

Preface

Hypothesis Testing: A Challenge for Undergraduate Researchers

Richard G. Coss

Professor, Department of Psychology

Recipient of the 2008 Chancellor's Award for Excellence in Mentoring Undergraduate Research

Ever since Darwin first broached the subject directly (Darwin, 1871), there has been a continuing debate in psychology over the role of evolutionary processes in shaping human evolution and the manner in which evolutionary hypotheses are tested empirically. I have noticed in the course of teaching research methodology and hypothesis testing to undergraduate students over the last three decades that the frequent criticisms leveled at evolutionary hypotheses have increased students' confusion about how to evaluate the merits and successes of evolutionary hypotheses.

But this confusion is not unique to psychology as a discipline nor to the study of evolutionary processes in a variety of fields. Rather, the challenge is one that students in many disciplines face: how to distinguish between research that aims to construct *new* hypotheses and research that aims to confirm or disconfirm *existing* hypotheses.

In the beginning phase of conducting research, students introduced to the scientific method are often confused about the complementarity in these two avenues of research. In my field, I typically assign journal articles to read that describe field studies that emphasize quantitative animal-behavior research. With fewer constraints on page restrictions, older scientific articles often explored different facets of a problem that typically included elaborate anecdotal observations. These studies often addressed scientific questions without the researcher(s) having framed specific predictions that could be confirmed or disconfirmed by the experimental findings. Students often accept the verisimilitude of anecdotal behavioral observations, or of quantitative data with statistical indices of reliability, without realizing that these findings only provided the preliminary backdrop for a particular line of enquiry. Although these studies might have insufficient information to allow novel predictions, due to their preliminary nature, they nevertheless might provide valuable information for developing follow-up projects that address specific hypotheses derived from this preliminary research.

For example, Kenneth Armitage, a well-respected researcher who has spent many years in the field studying marmot behavior, published a descriptive paper on how badgers hunt marmots in Colorado (Armitage, 2003). Armitage compared the time of day badgers hunted and how marmots changed their pattern of alarm calling during and after they encountered badgers. Both comparisons show reliable differences statistically. This seminal paper addressed questions about how badgers hunted and whether marmots became markedly less vigilant after badgers left the area. No specific hypotheses were tested, so while this paper remains purely descriptive, it contains reliability indices of behavioral changes that are useful for future hypothesis development. Nevertheless, students reading this paper might not differentiate the examination of exploratory questions using statistical analyses from true hypothesis testing, in which predictions were made before the study was conducted and statistical tests were used to examine these predictions.

In essence, the Armitage paper provides a context in which students might consider the statistical analyses of data as empirical results, illustrating the difficulty of distinguishing *statistical hypotheses*—those relying on statistical analysis to determine measurement reliability—from *theoretical hypotheses*, which

are either supported or refuted by experimental findings. This conflation of statistical and theoretical hypotheses is not uncommon in published research articles in many scientific disciplines. A case in point is typified by experimental papers that use the term *null hypothesis* as the alternative theoretical hypothesis, when this term in statistics really has the precise meaning that there is no relationship between two or more variables. If experienced researchers fail to make this distinction and editors and reviewers do not flag this discrepancy, it is especially difficult for students to avoid this conflation of statistical and theoretical hypotheses.

A third facet of hypothesis testing relevant to student learning is the idea that theoretical hypotheses are never entirely rejected by current experiments; rather, current experiments refute previous hypotheses, weakening them severely enough that they are ignored by the scientific community unless newer experimental findings alter this assessment (Meehl, 1990a; Popper, 1959). But the value of not rejecting previous research findings—even those that appear to have been refuted—can be seen in the case of Freud’s theories on how to guide the free associations of patient dreams during psychoanalytic interviews. These complex interview processes still await the development of more suitable statistical analyses to evaluate their validity (Meehl, 1990b).

These are the cautions. Even as we teach students what *not to do*, however, we must simultaneously teach them what *to do*. The approach that I recommend to my students in evolutionary research has been well phrased by Conway and Schaller (2002):

Start with a plausible assertion about some set of...constraints and consider the various consequences that follow logically from that assertion. In doing so, it is valuable to ignore easy temptations to arrive at previously discovered or intuitively obvious destinations and instead to try to find routes to subtle, surprising, previously uncharted...territory. (p. 163)

The testing of formal hypotheses was originally proffered by the philosopher Francis Bacon (1620/1939) and later by Chamberlin (1890), who reasoned that testing alternative hypotheses reduced the likelihood of researchers becoming attached to a single hypothesis. In further elaboration, Platt (1964) argued that science advances more quickly when two competing hypotheses are tested simultaneously, leading to “strong inferences” of support for one hypothesis if the other is disconfirmed. Although testing two alternative hypotheses can guide the direction of experimental research, notably in the physical sciences, its utility in behavioral research can be muddled by incomplete information for framing the alternative hypothesis. In my experience mentoring undergraduate students, the testing of a single, well-structured evolutionary hypothesis requires a considerable investment by students in reading the relevant literature.

With regard to the evolutionary theme of my research-mentoring activities, it is the investment of time students spend in perusing the literature—and the nature and quality of the process of how students develop evolutionary hypotheses from the research they read, especially when they pay close attention to those meaningful conjectures that researchers sometimes make while interpreting the experimental findings in the discussion sections of their papers (see Meehl, 1990b)—that determines the soundness of their own hypotheses. An additional facet that students need to be aware of is that evolutionary hypotheses may be generated through induction, conjecture, or even in some fields, *post-hoc* evolutionary explanations. However they are derived, good evolutionary hypotheses have the ability to “move beyond their explanatory role and generate novel predictions that can be tested empirically” (Coss & Charles, 2004) and to lead to progressive research that continues to make novel predictions (see Lakatos, 1970).

The mentoring of undergraduate students who are conducting hypothesis-driven research requires consistent interactions between mentors and students to ensure that these students clearly understand the aforementioned attributes and pitfalls when they create testable hypotheses. Both the mentoring and

student responsiveness pose important pedagogical challenges if we are to say that we are successful in teaching the scientific method.

* * *

The seven papers presented in this year's edition of *Explorations* attest to the success of teachers across many disciplines in helping students learn the challenges of the experimental process. And they testify to the fact that sound research is being produced by undergraduates at UC Davis.

Each of the authors in this year's volume explores some aspect of the human interplay with/in the rest of nature. Theresa DiMarco examines ultraviolet radiation and its effects on the behavior of sea anemones, while Shikha Chetal analyzes the efficacy of recycling programs and Anthony Ng considers the debate over how best to reduce carbon emissions. Looking backward in time, Eric Sneathen grapples with the political ecology of a nineteenth century work of social satire, while Lisa Lin dissects European Enlightenment standards of taste to explain the reaction of western observers to the Indian practice of widow-burning. Exhibiting two of the many applications of computer technology, Matthew Schultz models circadian rhythms in plants, while Megan Gordon-Gilmore examines a piece of art that both represents and creates the computer-mediated environment in which so much human interaction occurs today.

The contributions of these students to the bodies of research examining the natural and human worlds would not have been possible without the dedicated efforts of many professors, lecturers, researchers, and graduate students at UC Davis and elsewhere. This year, the contributors and editors wish especially to acknowledge Professors Terence Murphy and Julin Maloof (Plant Biology), Frances Dyson (Cultural Studies), Omnia El Shakry (History), Larry Berman (Political Science, UCD Washington Program), Timothy Morton (English), Steven Morgan (Environmental Science and Policy, Bodega Marine Laboratory), Gary Cherr (Nutrition, Environmental Toxicology, Bodega Marine Laboratory), and Brent Sewall (PhD candidate, Ecology; UCD Washington Program).

Funding for *EXPLORATIONS* has been generously provided by the Undergraduate Research Fund of the UC Davis "Students First" campaign of 1993.

References

- Armitage, K. B. (2003). Badger predation on yellow-bellied marmots. *American Midland Naturalist* 151, 378–387.
- Bacon, F. (1620/1939), *Novum Organum*. In E. A. Burt (Ed.), *The English philosophers from Bacon to Mill* (pp. 24-128). New York: Random House.
- Chamberlin, T. C. 1890. The method of multiple working hypotheses. *Science*, 15, 92-96.
- Conway, L. G., III, & Schaller, M. (2002). On the verifiability of evolutionary psychological theories: An analysis of the psychology of scientific persuasion. *Personality and Social Psychology Review*, 6, 152–166.
- Coss, R. G., & Charles, E. P. (2004). The role of evolutionary hypotheses in psychological research: Instincts, affordances, and relic sex differences. *Ecological Psychology*, 16, 199-236.
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: Murray.
- Lakatos, I. (1970). Falsificationism and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). Cambridge, England: Cambridge University Press.

- Meehl, P. E. (1990a). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108-141.
- Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195-244.
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347-353.
- Popper, K. R. (1959). *The logic of scientific discovery*. London: Hutchinson.