## The Seemingly Quixotic Pursuit of a Cumulative Psychological Science: Introduction to the Special Issue

## Patrick J. Curran

University of North Carolina at Chapel Hill

The goal of any empirical science is to pursue the construction of a cumulative base of knowledge upon which the future of the science may be built. However, there is mixed evidence that the science of psychology can accurately be characterized by such a cumulative progression. Indeed, some argue that the development of a truly cumulative psychological science is not possible with the current paradigms of hypothesis testing in single-study designs. The author explores this controversy as a framework to introduce the 6 articles that make up this special issue on the integration of data and empirical findings across multiple studies. The author proposes that the methods and techniques described in this set of articles can significantly propel researchers forward in their ongoing quest to build a cumulative psychological science.

Keywords: cumulative science, philosophy of science, meta-analysis, integrative data analysis

The novel *El Ingenioso Hidalgo Don Quijote de la Mancha* by Miguel de Cervantes tells the story of the relationship between the delusionally idealistic Don Quixote and his world-weary yet practical companion Sancho Panza. Their adventures juxtapose the eager pursuit of unrealistic ideals with practical grounding in the reality of day-to-day life. The main character succeeded in capturing the very nature of idealistic pursuit to the point that over time his name developed into an adjective that describes something as foolishly impractical. So, is it fair to describe the pursuit of a cumulative psychological science as quixotic? On some days, I believe that it is, whereas on others I do not.

Without question, psychology as a field must in some way or another be motivated by a sincere idealistic quest to systematically build a cumulative base of knowledge upon which the science of psychology can progress. Psychology simply cannot survive as a viable science in the absence of these ideals. Yet, at the same time the vicissitudes of real life intervene in many forms, including the need to draw timely conclusions from single-study results as well as the

This work was partially supported by Grant DA013148 awarded to Patrick J. Curran and by Grant DA15398 jointly awarded to Andrea Hussong and Patrick J. Curran. I am indebted to Scott Maxwell for his unwavering support of this project, to Andrea Hussong and Dan Bauer for being argumentative, and to all of the authors who contributed to this special issue for their unwavering dedication in the quest to improve the science of psychology.

Correspondence concerning this article should be addressed to Patrick J. Curran, Department of Psychology, University of North Carolina, Chapel Hill, NC 27599-3270. E-mail: curran@unc.edu

pressures of publication, grant acquisition, and tenure review; psychologists must make practical decisions that may serve them on an individual level yet may not necessarily move the science forward in a cumulative fashion. We thus find ourselves in the position of needing to serve our own interests while contributing to the construction of a more cohesive and cumulative psychological science. In thinking how this might be best accomplished, it is helpful to first consider how these issues play out in the general endeavor of scientific inquiry. An excellent starting point is to consider the work of Thomas Kuhn.

Arguably one of the most impactful works on the philosophy of science in the past century is Kuhn's *The Structure of Scientific Revolutions*, first published in 1962 and revised in 1970 and again in 1996. Kuhn was writing in response to the dominant view of *logical empiricism* that was primarily held in the first half of the 20th century. According to this perspective, science was a highly objective and logical process that led to the systematic construction of a cumulative understanding about the world (e.g., Howard, 1991). This was consistent with the notion that contemporary scientists were "standing on the shoulders of giants," in that the generation of new knowledge was systematically constructed upon the foundation of prior discoveries.

Within this dominant context, Kuhn (1996) offered the provocative (and what was to become highly controversial) perspective that science is indeed cumulative, but only during normal periods of relative calm when existing para-

<sup>&</sup>lt;sup>1</sup> See Mischel (2008a, 2008b) for a wonderful exploration of these same issues.

78 CURRAN

digms allow the creation of knowledge to progress in a systematic fashion. However, as scientific anomalies for which existing paradigms cannot account begin to accrue, tension builds. Existing paradigms attempt to suppress these anomalies, but at some point the tension becomes too great, and there is a paradigm shift in which the old paradigms are not simply augmented by the new ones but instead are outright replaced. Kuhn considered these paradigm shifts to be scientific revolutions and used examples such as moving from an earth-centered to a sun-centered understanding of the solar system. Even nearly half a century after the publication of his original work, skirmishes still break out over whether Kuhn's work implies that the very nature of science is or is not a cumulative process (see, e.g., D'Espagnat, 2008, and Weber, 2008, for recent competing perspectives) or whether Kuhn really set out to kill logical empiricism in the first place (Reisch, 1991).

In the presence of the broader issues underlying the progression of the endeavor of science, psychologists must remain diligently focused on the more practical issues of how they can best support and build a cumulative science. It is interesting that there is as much disagreement about whether psychologists have achieved (or even can achieve) a cumulative science as there is about whether the process of science is a cumulative endeavor. It is even more interesting that some of psychology's harshest critics are, quite appropriately, psychologists. For example, in commenting on the state of psychological inquiry, Meehl (1978) remarked,

It is simply a sad fact that in soft psychology theories rise and decline, come and go, more as a function of baffled boredom than anything else; and the enterprise shows a disturbing absence of that *cumulative* character that is so impressive in disciplines like astronomy, molecular biology, and genetics. (p. 807, italics original)

More recently, Schmidt (1996) stated that the reliance on significance tests in psychological research "has systematically retarded the growth of cumulative knowledge in psychology" (p. 115). Schmidt and Hunter (1997) recommended that researchers "try to show enough intellectual courage and honesty to reject the use of significance tests despite the pressures of social convention to the contrary. Scientists must have integrity" (p. 61). Thus, the dominant issue thought to be responsible for the failure of psychology to progress in a cumulative fashion is the overreliance (or reliance in any fashion) on significance testing in singlesample analysis. Cohen (1990, 1994), Hunter and Schmidt (1996), and Schmidt (1996) all provided excellent discussions of the inherent limitations of this strategy. Yet, singlestudy hypothesis tests remain psychologists' dominant paradigm: Some statistical model is fitted to a single sample of data obtained through a single study, and a go/no-go decision is made about a given effect primarily on the basis of the obtained p value. One widely recommended (although arguably palliative) solution to this problem is to replace significance testing with point estimates and confidence intervals (e.g., Reichardt & Gollob, 1997) and effect sizes (e.g., Cohen, 1988). However, a strong counterargument can be made that this strategy addresses only part of the problem (e.g., Krantz, 1999; Schmidt, 1996). A broader framework for overcoming these issues is to move beyond single-sample studies to the synthesis of findings drawn from multiple studies (e.g., Green & Hall, 1984; Hunter & Schmidt, 1996; Oakes, 1986; Schmidt, 1996).

Among several strategies, meta-analysis is a particularly important approach to research synthesis, and great strides have been made over the past two decades in the development and application of these techniques within the social sciences (e.g., Cooper, Hedges, & Valentine, 2009; Schmidt & Hunter, 1977; Smith & Glass, 1977). Broadly defined, meta-analysis is a set of formalized procedures that allows for the synthesis of summary statistics drawn from a large number of existing studies. One of the original motivations for meta-analysis was that these techniques would further support the creation of a cumulative knowledge within the social sciences, particularly in psychology (e.g., Hunter & Schmidt, 1996; Schmidt, 1984). Indeed, Schmidt (1996) concluded that "unlike traditional methods based on significance tests, meta-analysis leads to correct conclusions and hence leads to cumulative knowledge" (p. 119). There is no doubt that meta-analysis has substantially advanced the science of psychology toward this goal.

However, recall that one of the original motivations underlying the development of meta-analysis is the argument that single-study hypothesis testing has significantly impeded the progress of psychology. Yet, the logical conclusion to this argument is that single-study analyses are useful only to the extent that they might represent a data point in some future meta-analysis. Further, the generalization of any single-study finding supported by traditional hypothesis tests may at best be irrelevant and may at worst be impeding psychology's progress as a science. Thus, psychologists find themselves in a situation in which they have their dominant paradigm of hypothesis testing in single-sample analysis on one end and the meta-analysis of a large number of studies on the other. And, at least in my own reading of the literature, they are each poked with a rather sharp stick to move to one of these two extremes.

Must we all be herded to one end of the pen or the other, or might there be options? Recent analytical and methodological developments reveal a progressively widening and highly captivating interior space that falls between the boundaries defined by single-study analysis and meta-analysis, a space on which psychology has yet to fully capitalize. More specifically, in an increasing number of situations, psychologists may have access to the original raw data drawn from a modest number of individual studies that would be too small to support a meta-analysis but that might

still be combined in some integrated and collaborative fashion to move meaningfully beyond single-study analysis. Could psychologists in some way capitalize on this interior space to progress as a cumulative science while being cognizant of the very real practical issues that arise in applied research? The simple answer is *yes*, but how is such an integration best accomplished?

This seemingly innocuous question quickly morphs into a rather vexing problem. For example, could one simply fit models to data that have been pooled across two or more independent samples? How can this best be accomplished in a reliable and valid way? Would existing concerns about significance testing associated with single-sample analysis simply be generalized to the pooled data analysis? Alternatively, what if independent teams of researchers were pursuing similar programs of research at the same time? Could these be coordinated in a collaborative way to maximize the resulting scientific contributions? Further, might singlestudy analysis, pooled-study analysis, and meta-analysis be conducted so that each would support the others, thus allowing inferences that might not otherwise be possible? The goal of this special issue is to thoroughly explore these questions.

This special issue consists of six articles each of which addresses a different dimension of the methods for combining data in ways that move beyond single-study designs and help contribute to the pursuit of a cumulative psychological science. In the first article, Andrea Hussong and I (Curran & Hussong, 2009) explore issues that arise when fitting models to data that have been pooled over two or more independent samples; we refer to this as integrative data analysis. Next, Dan Bauer and Andrea Hussong (2009) propose a novel application of nonlinear factor analysis that allows for the estimation of measurement models for items that may be scaled differently and drawn from separate samples; they demonstrate this model, using alcohol involvement measures pooled over two different studies. Jack McArdle, Kevin Grimm, Fumiaki Hamagami, Ryan Bowles, and William Meredith (2009) then present a rigorous analysis of life span growth curves of cognition through the simultaneous estimation of item response theory models and latent curve models that are based on multiple scales drawn from multiple samples. Scott Hofer and Andrea Piccinin (2009) describe the design and execution of a collaborative framework for building a cumulative science through the coordinated replication and integration of multiple studies and multiple data sets with a particular focus on research in aging. Next, Harris Cooper and Erika Patall (2009) provide a careful comparison of the relative advantages and disadvantages of current meta-analytic procedures with methods for combining raw data drawn from multiple studies and describe the conditions under which each approach may be beneficial and when they might be used in combination. Finally, Pat Shrout (2009) concludes with a thoughtful discussion of the core issues raised in the prior articles and describes future directions for both quantitative and substantive researchers.

Taken together, this set of articles explores the promising interior space that is defined by the single-study analysis on one extreme and the synthesis of summary statistics from a large set of studies on the other. Of course, both single-study analysis and meta-analysis will always play a critical role as psychology continues to progress as a science. However, the following six articles reveal that much can be gained by navigating the area that lies between these two boundaries. Indeed, I believe that capitalizing on this middle ground will help balance the lofty ideals of Don Quixote with the worldly practicality of Sancho Panza as psychologists continue on their vitally important quest to build a truly cumulative psychological science.

## References

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, 14, 101–125.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (thus far). *American Psychologist*, 45, 1304–1312.

Cohen, J. (1994). The earth is round (p < .05). American Psychologist, 49, 997–1003.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The hand-book of research synthesis and meta-analysis* (2nd ed.). New York: Sage.

Cooper, H., & Patall, E. A. (2009). The relative benefits of metaanalysis conducted with individual participant data versus aggregated data. *Psychological Methods*, 14, 165–176.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14, 81–100.

D'Espagnat, B. (2008). Is science cumulative? A physicist view-point. In L. Soler, H. Sankey, & P. Hoyningen-Huene (Eds.), Rethinking scientific change and theory comparison: Stabilities, ruptures, incommensurabilities? (pp. 145–151). New York: Springer.

Green, B. F., & Hall, J. A. (1984). Quantitative methods for literature reviews. Annual Review of Psychology, 35, 37–53.

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*, 14, 150–164.

Howard, D. (1991). Einstein, Kant, and the origins of logical empiricism. In W. Salmon & G. Wolters (Eds.), *Logic, language, and the structure of scientific theories* (pp. 45–106). Pittsburgh, PA: University of Pittsburgh Press.

Hunter, J. E., & Schmidt, F. L. (1996). Cumulative research

80 CURRAN

knowledge and social policy formulation: The critical role of meta-analysis. *Psychology, Public Policy, and Law, 2,* 324–347.

- Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association*, 44, 1372–1381.
- Kuhn, T. S. (1996). The structure of scientific revolutions (3rd ed.). Chicago: University of Chicago Press.
- McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods*, *14*, 126–149.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow process of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806–834.
- Mischel, W. (2008a, October). Our urban legends: Publishing. *APS Observer*, *21*, Retrieved April 9, 2009, from http://www.psychologicalscience.org/observer/getArticle.cfm?id=2403
- Mischel, W. (2008b, December). The toothbrush problem. *APS Observer*, 21. Retrieved April 9, 2009 from http://www.psychologicalscience.org/observer/getArticle.cfm?id=2430
- Oakes, M. L. (1986). Statistical inference: A commentary for the social and behavioral sciences. New York: Wiley.
- Reichardt, C. S., & Gollob, H. F. (1997). When confidence intervals should be used instead of statistical significant tests, and vice versa. In L. Harlow, S. Mulaik, & J. Steiger (Eds.), What if there were no significance tests? (pp. 259–286). Mahwah, NJ: Erlbaum.
- Reisch, G. A. (1991). Did Kuhn kill logical empiricism? *Philosophy of Science*, 58, 264–277.

- Schmidt, F. L. (1984, August). Meta-analysis: Implications for cumulative knowledge in the behavioral and social sciences.
   Paper presented at the annual convention of the American Psychological Association, Toronto, Ontario, Canada.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, *1*, 115–129.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, 62, 529–540.
- Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. Harlow, S. Mulaik, & J. Steiger (Eds.), What if there were no significance tests? (pp. 37–64). Mahwah, NJ: Erlbaum.
- Shrout, P. E. (2009). Short and long views of integrative data analysis: Comments on contributions to the special issue. *Psychological Methods*, *14*, 177–181.
- Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752–760.
- Weber, M. (2008). Commentary on "Is science cumulative? A physicist viewpoint," by Bernard D'Espagnat. In L. Soler, H. Sankey, & P. Hoyningen-Huene (Eds.), Rethinking scientific change and theory comparison: Stabilities, ruptures, incommensurabilities? (pp. 153–156). New York: Springer.

Received March 31, 2009
Revision received March 31, 2009
Accepted March 31, 2009