted by Bargh, Chen, and Burrows (1996) showing that participants primed with an elderly stereotype walked more slowly upon leaving the experiment. Despite using twice as many participants as the original study and a more precise method to measure speed of walking, they were unable to reproduce the effect, even though the original effect was quite large. Doyen et al. then conducted a follow-up study in which experimenter expectancies were manipulated and the effect was obtained. Although controversy over the Doyen study abounds (see Srivastava, 2012, for a thoughtful discussion), the general point is that whether the exact replication attempt obtains or fails to obtain the originally published effect, subsequent studies in the package can explore the implications of whatever outcome was obtained.

Toward a Broader Definition of What Constitutes Good Scholarship

In closing, I would like to underscore that the goal of these policies is not a lessening of standards, but rather a broadening of the vision of what constitutes good science. Certainly there is much to appreciate and celebrate about the current state of the field, as discussed above, but this should not blind us to the important issues that face our field today, nor deter us from taking needed steps toward a slightly different course. Thus, it seems clear to me that it is time to seriously consider broadening the form, style, and substance of research that our top journals explicitly value and encourage. This is in no way intended to shut out the more traditional theoretically driven, multistudy experimental package that has been *JPSP*s bread and butter, but rather to communicate that a broader and more inclusive spectrum of methodologies and goals are both needed and welcome.

—M. Lynne Cooper, Editor

References

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7,* 543–554. http://dx.doi.org/10.1177/1745691612459060

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71,* 230–244. http://dx.doi.org/10.1037/0022-3514.71.2.230

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100,* 407–425. http://dx.doi.org/10.1037/a0021524

Campbell, D. T., & Cook, T. D. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Cooper, M. L., Krull, J. L., Agocha, V. B., Flanagan, M. E., Orcutt, H. K., Grabe, S., . . . Jackson, M. (2008). Motivational pathways to alcohol use and abuse among Black and White adolescents. *Journal of Abnormal Psychology, 117,* 485–501. http://dx.doi.org/10.1037/a0012592

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE, 7,* e29081. http://dx.doi.org/10.1371/journal.pone.0029081

Funder, D. C. (2009). Naive and obvious questions. *Perspectives on Psychological Science, 4,* 340–344. http://dx.doi.org/10.1111/j.1745-6924.2009.01135.x

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science, 7,* 562–571. http://dx.doi.org/10.1177/1745691612457576

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93*, 216–229. http://dx.doi.org/10.1037/0033-295X.93.2.216

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33,* 61–83. http://dx.doi.org/10.1017/S0140525X0999152X

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7,* 645–654. http://dx.doi.org/10.1177/1745691612464056

LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in the modal research practice. *Review of General Psychology, 15,* 371–379. http://dx.doi.org/10.1037/a0025172

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537–563. http://dx.doi.org/10.1146/annurev.psych.59.103006.093735

McGrath, J. E., Martin, J., & Kulka, R. A. (1981). Some quasi-rules for making judgment calls in research. *American Behavioral Scientist, 25,* 211–224. http://dx.doi.org/10.1177/000276428102500206

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7,* 531–536. http://dx.doi.org/10.1177/1745691612463401

Roberts, B. W. (2011). Personality psychology has a serious problem (and so do many other areas of psychology). *P: The Online Newsletter for Personality Science, 4*. Retrieved from http://www.personality-arp.org/html/newsletter06/problem.html

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist, 44,* 1276–1284. http://dx.doi.org/10.1037/0003-066X.44.10.1276

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review, 5,* 2–14. http://dx.doi.org/10.1207/S15327957PSPR0501_1

Rozin, P. (2009). What kind of empirical research should we publish, fund, and reward? A different perspective. *Perspectives on Psychological Science, 4,* 435–439. http://dx.doi.org/10.1111/j.1745-6924.2009.01151.x

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47,* 609–612. http://dx.doi.org/10.1016/j.jrp.2013.05.009

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Srivastava, S. (2012). *Some reflections on the Bargh-Doyen elderly walking priming brouhaha*. Retrieved from https://hardsci.wordpress.com/2012/03/12/some-reflections-on-the-bargh-doyen-elderly-walking-priming-brouhaha/