Special Section: Statistical Training and Curricular Revision

What Is Statistics?

Emery N. BROWN and Robert E. KASS

We use our experience in neuroscience as a source of defining issues for the discipline of statistics. We argue that to remain vibrant, the field must open up by taking a less restrictive view of what constitutes statistical training.

KEY WORDS: Cross-disciplinary statistical research; Statistical paradigm; Statistical thinking.

1. SHORT SUPPLY

Our field faces fundamental challenges. The statistical needs of science, technology, business, and government are huge and growing rapidly, producing a shortfall in statistical workforce production. In their summary of an National Science Foundation workshop, *The Future of Statistics*, Lindsay, Kettenring, and Siegmund (2004) reported that

> Workshop participants pointed repeatedly to shortages in the pipeline of students and unmet demand from key industries and government laboratories and agencies. . . . The shortage may prove quite damaging to the nation's infrastructure.

The growth in demand for data analysis may be attributed in large part to the exponential increase in computing power and data collection capabilities. At the same time, there is a worrisome tendency for quantitative investigators or technical staff to attack problems using blunt instruments and naive attitudes. Our discipline as a whole has been gloriously productive, making available a wide variety of tools. But we have been less successful in producing easy-to-master operating instructions and training programs. We have effectively created a supply side of the problem: Statistical education has not been sufficiently accessible. Curricula in statistics have been based on a now-outdated notion of an educated statistician as someone knowledgeable about existing approaches to handling nearly every kind of data. Degrees in statistics have emphasized a large suite of techniques, and introductory courses too often remain unappetizing. The net result is that at every level of study, gaining statistical expertise has required extensive coursework, much of which appears to be extraneous to the compelling scientific problems students are interested in solving.

We also must acknowledge that some of the most innovative and important new techniques in data analysis have come from researchers who would not identify themselves as statisticians. Computer scientists have been especially influential in the past decade or so. The influx of methodology from outside the discipline is not new; indeed, the field of statistics itself is relatively young, with much foundational achievement predating the advent of departments of statistics. But an undeniable fear lurks in the hearts of many statistics professors: As others leap daringly into the fray, attempting to tackle the most difficult problems, might statistics as we know it become obsolete?

The two of us recently co-organized the fourth international workshop on statistical analysis of neural data. This series of conferences has brought together quantitatively oriented experimenters and cutting-edge data analysts working in the field of neuroscience, offering new challenges for statistical science in the process. We and others have found the high quality of statistical application gratifying and the articulation of new ideas very stimulating. One of the reactions from readers of our grant proposal to the National Science Foundation took us by surprise, however. Only a relatively small minority of our speakers and participants came from departments of statistics, and as a result, some reviewers questioned whether the Division of Mathematical Sciences should be supporting this activity. Luckily, the program officers handled this issue adeptly, in part by getting cosponsorship from Computational Neuroscience. But the issue is an aspect of the existential identity crisis; the reviewers were grappling with the vexing question, raised by institutional structures, of who should be counted as a statistician.

The participation in neuroscientific research of many non-statisticians doing sophisticated data analysis is not surprising. The brain is considered a great scientific frontier. Studying it creates many technological challenges, and because neuronal networks form electrical circuits, fundamental contributions to neurophysiology have been made by physical arguments, in the

Emery N. Brown is Warren M. Zapol Professor of Anaesthesia, Harvard Medical School, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA 02114 and Professor of Computational Neuroscience and Health Sciences and Technology, Department of Brain and Cognitive Sciences, MIT-Harvard Division of Health Science and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139 (E-mail: *enb@neurostat.mit.edu*). Robert E. Kass is Professor, Department of Statistics, Center for the Neural Basis of Cognition, and Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15217 (E-mail: *kass@stat.cmu.edu*). The thoughts herein have resulted from many extended discussions with colleagues, especially in the Department of Statistics at Carnegie Mellon, where Kass was Department Head from 1996 to 2005 and Brown serves on the external advisory board. Brown's research was partially supported by Grants DP1 OD003646, R01 MH59733, and R01 MH071847. Kass' research was partially supported by Grants RO1 MH064537, RO1 EB005847, and R90 DA023426.

 DOI: 10.1198/tast.2009.0019

form of differential equations. Furthermore, brain science is where artificial neural network models arose, not as machines for nonparametric multiple regression, but rather as descriptors of cognitive mechanisms. For these reasons, neuroscience has attracted many researchers trained in quantitative disciplines, especially physics and engineering. Although their activities might make some statisticians nervous when it comes to federal grants and other resources, a more serious threat is a disciplinary attitude that contrasts strikingly with what we see among many statisticians. Physicists and engineers very often become immersed in the subject matter. In particular, they work hand in hand with neuroscientists and often become experimentalists themselves. Furthermore, physicists and engineers (and likewise computer scientists) are ambitious; when faced with problems, they tend to attack, sweeping aside impediments stemming from limited knowledge about the procedures that they apply. In seeing this, we often shudder, and we criticize this cavalier attitude later in this article. But there is a flip side to our reaction; in contrast, we find that graduate students in statistics often are reticent to the point of inaction. Somehow, in emphasizing the logic of data manipulation, teachers of statistics are instilling excessive cautiousness. Students seem to develop extreme risk aversion, apparently fearing that the inevitable flaws in their analysis will be discovered and pounced upon by statistically trained colleagues. Along with communicating great ideas and fostering valuable introspective care, our discipline has managed to create a culture that often is detrimental to the very efforts it aims to advance.

We are worried. While we expect that in many institutions—perhaps most—there may exist specific courses and programs that are exemplary in certain respects, in the aggregate, we are frustrated with the current state of affairs. The concerns that we have articulated here are not minor matters to be addressed by incremental improvement; rather, they represent deep deficiencies requiring immediate attention.

2. CHANGING TIMES

In making critical comments, we hope to stir discussion and debate. We do not wish to be misunderstood, however; our most fundamental loyalty is to the discipline of statistics. We appreciate its role in technical advances over the past century, and see even greater opportunities for essential contributions in the future, as scientific investigations rely on more massive and intricate data sets to examine increasingly complex phenomena. Furthermore, besides utility, there is great beauty in the subject. We have spent considerable effort learning and trying to advance neuroscience. But even after substantial exposure to one of the most exciting and rapidly developing areas of science, we still believe that statistics, with its unique blend of real-world mathematics, epistemology, and computational technique, is the most deeply interesting and rewarding of all intellectual endeavors. There are strong arguments to suggest that much of cognition is based on pattern learning, and that humans have well-developed neural machinery for making inferences implicitly, without conscious recognition. Perhaps part of the pleasure that we get from statistical reasoning comes from bringing a harmonious coherence to otherwise unappreciated brain processes. Regardless of its biological explanation, however, there is certainly an inspiring aesthetic of statistics driven in part by the emotional overlay of trying to tame uncertainty. The problem is not with the nature of the discipline. There are compelling reasons to love statistics and to pass on to others both knowledge of its methods and appreciation of its powerful logic.

So where have things gone wrong? We believe that the primary source of the current difficulties is an anachronistic, yet pervasive conception of statistics. The problem is that departments of statistics often act as if they are preparing students to be short-term consultants, able to answer circumscribed methodological questions based on limited contemplation of the context. This short-term consultant model relegates the statistician to a subsidiary position, and suggests that applied statistics consists of handling well-formulated questions, so as to match an accepted method to nearly any kind of data. This notion may have developed partly because—at least in the United States—statistics evolved from mathematics with its lone investigator, and partly because a qualified statistician could know the entire field. The large majority of senior statisticians began their academic careers as math majors. Within statistics departments, mathematical thinking influenced both research and infrastructure, whereas the mathematics involved was relatively limited, so that Ph.D. statisticians could master the technical details in diverse areas of statistics. Graduate programs thus emphasized mathematically thorough knowledge of multiple branches of the field. At one time, this served a useful purpose. But statistics has expanded and deepened, so that individuals rarely have state-of-the-art, rigorous expertise in more than a few well-developed subdomains. Furthermore, in today's dynamic and interdisciplinary world, success in confronting new analytical issues requires both substantial knowledge of a scientific or technological area and highly flexible problem-solving strategies. In neuroscience, for example, a statistician will have far more impact once he or she is able to generate ideas for scientific investigation. In other fields, the situation is surely analogous. The discipline of statistics needs to recognize our new situation and act accordingly. We suggest two overarching principles of curricular revision.

3. A FOCUS ON STATISTICAL THINKING

According to syllabi and lists of requirements, statistics courses and degree programs tend to emphasize mastery of technique. But statisticians with advanced training and experience do not think of statistics as simply a collection of methods; like experts in any field, they consider their subject highly conceptual. This deserves emphasis, because it distinguishes a disciplinary approach from efforts that might be disparaged as the work of amateurs. In neuroscience, we have seen many highly quantitative researchers trained in physics and engineering, but not in statistics, apply sophisticated techniques to analyze their data. These often are appropriate and sometimes are inventive and interesting. In the course of perusing many, many articles over the years, however, we have found ourselves critical of much published work. Starting with vague intuitions, particular algorithms are concocted and applied, from which strong scientific statements are made. Our reaction too often is negative;

we are dubious of the value of this approach, believing that alternatives are preferable. Or we may concede that a particular method possibly may be a good one, but the authors have done nothing to indicate that it performs well. In specific settings, we often come to the conclusion that the science would advance more quickly if the problems were formulated differently—in a manner more familiar to trained statisticians. As an example, neuroscientists developed the highly intuitive "spike-triggered average" to identify an association between a neural spike train, which may be considered a point process, and a continuous stimulus. Point process analysis by a member of Columbia's Department of Statistics (Paninski 2003) has shown that spike-triggered averaging can be inconsistent in some realistic settings, but that consistent estimators may be constructed using generalized linear (or nonlinear) regression models, an approach first championed by Brillinger. (For related references and other examples, see Brown, Kass, and Mitra 2004; Kass, Ventura, and Brown 2005.)

The statistician's perspective, missing from much analysis of neural data, is the most important thing that we can provide. Once students have it, they will be empowered in diverse situations. Thus, we suggest that the primary goal of statistical training at all levels should be to help students develop *statistical thinking.*

What exactly do we mean by this? Different statisticians would use somewhat different words to describe what defines the essential elements of our discipline's approach, but we believe there is general consensus about the substance, which can be stated quite concisely. Statistical thinking uses probabilistic descriptions of variability in (1) inductive reasoning and (2) analysis of procedures for data collection, prediction, and scientific inference. For instance, a prototypical description of variability among data pairs $(x_1, y_1), \dots, (x_n, y_n)$ is the nonparametric regression model

$$Y_i = f(x_i) + \varepsilon_i,$$

in which each ε_i is a random variable. This may be used to suggest methods of smoothing the data and to express uncertainty about the result [both of which are part of item (1)] and also to evaluate the behavior of alternative smoothing procedures [item (2)]. One can dream up a smoothing method, and apply it, without ever referencing a model—indeed, this is the sort of thing that we witness and complain about in neuroscience. Meanwhile, among statisticians there is no end of disagreement about the details of a model and the choice among methods (What space of functions should be considered? Should the ε_i random variables enter additively? Independently? What class of probability distributions should be used? Should decision-theoretic criteria be introduced, or prior probabilities?). The essential component that characterizes the discipline is the introduction of probability to describe variation in order to provide a good solution to a problem involving the reduction of data for a specified purpose. This is not the only thing that statisticians do or teach, but it is the part that identifies the way they think. We provide a bit more discussion of this notion in the Appendix.

Currently, statistical thinking is internalized as a byproduct of extensive statistical training. Elevating it to an overarching goal allows curricula to be assessed according to the way in which statistical thinking is engendered.

4. FLEXIBLE CROSS-DISCIPLINARITY

Contemporary students see before them a world dominated by "big science," with a host of exciting paths to participate in progress. Many students recognize a fundamental role for statistics, and most see great value in learning statistical methods, but they are increasingly motivated by a desire to solve important problems. In this context, the very best quantitatively oriented students often come from other quantitative disciplines, including computer science, physics, and engineering, and they have many options.

As an example, because of his involvement in computational neuroscience at Carnegie Mellon, one of us (Kass) became aware of an outstanding senior undergraduate, a young woman majoring in computer science at one of the top liberal arts colleges, with nearly perfect GPA and GRE score. She was very interested in computational aspects of neuroimaging and wanted to pursue a Ph.D. However, she had never taken a statistics course, and in fact had taken only one math course beyond calculus. It had not occurred to her that statistics might be a good option, and, from the standpoint of admission to a graduate program in statistics, she presented logistic complications; it was not clear exactly what she would study, or how many years it would take to complete her degree. We must make room for students like this and recruit them.

To attract students with nontraditional quantitative backgrounds, statistics programs must guide these students toward making important contributions in a timely manner. Cross-disciplinary projects will have to play a major role. Once a department accepts as its primary mission helping students develop an ability to think like statisticians, it is freed from the constraints of excessive content and can recognize alternative ways that students can demonstrate their abilities and achievements. On the one hand, we see cross-disciplinary work as essential to anyone with any kind of statistical credentials—and thus to statistical training at every level. On the other hand, we view cross-disciplinary research as an opening to students of varied backgrounds—a way of welcoming them into the fold and a mechanism for streamlining training, making programs more manageable and the discipline more inviting.

To satisfy different kinds of students, programs also must allow multiple pathways toward degrees. Increasing the emphasis on cross-disciplinarity goes hand in hand with reducing the importance of particular courses and thereby decreases programmatic rigidity. Flexibility is paramount. We do not wish to remove theoreticians from our midst; indeed, many nonmathematicians will blossom in theoretical directions. Rather, our aim is to allow a broader notion of who counts as a statistician.

5. IMPLICATIONS

If someone is able to (i) appreciate the role of probabilistic reasoning in describing variation and evaluating alternative procedures and (ii) produce a cutting-edge cross-disciplinary analysis of some data, should we feel comfortable calling that person a statistician? We think so, and we would like to see our profession broaden its perspective to a sufficient degree to make this possible.

We further believe that it is consequential to declare (i) and (ii) to be defining goals for a training program. In applying this at the graduate level, however, we presume that to do "cutting edge" work, along the way a trainee would have had to have learned something about classical techniques, such as linear regression, some area of modern statistics (e.g., nonparametric regression, dimensionality reduction, graphical models), and also general inferential tools, such as the bootstrap and Bayesian methods. Furthermore, appreciation of probabilistic reasoning comes from repeated exposure to it in varied contexts. Both of these require mathematical and computational skills. Thus, we are proposing variations on what is currently in place in training programs throughout the country; each training program formulates (explicitly or implicitly) a list of skills and units of knowledge that are truly essential, and figures out how the items on the list are to be taught and evaluated. What constitutes inculcation of statistical thinking may be in the eye of the beholder—in this case, the departmental training program. On the other hand, we have argued that the status quo is unacceptable. Here are four recommendations.

1. *Minimize prerequisites to research*. There are continual disagreements about the stage at which trainees should do research. We strongly favor making cross-disciplinary projects widely available, even to those with minimal backgrounds. Although advanced trainees will have more tools at their disposal, talented quantitatively oriented students can quickly learn how to apply and interpret statistical techniques without formal coursework—indeed, we witness this repeatedly in neuroscience. There has been a tendency in statistics to have students first understand, then do. But this sequence can be reversed, giving a statistical faculty supervisor the opportunity to demonstrate in practice the value of knowing the theoretical underpinnings of methodology. Perhaps most importantly, as we stated earlier, students who want to solve real problems will be attracted to cross-disciplinary research. At both the graduate and undergraduate levels, exciting research opportunities are likely to be among the best recruitment tools.

2. *Identify ways of fostering statistical thinking*. How should we help our students internalize a principled approach to data collection, prediction, and scientific inference? Appreciation of statistical thinking should begin in introductory courses. Each instructor of a first course in statistics grapples with ideas behind reasoning from data, and much effort has gone into texts for such classes. Although we recognize the many great strides taken by textbook authors, we are not entirely satisfied with the typical content of introductory courses. For example, in teaching young neurobiologists, we have found it helpful to stress the value of probabilistic reasoning through propagation of uncertainty via simulation methods—as in bootstrap confidence intervals or Bayesian inference—and to emphasize "principles" by including explicit discussion of mean squared error. Both topics seem more advanced than what is usually found in elementary texts. To be attracted to the subject, however, the most gifted students must see it as deep, with serious theoretical content. Courses tend to be categorized as either theoretically oriented for math/statistics majors or method-oriented "service courses" for other disciplines, and we find too little similarity between the two. The main point here is that the first college-level exposure to statistics matters. Although for pedagogical purposes, central ideas must remain simple and approachable, we believe that it is important to represent the discipline as being rich in profound concepts. More fundamentally, one goal of every first course in statistics for quantitatively capable students should be to interest some of the students in further study.

At the graduate level, existing curricula succeed in getting students to think like statisticians, but focus on this goal is necessary if programs are to be streamlined. Students will still need exposure to statistical reasoning in multiple diverse settings, together with emphasis on (a) the roles of heuristics, computational considerations, and/or generative models in producing procedures and (b) theoretical performance, balanced by convenience, computational efficiency, and interpretability. Many excellent books on such topics as nonparametric regression, density estimation, time series analysis, and Bayesian methods offer very good comparative discussions combining both theoretical and practical concerns. The only problem we see is that they are designed for full-semester courses, whereas in many cases the modern student may wish to devote only a couple of weeks to each within formal course work. We believe that there is an important place for courses, and texts, that give quick impressions while reinforcing underlying principles.

We also take it for granted—but nonetheless believe it worth mentioning— that training programs at every level should include many opportunities for trainees to interact with experienced statisticians (in, e.g., journal clubs, informal seminars, social events), partly to see how they think about problems, but also to have role models reinforce the joys and benefits of pursuing statistics.

3. *Require real-world problem solving*. Experienced statisticians spend much of their collaborative time trying to understand the nature of the data collection process and its relationship to scientific or technological issues. Some students, especially those with backgrounds in experimental science, tend to be well prepared in this dimension, asking appropriate questions, digging up background material, and readily grasping the big picture. Many others, however, have difficulty making connections among scientific ideas, the resulting data, and appropriate analytic strategies. Having recognized this basic skill for applied statistics, we must help our students develop it. Several methods for doing so exist. Project courses, especially at the undergraduate level, can be helpful. Extended research projects—learning by doing—can of course be among the best ways to develop problem-solving skills. An important caveat, however, is that some projects are so well formulated that execution becomes straightforward, and little effort toward big-picture comprehension is needed. We come across students who in the course of doing statistical analyses exhibit remarkably little curiosity about the material they are analyzing. Most likely this is because they have not been taught a systematic approach to problem solving and do not appreciate the payoff from pursuing it.

4. *Encourage deep cross-disciplinary knowledge*. In neuroscience, as elsewhere, statistical training can shape how data lead to useful knowledge. Once the information obtainable from an experiment is clearly understood, a new aspect of the scientific landscape may come into view. Consequently, statisticians

can make major contributions by redefining problems and redirecting data-collection efforts.

In this regard, we distinguish two alternative roles. The first role has been played by both of us; like other senior statisticians in varied domains, we have spent many years learning scientific principles and methods and building collaborations with colleagues, so that our suggestions for research problems and approaches are taken seriously and often followed. The second role requires a deeper commitment to cross-disciplinary training, however. One of us (Brown) became a practicing anesthesiologist in addition to being a statistician. As a result of his extensive physiological knowledge and expertise, he has been able to create a laboratory and is undertaking a series of experiments on brain activity to describe how anesthetic drugs produce the state of anesthesia. Many others in the profession play a similar "principal investigator" role. Two examples are John Quackenbush in the Biostatistics Department in the Harvard School of Public Health and the Dana Farber Cancer Institute, who formulates and executes experiments that use genomic and computational approaches to study networks and pathways in cancer development and progression, and Wing Wong in the Department of Statistics at Stanford University, who conducts experiments on developmental genomics and signal transduction that are informed by statistical considerations.

Faculty who run extradisciplinary experiments and contribute to disciplinary methodology are becoming fairly common in engineering and physics, but not in statistics. The change in attitude that we advocate should in time produce more such people in departments of statistics. In addition to accepting the desirability of these appointments, however, more joint training programs are needed. As models in neuroscience, we can point to our own institutions. The Harvard/MIT Health Sciences and Technology Ph.D. program trains students in quantitative subjects while also having them take substantial medical school courses and serve on rotations in the hospital as a medical student would. Carnegie Mellon's Ph.D. Program in Neural Computation is similar, requiring mastery of a technical discipline (e.g., computer science or statistics) together with multiple courses in the brain sciences, and rotation through experimental laboratories. Again, to attract large numbers of students, course requirements in interdisciplinary programs must be stripped down to manageable essentials. We would like to see more such joint programs that offer credentials in statistics.

6. DISCUSSION

The report by Lindsay, Kettenring, and Siegmund (2004) was aimed at the general community of mathematical scientists. Our discussion has been inward-looking, and critical. Although there is much to be admired in statistical training programs throughout the world, we accuse them of harboring obsolete attitudes about the nature of statistics. Statistics is a wonderful field, but the way in which statisticians view it must evolve. We have suggested defining what our discipline brings to the table, labeling the perspective that we believe to be so fundamentally valuable "statistical thinking." We also have advocated greater encouragement of cross-disciplinary training. Deepening cross-disciplinary involvement and welcoming more experimentalists and other practitioners into the clan of statisticians need not diminish the importance of the theoretical core. Quite the contrary; those with hands-on knowledge of context-driven issues can help identify methodological problems, prodding theory to advance in productive new directions.

Our first main message is that training programs should have a clearer notion of what they intend do. The second message is that these programs generally need to strengthen and deepen their commitment to cross-disciplinary work. In this, we follow many others. We have emphasized the contrast between short-term consulting and deeper, long-term engagement, which require different attitudes and skills. We are sympathetic to the promise made by Birnbaum (1971) that "each student of statistics working with me at any level shall also work systematically with another study adviser representing a scientific or technological research discipline of interest to the student," and we agree with Gnanedesikan (1990) that training should focus less on defining the appropriate encompassing content and more on instilling a relevant sense of cross-disciplinary curiosity: "We need a switch turned on, a value established, for impelling statisticians to be challenged intellectually and through a desire to contribute to solving major problems in other fields."

The worth of cross-disciplinary work, and its essential role in stimulating new statistical theory and methods, seems to be much more widely appreciated now than in the past. We want to push harder, partly because we feel that curricular ramifications have not been given sufficient attention, but also because the world needs more statistically oriented scientific principal investigators. Such scientific leadership is, again, not just a recent development. As one example, in the mid-1970s, Fred Mosteller, a master at initiating interdisciplinary collaborations on topics he deemed scientifically important, became interested in the benefits of surgical therapies, which typically are not studied using randomized controlled clinical trials. This led to his formulation of a large research effort involving statisticians, surgeons, anesthesiologists, and public health specialists to investigate the costs, risks, and benefits of surgery (Bunker, Barnes, and Mosteller 1977). Mosteller was not trained in surgery, but he was clearly the intellectual leader of the project. This kind of leadership is not limited in any way to areas in which "principal investigator" has a literal meaning in a biomedical context. As emphasized by Keller-McNulty (2007), many of today's big challenges throughout society are tackled by large teams, and these teams are in desperate need of statistical thinking at the very top levels of management. We suggest that a way forward begins with a focus on the fundamental goals of training, combined with a broad vision of the discipline of statistics.

APPENDIX: WHAT IS STATISICAL THINKING?

Snee (1990) noted that "many of us talk about statistical thinking but rarely define it." Although the field is so broad that a single notion of statistical thinking cannot possibly be universally applicable, we provided above a succinct definition coming from our own experience that we believe articulates a widely held consensus. We are, at least, in line with Rubin (1993) when he said that

> The special training statisticians receive in mapping real problems into formal probability models, computing inferences from the data and models, and exploring the adequacy of these inferences, is not really part of any other formal discipline, yet is often crucial to the quality of empirical research.

Similarly, Mallows (1998) wrote that

> Statistical thinking concerns the relation of quantitative data to a real-world problem, often in the presence of variability and uncertainty. It attempts to make precise what the data has to say about the problem of interest.

In combining these points of view, we wished to recognize the centrality of probabilistic reasoning while distinguishing two roles for it. First, there is the inductive movement from description of variation to expressions of knowledge and uncertainty. A probabilistic description of variation would be "the probability of rolling a 3 with a fair die is $1/6$," whereas an expression of knowledge would be "I'm 90% sure that the capital of Wyoming is Cheyenne." These two sorts of statements, which use probability in different ways, are sometimes considered to involve two different kinds of probability, called "aleatory probability" and "epistemic probability." Bayesians merge these, applying the laws of probability to go from quantitative description to quantified belief, but in every form of statistical inference, alleatory probability is used somehow to make epistemic statements. This is the first role of probabilistic reasoning. The second role is in evaluating procedures. We understand statistical thinking to be based on these two roles for probabilistic reasoning. This allows us to elaborate our definition of statistical thinking by stating that it involves two principles:

1. Statistical models of regularity and variability in data may be used to express knowledge and uncertainty about a signal in the presence of noise, via inductive reasoning.
2. Statistical methods may be analyzed to determine how well they are likely to perform.

The downside of spelling out a definition is that it can be easy to get sidetracked on the details. For starters, we intend "signal" to denote general underlying phenomena and relationships of interest, whereas "noise" refers to sources of variation that are being separated from the signal. We find these terms helpful partly because the nonparametric regression model, where they become explicit, is a useful archetype. Furthermore, we believe that there is at least some modest historical evidence to support the importance of such a basic dichotomy. Stigler (1999) considered why psychology adopted statistical methods so much earlier than economics or sociology, and why astronomy did do so even earlier. His answer was that the theory of errors, arising in astronomy, was based on a conceptualization encapsulated by "observation $=$ truth $+$ error," and that psychophysics was able to introduce this to psychology via careful experimental design. Using our words, this suggests that the idea of considering data to be generated by combining signal and noise was essential to the historical development of statistical thinking.

A related detail is that, just as there are disagreements about the subtleties of the nonparametric regression model and its application, there are important issues surrounding the role of modeling in statistics. We intend to use "statistical model" very broadly, with the only restriction being that probability is involved, so that the notion covers models with relatively weak assumptions, as in a two-sample permutation test, or strong assumptions, as in many Bayesian multilevel hierarchical models. Our formulation cannot accommodate the perspective of Breiman (2001), but we believe that it is entirely consistent with the views given in discussions of that article by Cox (2001) and Efron (2001). Here we are also remaining agnostic about the extent to which a model may be "explanatory" or "empirical," as discussed by Cox (1990) and Lehmann (1990), recognizing that "[these descriptions] represent somewhat extreme points of a continuum" (Kruskal and Neyman 1956). Rather, we believe that when Box (1979) stated that "all models are wrong, but some are useful," he was expressing a quintessential statistical attitude.

[Received September 2008. Revised September 2008.]

REFERENCES

Birnbaum, A. (1971), "A Perspective for Strengthening Scholarship in Statistics," *The American Statistician*, 25, 14–17.

Box, G. E. P. (1979), "Robustness in the Strategy of Scientific Model Building," in *Robustness in Statistics*, eds. R. L. Launer and G. N. Wilkinson, New York: Academic Press.

Breiman, L. (2001), "Statistical Modeling: The Two Cultures" (with discussion), *Statistical Science*, 16, 199–231.

Brown, E. N., Kass, R. E., and Mitra, P. (2004), "Multiple Neural Spike Train Analysis: State-of-the-Art and Future Challenges," *Nature Neuroscience*, 7, 456–461.

Bunker, J. P., Barnes, B. A., and Mosteller, F. (1977), *Costs, Risks, and Benefits of Surgery*, Oxford: Oxford University Press.

Cox, D. R. (1990), "Role of Models in Statistical Analysis," *Statistical Science*, 5, 169–174.

——— (2001), Comment on "Statistical Modeling: The Two Cultures," by L. Breiman, *Statistical Science*, 16, 216–218.

Efron, B. (2001), Comment on "Statistical Modeling: The Two Cultures," by L. Breiman, *Statistical Science*, 16, 218–219.

Gnanedesikan, R. (1990), "Looking Ahead: Cross-Disciplinary Opportunities for Statistics," *The American Statistician*, 44, 121–125.

Kass, R. E., Ventura, V., and Brown, E. N. (2005), "Statistical Issues in the Analysis of Neuronal Data," *Journal of Neurophysiology*, 94, 8–25.

Keller-McNulty, S. (2007), "From Data to Policy: Scientific Excellence Is Our Future," *Journal of the American Statistical Association*, 102, 395–399.

Kruskal, W., and Neyman, J. (1956), "Stochastic Models and Their Applications to Social Phenomena," unpublished lecture at Joint Statistical Meetings, Detroit; referenced by E. L. Lehmann (1990).

Lehmann, E. L. (1990), "Model Specification: The Views of Fisher and Neyman, and Later Developments," *Statistical Science*, 5, 160–168.

Lindsay, B. G., Kettenring, J., and Siegmund, D. O. (2004), "A Report on the Future of Statistics," *Statistical Science*, 19, 387–413.

Mallows, C. (1998), "The Zeroth Problem," *The American Statistician*, 52, 1–9.

Paninski, L. (2003), "Convergence Properties of Three Spike-Triggered Analysis Techniques," *Network: Computation in Neural Systems*, 14, 437–464.

Rubin, D. R. (1993), "The Future of Statistics," *Statistics and Computing*, 3, 204.

Snee, R. D. (1990), "Statistical Thinking and Its Contribution to Total Quality," *The American Statistician*, 44, 116–121.

Stigler, S. M. (1999), *Statistics on The Table: The History of Statistical Concepts and Methods*, Cambridge, MA: Harvard University Press, Chap. 10.

This article has been cited by:

1. Julie Legler, Paul Roback, Kathryn Ziegler-Graham, James Scott, Sharon Lane-Getaz, Matthew Richey. 2010. A Model for an Interdisciplinary Undergraduate Research ProgramA Model for an Interdisciplinary Undergraduate Research Program. *The American Statistician* **64**:1, 59-69. [Abstract] [PDF] [PDF Plus] [Supplementary material]
2. Elart von Collani. 2010. Response to 'Desired and Feared—What Do We Do Now and Over the Next 50 Years' by Xiao-Li MengResponse to 'Desired and Feared—What Do We Do Now and Over the Next 50 Years' by Xiao-Li Meng. *The American Statistician* **64**:1, 23-25. [Abstract] [PDF] [PDF Plus]
3. Brian C. Kotz. 2010. Thoughts on the Importance of the Undergraduate Statistics Experience to the Discipline's (and Society's) FutureThoughts on the Importance of the Undergraduate Statistics Experience to the Discipline's (and Society's) Future. *The American Statistician* **64**:1, 15-18. [Abstract] [PDF] [PDF Plus]
4. Roger W. Hoerl, Ronald D. Snee. 2010. Moving the Statistics Profession Forward to the Next LevelMoving the Statistics Profession Forward to the Next Level. *The American Statistician* **64**:1, 10-14. [Abstract] [PDF] [PDF Plus]
5. Robert G. Easterling. 2010. Passion-Driven StatisticsPassion-Driven Statistics. *The American Statistician* **64**:1, 1-5. [Abstract] [PDF] [PDF Plus]
6. Xiao-Li Meng. 2009. Desired and Feared—What Do We Do Now and Over the Next 50 Years?Desired and Feared—What Do We Do Now and Over the Next 50 Years?. *The American Statistician* **63**:3, 202-210. [Abstract] [PDF] [PDF Plus]
7. Robert A. Oster. 2009. Section Editor's NotesSection Editor's Notes. *The American Statistician* **63**:2, 187-188. [Citation] [PDF] [PDF Plus]

Comment

Iain JOHNSTONE

First, I would like to congratulate the authors for a stimulating article on the vital challenge posed by renewing and expanding our graduate education programs in statistics. There can be no doubt that our subject is changing rapidly, and that our training should periodically adapt accordingly.

There is much to agree with in the particular desiderata that they cite, although it is hard to imagine any blanket mandates that must apply in every departmental situation: combine the dazzling variety of institutions (e.g., public/private, research/teaching emphasis, urban/rural, large/small) with the fact that statistics, perhaps uniquely, actually or potentially interacts with *every* discipline of the modern university (in, say, medieval history [see, e.g. Feuerverger et al. 2008]) and surely in varying ways.

Having (mis)spent a few years in university administration, I suggest that inertia and resistance to change of graduate programs are generic traits of organizations, and that statisticians need not feel *uniquely* guilty. Indeed, a physics department chair recently complained to me of similar phenomena! Naturally, having company does not let us off the hook. My comments now turn to challenges of implementation.

I have in mind an energetic statistics department chair who wishes to lead his or her department through discussion and implementation of a curriculum review, likely along the lines propounded by Brown and Kass. This can be a daunting exercise, demanding much time from many people and fraught with land mines, both known and unknown, that if detonated might doom the best-intentioned effort. Perhaps the editors of this journal could identify three or more recent examples of successful innovations in graduate statistics programs and invite the chairs (or other leaders) to collaborate on an article that collects advice, experiences, and good practices. (One example of such an article from the early 1990s is Kettenring (1994).) Such an article might grapple with issues including, but of course not limited to:

- How to define the reformed program's niche and comparative advantage in the diverse graduate statistics marketplace
- How to engage the attention of faculty who may think they are already too busy with research, teaching, and service
- Having gotten this attention, how to avoid being stymied by divergent and often strongly held opinions
- How to build support beyond the department for the new program, for example, with the administration and with other departments in the institution that may be impacted
- How to find resources, within or beyond the institution, for supporting the development of the new curricula
- How to optimize recruitment of well-matched students, including strategies for attracting members of underrepresented groups
- How to maintain focus on the reform over a period of years
- How to measure progress and success.

A second, often thorny issue is tenure review for scholars whose work is primarily interdisciplinary. In many universities, the challenge will be how to make a compelling argument that the interdisciplinary research is done *well*—that is, at a quality level consistent with the institution's criteria for tenure. Even if the statistics department has good procedures for such an assessment, there can be a real issue with "higher" committees that lie in wait along the path to final approval of tenure. This topic was not addressed in the otherwise wide-ranging IMS Panel on Cross-Disciplinary Research in the Statistical Sciences (1990). It could be helpful if one or more of our professional societies, or even the NRC's Committee on Applied and Theoretical Statistics, would produce a consensus document providing some guidelines on various appropriate ways to recognize and document effective and high-quality cross-disciplinary work by statisticians. Such a document could be referenced as a benchmark in departments' deliberations and perhaps also later as they make the case to university administrations.

REFERENCES

Feuerverger, A., Hall, P., Tilahun, G., and Gervers, M. (2008), "Using Statistical Smoothing to Date Medieval Manuscripts," in *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. IMS Collections*, Vol. 1, Beechwood, OH: Institute of Mathematical Statistics, pp. 321–331.

IMS Panel on Cross-Disciplinary Research in the Statistical Sciences (1990), "Cross-Disciplinary Research in the Statistical Sciences," *Statistical Science*, 5 (1), 121–146.

Kettenring, J. R. (ed.) (1994), *Modern Interdisciplinary University Statistics Education: Proceedings of a Symposium*, Washington, DC: National Academy Press.

Iain Johnstone is Professor of Statistics and Biostatistics, Department of Statistics, Stanford University, Standford, CA 94305 (E-mail: *imj@stanford.edu*).

Comment

Alison GIBBS and Nancy REID

We extend our thanks for an extremely interesting and provocative article. It will surely influence our thinking as we approach future curriculum reform. The authors' description of the "risk-averse" statisticians produced by our undergraduate programs immediately caught our attention. These graduates are all too familiar to us, and we were happy to see the emphasis on this problem. One of us teaches a graduate course in statistical consulting, and every year our first challenges are to help students find the confidence to tackle an unfamiliar problem and help them learn to acquaint themselves with the data and the meaning of nonroutine questions phrased in the language of a client before attempting to apply sophisticated statistical methods.

Teaching statistical reasoning is difficult. We agree with Brown & Kass that we need to rethink our undergraduate curricula, with a focus on statistical thinking. Their two principles can be used as the underlying theme in all of our courses, both applied and theoretical. But for students to move beyond their reticence to plunge into a real-world problem, they also must have experience in the statistical process, which is the intuition and experience that many of us have acquired only through years of experience. To give our students a running start, we need to engage them in varied rich, real problems. We applaud recent contributions on how to carry this out; for example, Weldon (2008) has argued that formalization and structure are necessary, but only after immersion in data, and Nolan and Temple Lang (2007) have introduced a method for documenting the statistical problem solving process, including the explorations and false starts that typically are not reported.

Brown & Kass cite the difficulty of finding a textbook for a broad course emphasizing statistical reasoning. We have been profoundly influenced by the classic text by Cox and Snell (1981), which we consider the embodiment of statistical thinking, although it must be admitted that students find it rather concise. The text of Davison (2003) has something of the same spirit, while covering a wider range of topics.

Brown & Kass differentiate between the training of short-term statistical consultants and statisticians who will be involved in long-term engagements. Statisticians working in either of these activities will improve any project to which they contribute. And we believe that it may not be important to differentiate between them in the training that statisticians need. The repertoire of a short-term consultant requires much more than mastery of an encyclopedia of techniques and the ability to reduce each practical problem to an approximately appropriate technique. Statistical consultants also must understand the nature and sources of variability in data and the logic of quantitative methods, and have the ability to generate ideas for sensible approaches to understanding the messages hidden in data.

Our experience in statistical consulting only rarely involves straightforward application of a parametric technique in a statistical software package. Our consulting students come to the course thinking that they will succeed with a well-stocked statistical arsenal, and they are surprised by the initial lessons they learn—that almost all of an initial meeting will be occupied by the statisticians understanding the clients' science, and that even if they do not have techniques at their fingertips, their training in statistical thinking makes them surprisingly well prepared to advise on directions to consider in someone else's project. Much more important than acquiring an encyclopedic knowledge of techniques is acquiring experience in applying statistical thinking in a substantial number of situations encompassing a broad range of applications.

Whether training as short-term consultants or long-term collaborators, our students need exposure to lots of real data with interesting contexts. While deep engagement may be difficult to experience in the time-frame of a course, even fairly small projects can teach the important lesson that analysis carried out by a statistician in isolation without continued communication with the client is unlikely to be satisfactory to anyone.

Both of us have had the opportunity to focus on the teaching of statistical literacy to small classes in our freshman seminars. A major focus of our seminars is on using current reports from the news media to teach our students to think critically about the use and abuse of statistics as reported. There are no mathematics or statistics prerequisites for these seminars; many students take them as part of their breadth requirement, and very few students are from quantitative disciplines. Because one aim of the course is to help our students acquire quantitative literacy, we focus on an accessible conceptualization of statistical thinking that encompasses the role of chance in natural phenomena, how we can reason sensibly in the presence of this uncertainty, the influence of the assumptions that we make, and the possibility that we might be wrong. This nontechnical emphasis on statistical thinking has profoundly influenced how we approach our upper-division courses (although it must be said that the upper-division courses continue to emphasize statistical methods and statistical theory). Brown & Kass suggest that helping students develop statistical thinking should be a primary goal of statistical training at all levels. We agree, but have found that the more advanced the course, the more difficult this is to teach. The students seem to struggle with an expectation of some component of evaluation based on these skills, as do the graders.

Emphasis on statistical thinking goes hand in hand with what Brown & Kass call "flexible cross-disciplinarity." For statistics students to engage in interdisciplinary research, they must be able to call on a store of statistical intuition, and to communicate some of this intuition to other scientists. To learn this, there seems to be no substitute for doing—so how do we get there from here? There are various administrative and organizational hurdles, and the difficulty of overcoming them will vary from

Alison Gibbs is Senior Lecturer (E-mail: *alison.gibbs@utoronto.ca*) and Nancy Reid is University Professor (E-mail: *reid@utstat.utoronto.ca*), Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada.

one institution to another, but these are probably relatively minor compared with the hurdle for statistical science of playing a major role in interdisciplinary teams while at the same time maintaining and strengthening the core discipline of statistics. Without this core, it is hard to see where the advancement of statistical thinking will come from. Thus, while admitting an outstanding student with little background in statistics to a graduate program requires administrative flexibility and some determination, it is also the case that a graduate program in statistical science, however broadly construed, should probably require that the vast majority of its students have a fairly strong background in quantitative work. It would be unthinkable to admit students with one undergraduate course in chemistry, or English literature, to graduate study in the field; if statistical science is to be a field, surely a basic set of quantitative skills is needed along with the statistical thinking that we wish to develop.

Many of us were attracted to the field of statistics because of the opportunity, as described by Tukey, to "play in everyone's backyard." We are not sure how one acquires the curiosity necessary to engage in these excursions, but we suspect that it is innate in many mathematically talented students who choose to study statistics. We want to attract quantitatively talented students with this curiosity regardless of their background, but feel that we have the responsibility to ensure they receive training that gives them the flexibility and broad range of interests that is a particular joy of statistical science. To be concrete, if a student's research involves a collaboration in marine biology, with the attendant investment in effort to learn this science and develop a common language for the collaboration, will that student also have the opportunity to develop other collaborations in the future, or is this a commitment to becoming a "statistical marine biologist"? This again comes back to the core; presumably someone well trained in statistical thinking and statistical theory will indeed have this flexibility (and interest), but again, how do we adjust our curricula for students to cover all of these bases?

We close by describing two initiatives in Canada that are related to the themes of this discussion. The need for statisticians to have deep cross-disciplinary knowledge was recognized in the development of the Canadian professional accreditation program for statisticians. There are two levels of accreditation, beginning with Associate Statistician (A.Stat.) and progressing to Professional Statistician (P.Stat.). The requirements for the A.Stat. designation are university courses in statistics, along with demonstrated knowledge in a substantive field other than statistics, as could be obtained by coursework corresponding to a minor in the field, work experience, or a combination of coursework and work experience. The National Institute for Complex Data Structures (NICDS) was formed to develop interdisciplinary teams of researchers with statistical leadership. The conception of the institute was influenced by the success of SAMSI, as well as by the successful and varied infrastructure initiatives in mathematical sciences in Canada. The buy-in from our scientific collaborators has been extraordinarily positive, and research projects have been established in areas of social, health, and physical science. Limitations have included the slow growth of funding, the shortage of statisticians in the pipeline, and, it must be said, some resistance from our statistical colleagues. Nonetheless, we sense an emerging consensus that collaborations like those undertaken by the NICDS are absolutely essential not only for science, but also for the strength of statistical science.

These two initiatives recognize the need for statisticians to acquire cross-disciplinary expertise, but they have a somewhat top-down pressure. The challenge for our discipline is to build from the bottom up; to embed the development of cross-disciplinary expertise in our statistics programs from the outset. This indeed requires a culture shift, and we are grateful to Brown & Kass for initiating what we hope will be an ongoing and passionate discussion across the field. Their article challenges us to rethink our approach to statistical training and statistical research, and we look forward to engaging in this challenge.

REFERENCES

Cox, D. R., and Snell, E. J. (1981), *Applied Statistics: Principles and Examples*, Boca Raton: Chapman & Hall/CRC.

Davison, A. C. (2003), *Statistical Models*, Cambridge: Cambridge University Press.

Nolan, D., and Temple Lang, D. (2007), "Dynamic, Interactive Documents for Teaching Statistical Practice," *International Statistical Review*, 75, 395–321.

Weldon, L. (2008), "Experience Early, Logic Later," in *OZCOTS-2008 Conference in Melbourne, July 3–4, 2008*. Available at *http://www.stat.sfu.ca/~weldon/papers/49.experience.pdf* (accessed November 19, 2008).

Comment

David MADIGAN and Andrew GELMAN

We agree wholeheartedly with Brown and Kass that something has indeed gone wrong with the way in which we attract and educate students in statistics. The problems begin with the standard unappealing and outdated introductory undergraduate course and persist through many, if not most, graduate programs. Our undergraduate courses focus on an exquisitely narrow set of topics that has changed little in 30 or more years. At the graduate level, we still persist with the increasingly untenable notion that there should be a core (and rather large) body of knowledge that *all* statistics students should know.

We see parallels with the discipline of engineering. Specialization into subdisciplines, such as civil engineering and chemical engineering, has existed for over a century, and while all engineers may share a certain mode of thinking, specific technical knowledge and skills divide along subdisciplinary lines. It is surely premature for statistics to subdivide into hard and fast subdisciplines, but we believe that some degree of specialization is in order. However, we also believe that specialization along applied versus theoretical lines is precisely the *wrong* type of specialization; this particular distinction reinforces the notion of the theoretical statistician developing mathematical artifacts without reference to any scientific enquiry while the applied statistician conducts the intellectually less challenging task of implementing the theory. The compleat statistician must span both aspects.

We believe that the characterization of statistics as a branch of mathematics also underlies many of the problems Brown and Kass describe. According to the Wikipedia entry for "statistician," the core work of a statistician is "to measure, interpret, and describe the world and human activity patterns within it." This seems about right to us—so how is it then that statistics came to be seen as a branch of mathematics? It makes no more sense to us than considering chemical engineering as a branch of mathematics. Both are highly quantitative subjects, and both use mathematics extensively. But in statistics, a purely mathematical agenda is often at the forefront. A statistics department attempting to go against these forces may meet resistance. (A story: We know of a top statistics department that had an interesting applicant with a math GRE of 650 (out of a possible 800). The dean tried to talk the department out of admitting this student. The department stuck to its guns, and the student is doing well.) Statistics departments often recruit mathematically adept students without regard to, for example, their potential to take leading roles in scientific teams. The net result is that our discipline has many outstanding mathematicians but few scientists in the mold of Fred Mosteller.

An example of the new style of statistical thinking described by Brown and Kass appears in the formula $y = f(x) + \varepsilon$. What is appealing about this expression is that the focus is on the deterministic model $f(x)$, rather than (as is traditional in statistics) the error distribution. Recall that in standard statistical notation, the notation f (generic mathematical notation for "function") has the privileged meaning of "probability density function." We believe that it is generally more important to model the mean than the error function, and moving to the generic "f" is a good start.

Statistics faculty recruiting provides another opportunity to effect change. Departments that kick-start the discipline out of its current rut will have many faculty deeply engaged in *different* interdisciplinary endeavors. Skilled "statistical thinking" cannot derive from experience in just one area. Indeed, one of the difficulties in our occasional efforts within statistics to discuss the future of our discipline is the often-narrow perspective that each of us brings to the table. Brown and Kass have done an outstanding job of generalizing from their neuroscience perspective, but nonetheless, the perspective of a social science statistician or a clinical trials biostatistician, to pick two examples, inevitably would be different and no less important.

Finally, as statisticians we should show some humility when recommending methods to others. For example, education researchers have long accepted the importance of randomization and other methods for facilitating "evidence-based" inference. But when devising our own educational plans, we resort to the usual mixture of introspection and anecdote that we deplore in others. We know of no easy way around this incoherence, but it should at least make us wary about over-certainty in our recommendations.

David Madigan is Professor (E-mail: *madigan@stat.columbia.edu*) and Andrew Gelman is Professor, Department of Statistics, Columbia University, New York, NY 10027.

  DOI: 10.1198/tast.2009.0022

Comment

Samad HEDAYAT and John STUFKEN

Brown and Kass present a set of recommendations for changes in how we should train statisticians. Based on their experience in neuroscience, they lay out a plan that promotes a strong emphasis on interdisciplinary training and more flexible admission standards for graduate students. They also propose changes in graduate programs and curricula to enable other components of their plan. They suggest that these changes will be critical for our discipline to remain vibrant.

Keeping our field vibrant by making important contributions to scientific problems is obviously a highly desirable and vitally important goal. But the proposed path toward reaching this goal is highly challenging, potentially perilous, and, if we are not extremely careful, possibly even counterproductive. We hope that the article and comments will stimulate a broader discussion on this important topic.

As Brown and Kass point out, they are not the first to draw attention to the inadequate training that our students receive and the need for action. "Good statistics is not equated with mathematical rigor or purity but is more closely associated with careful thinking.... We must help recruit strong students to statistics... by improving our courses and programs." Thus wrote Hogg (1991) in an article titled "Statistical Education: Improvements Are Badly Needed."

Indeed, a vibrant discipline should have a continual discussion on the adequacy of its programs in meeting new and continuing demands. As observed by Brown and Kass, there are many new demands. Although statisticians have long found their inspiration in applied problems, the massive amounts of data collected nowadays call for new statistical methods, innovative approaches, and often a deep understanding of the scientific problems. The opportunities for statisticians are virtually limitless, and many have found ways to benefit. So have many others with a quantitative knack. Although these opportunities come with gigantic challenges (see also Lindsay, Kettenring, and Siegmund 2004), it is absolutely critical that we embrace the opportunities and overcome the challenges.

Brown and Kass propose a path by which to do this. This is a very welcome contribution, although we believe that great care is needed when deciding whether to follow this path. In the remainder of this comment we focus on some of the issues that give us pause and that should be part of the envisioned broader discussion.

Workload. Successful implementation of the proposed path implies a significant increase in faculty workload. For details, we refer to some of the comments below. Active faculty do not have spare time, so something will have to give. This may require a change in how we view and account for our teaching loads, which could meet with major institutional objections.

Mentoring. Whereas dedicated mentoring of students is important under any circumstance, it becomes absolutely critical as we increase the flexibility of our programs and broaden the spectrum of students that enter these programs. Mentoring of students who traditionally have not been part of our programs will be extremely challenging. We wonder how many programs are well prepared for the additional burden that will come with this. Although some probably are, we expect that many are not.

Training for Faculty. While some statistics faculty are already heavily involved in interdisciplinary collaborations, the proposed path would require that many more move in that direction. Some will not be willing to do so (and others should not be, based on their areas of strength). But even for those who are willing, learning another field requires a major investment of effort and time. Although this investment is required to be an effective mentor, enabling faculty to accomplish this will require workload adjustments or other incentives.

Academic Requirements. For many departments, a successful transition to the model that Brown and Kass advocate will require drastic changes in how promotion and tenure cases are evaluated. If we expect a nontenured assistant professor to develop and maintain meaningful interdisciplinary collaborations, then we must recognize that this will require a huge investment of time and effort. Based on current expectations of research productivity, the nature of publication outlets, development of an independent research program, and so on, such a person may have a difficult time obtaining tenure. For many departments, this will require major changes.

Academic Inertia. It is hard to make major changes in academia, especially if these require money (even when there is the potential to bring in more money in the future) and involve structural changes in faculty workloads. There could be major almost insurmountable obstacles to achieving the vision expressed by Brown and Kass at many institutions.

Leadership. Successful implementation of the proposed plan will require exceptional leadership at many different levels, from local departments to national organizations. It can be argued that as a discipline, we do not exactly have a stellar record of providing this type of leadership.

Postdoctoral Programs. Statistics has only a limited number of successful postdoctoral programs. This is often attributed to the job opportunities for statisticians in industry and government. But if most statistics units at research universities were to develop postdoctoral programs and insist on filling their tenure-track positions with candidates who have postdoctoral experience, then we could make a major step forward in training future statisticians. The additional, say, 2 years of postdoctoral experience would, in combination with the years as a graduate student, enable training that could offer breadth and depth in

Samad Hedayat is Distinguished Professor, Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7045 (E-mail: *hedayat@uic.edu*). John Stufken is Professor and Head, Department of Statistics, University of Georgia, Athens, GA 30602 (E-mail: *jstufken@uga.edu*).

statistical methods and theory, knowledge of another field, and meaningful interdisciplinary experience.

The Core. In our drive toward meaningful interdisciplinary collaborations, we need to remain mindful to strengthen the core of statistical research (Lindsay, Kettenring, and Siegmund 2004). Without that core, we do not have a cohesive discipline. Our departments are not large enough, and never will be, to represent every "flavor" of statistics. Therefore, along with flexibility within programs, recognizing the value of variability among different units of statistics is critical. We need some balance in terms of the type of departments that we have.

The Pipeline. In view of the opportunities, it is clear that there are not enough of us. More attractive curricula advocated by Brown and Kass should help alleviate this problem to some extent, but the STEM initiatives demonstrate that there are fundamental pipeline issues across the sciences and engineering. We must work alongside other disciplines to attract more students to the sciences. There is also the question of how we would handle the desired growth in the number of graduate students; in this day and age, can we expect major increases in the number of faculty in our departments? Another issue that has a great potential impact is that AP high school statistics has exploded in recent years. The number of exams taken grew from 58,230 in 2003 to 98,033 in 2007 (College Board 2008). This also means that many students with quantitative interests will get their first exposure to statistics in high school, and will get turned on or off at that level. Consequently, one of the most critical steps in addressing the pipeline issue is better preparation of high school teachers for teaching statistics. Although small groups in our community have been very active in this area (e.g., the ASA/NCTM Joint Committee on Curriculum in Statistics and Probability and the ASA's initiative on Guidelines for Assessment and Instruction in Statistics Education), we have a very long way to go toward putting qualified teachers in our high school classes (not to mention teachers who provide even earlier exposure to statistical ideas).

Funding. There is not enough funding for statistics. Whatever the reason for it, this lack of funding lies at the very center of many of our problems. We need high school teachers that are better prepared, improved curricula and better training programs, more funds to support the larger number of students that we hope to attract, funds to develop postdoctoral programs, and funds for many of our current faculty to acquire depth of knowledge in another discipline. Funds are available for these purposes from both the NSF and NIH, with recent increases for interdisciplinary opportunities. One could argue that we have not taken sufficient advantage of the available opportunities, and there is probably some truth to that, but it also is easy to argue that the odds have been stacked against us. As advocated by Lindsay, Kettenring, and Siegmund (2004), it is critical that the leaders of our discipline work with funding agencies to find creative ways to fund training and research programs in statistics.

Many of the foregoing issues are intertwined. It will take a collective effort to address them, and the pace of any progress will necessarily be modest. Nonetheless, in this time of tremendous opportunity, it would be a shame if we were not able to make a concerted attempt to tackle the challenges, and we are hopeful that the contribution by Brown and Kass can be an inspiration for further discussion and action.

REFERENCES

College Board (2008), "The 4th Annual AP Report to The Nation," available at *http://professionals.collegeboard.com/profdownload/ap-report-to-the-nation-2008.pdf*.

Hogg, R. V. (1991), "Statistical Education: Improvements Are Badly Needed," *The American Statistician*, 45, 342–343.

Comment

Deborah NOLAN and Duncan TEMPLE LANG

We want to congratulate Emory Brown and Rob Kass on an important, timely, and compelling article. They challenge each of us to think very seriously about the future of statistics education and practice, and our role in that evolution. Along with Brown & Kass, we too believe it essential for the health of the field to make significant changes and additions to the content and focus of statistics training at all levels to attract, retain, excite and inspire students to become statisticians. To us, it seems "obvious" that we should broaden our view of statistics education to incorporate, alongside existing mathematical content, the process of real-world data analysis, skills in computing and data technologies, and statistical experience in scientific contexts. Based on our experience, we believe that statistical programs need to:

- Focus on statistical experience, reasoning, and applications *throughout* statistical training
- Recognize computing as an essential building block for statistical learning, creativity, exploration, and practice
- Design new courses and curricula to attract bright, motivated students to the field
- Change the culture of statistics training to engage students in active, participatory "effortful learning" in addition to critical study.

We continue this comment by providing some details and our thoughts on these four important aspects of statistics education. We then describe some of the activities that we have pursued in our research and our teaching on these topics, and suggest how these might provide possible practical solutions to some aspects of the significant challenges enumerated by Brown and Kass.

We have been thinking about and working on making changes in these directions for many years. We believe strongly that the field of statistics is at a crucial tipping point, and that bold measures of reform in statistics curricula are called for. The changes are necessary both to attract and prepare future statisticians and to keep pace with the rapidly changing "big science" fields. Our experiences over the last 10 years have shaped our views on the subject. These experiences include:

- Organizing a summer school that engages students in applied research projects of statisticians, with an aim of encouraging undergraduates to apply to statistics graduate programs
- Designing and teaching new courses in statistical computing and data technologies
- Teaching faculty how to teach computing
- Revamping a graduate program to broaden the curriculum and the graduate student population
- Exploring how to make research activities and results available through dynamic, reproducible, interactive documents.

Before continuing, it behooves us to make explicit a few parameters in this discussion. When considering statistics training programs, there are several different levels and career goals to take into account: undergraduate preparation for the workforce, undergraduate preparation for graduate school, Masters preparation for the workforce, and Ph.D. preparation for academia and careers outside of academia. With one exception related to an introductory course to attract freshmen into the major, our discussion focuses primarily on advanced undergraduate and graduate students, not the service-oriented introductory statistics class. Commonalities and differences can be found across these different levels of training. One important commonality is teaching data analysis. The collective perspective is that data analysis is taught in all statistics programs; however, the phrase "data analysis" has many connotations, and we believe that it is often the case that "data analysis" experience is simply illustrating a particular statistical method by applying it to a pedagogically chosen data set. We use the term quite differently, to refer to formulating a statistically oriented approach to a scientific question, which involves much more than just applying one or more statistical methods. Also, when we refer to "computing," we mean not simply programming or numerical algorithms, but rather the broader notions of computational concepts, ideas, and skills for statistical inquiry and working with data. Both of these concepts are core elements of statistical thinking.

Statistical Experience. For those learning statistics, the intuition and experience required for good statistical practice are the hardest things to learn (Wild and Pfannkuch 1999) and to teach. They involve very different types of concepts and a new *dimension* of thinking than in mathematics. After years of studying mathematics and statistics from textbooks, statistics students have learned a toolbox of methods, but not necessarily how to frame a question in a meaningful way, for example, balancing constraints, resources, and context. Students may know how to use one or more of these tools but are not masters of the tools, and often use them with trepidation. They need training and practice in mapping a scientific question into a statistical approach and developing understanding, experience, and intuition of when and how to use statistical methodology in the scientific context. These are essential skills in statistical thinking that involve many aspects beyond selecting and applying statistical methods to data. However, most courses focus explicitly on statistical methodology, either the theory or the "application," and very few address the essential larger context. A result of this

Deborah Nolan is Professor, Department of Statistics, University of California, Berkeley, CA 94720 (E-mail: *nolan@stat.berkeley.edu*). Duncan Temple Lang is Associate Professor, Department of Statistics, University of California, Davis, CA 95616 (E-mail: *duncan@wald.ucdavis.edu*). The authors were supported in part by NSF grants DUE 0618865 and DMS 0636667.

focus on techniques is to train students as confirmatory consultants rather than engaged scientific collaborators. To add this important dimension to statistics programs, we advocate regularly teaching statistics at all levels from the standpoint of statistical concepts flowing from contextual problem solving with data. We realize that this is challenging, but that does not excuse us from avoiding it.

Computing. Traditionally, education and research in statistics have relied primarily on mathematics. However, the enormous increase in computational power over the past 20 years provides numerous opportunities for the field for both statistical practice and statistical education. Computing represents an alternative, complementary medium to help students understand and explore statistical concepts and methods. The ability to simulate and compute gives students and researchers a tangible laboratory for exploring statistical concepts to concretize mathematical abstraction, and provides a forum for gaining insight and intuition about potential new methodologies. Through computing, students actively engage in constructively framing instructions to do a particular task, for example, designing experiments to explore or confirm understanding of concepts. This is quite different from mathematical exposure to statistical concepts, where the student is passively accepting the results of theorems or cautiously manipulating symbols to prove a concept known to be true. If students had as much background in computing as they do in algebra and calculus, we would be able to exploit this additional medium much more effectively.

Besides leveraging the computer for pedagogical purposes, computing in its own right is an essential part of statistical training. Statisticians use computers almost exclusively to access data; filter, process, and explore data; iteratively model the data; and report findings about the data. Each of these steps requires a computer, and in fact each requires very different computational skills. Computing also provides a source of new research problems, such as stochastic algorithms, understanding computer networks, and software reliability. Furthermore, it has changed the nature of other scientific problems by providing a medium for acquiring and exchanging both data and statistical methods in such areas as computational biology, astrophysics, aeronautics, transportation engineering, and medicine. Without computational skills, one simply cannot engage in the application and practice of statistics, regardless of one's knowledge of the concepts. In addition, a good foundation in concepts of scientific and statistical computing and data technologies is essential to the ability to continue to adapt to rapid technological changes. Because most statistics students go on to apply statistics rather than study it academically, computational skills are vital, but as with data analysis, it is a dimension omitted from many statistical curricula.

Attracting Students. We agree entirely with Brown & Kass that statistical thinking and interdisciplinary interaction (or better, immersion) is key for a statistics student to learn. Brown & Kass also recommend presenting statistics as a deep subject, with serious content. Again, we wholeheartedly agree, and also add that we must present it as vibrant and relevant in the modern world as well as for the future. The repeated focus throughout undergraduate and graduate courses on the same concepts at different levels of mathematical rigor presents the view that the important statistical ideas have all been developed. Indeed, many students, even graduate students, do not encounter methods developed within the last decade or two within their courses. Also, the repetition of the classical material is not a compelling approach to attracting good students to the major. Similarly, in our experience at the graduate level, this approach does not attract students to advanced study or prepare them for research. The traditional statistics curriculum is based on the need to first present an intellectual infrastructure for understanding the statistical method. But instead, statisticians need to lead with real and interesting scientific questions and show how statistics "saves the day." Early and continued exposure to statistics in this form we believe will excite and interest students. They will be eager to learn about the statistical theory and take the more traditional classes on the fundamentals that we offer.

Changing the Culture. As Brown & Kass note, the culture of statistics is more one of confirming other people's work, and often criticizing it. A culture of changing the world, attacking the very hard problems, and "dreaming big" is associated more with physics, computer science, and engineering and seems to be quite removed from our field. Perhaps this "caution" is the nature of statistics and a good thing. However, being cautious and circumspect is quite different from a "can't do" attitude. As Brown & Kass note, we must instill in our students the self-confidence to immerse themselves in the subject matter discipline and work alongside the content experts. In our view, statistics students equipped with the unique skills of computing and experience with data can gain this confidence and be welcomed into scientific teams because they have something unique to contribute.

1. CHALLENGES, EXPERIENCES, AND SOLUTIONS

The Role of Introductory Courses. Over the last decade, many educators have focused attention on improving introductory statistics courses. These courses service thousands of students who take only one statistics course, typically to fulfill some general education requirement of the university or their degree program. However, the introductory course can be viewed as a recruitment opportunity rather than solely a vehicle for providing basic statistical literacy to the masses. We believe the field and the students would be significantly better off if this course showed the challenges and applicability of statistics to important policy and scientific decision making in many contexts, and taught students how to think statistically and creatively in these contexts. How can we present the role of statistics in addressing "big science" problems in introductory courses? One possibility is to develop an "honors course" for a small group of students that is creative and bold in the research-like experiences it provides.

Our experience in developing and running a summer program in statistics with Mark Hansen (Hansen, Nolan, and Temple Lang 2006) provides ideas on how such a course might work. In the summer program, undergraduates with limited backgrounds in statistics and computing are exposed to important, topical scientific research problems presented by statisticians working on a scientific team. The program was held at UCLA in 2005 and 2006, and was funded primarily by the Institute for Pure and Applied Mathematics. Recently the NSF awarded a grant to continue this program for four summers, beginning in 2009 at UC Berkeley and then moving to the National Center for Atmospheric Research (NCAR), Columbia University, and UCLA.

The core of the program consists of three data analysis projects spread over 6 days. Each project is lead by a research statistician, who organizes 2 days of activities around an applied project. The researcher presents the scientific problem and explains its importance, provides data, and prepares short talks and computer investigations that introduce the students to the material in stages. At each stage, the instructor guides a discussion in which the students come up with with different approaches for the subproblem, work in groups to follow up on one or more of these approaches, and return to discuss their findings. Students use R (R Development Core Team 2006) to explore and visualize the data. The aim is to keep the interaction fluid and make it easy to move from individual and small group activities to a short presentation on a topic by the speaker to informal class discussion and group presentations. Overall, we found that the students were captivated and engaged by their interactions with the researcher. With the help of numerous instructors and teaching assistants, they quickly mastered the computing tools and were excited about using them to uncover the basic structure of the data, get to the statistical problems that the data present, and gain a sense of how statisticians approach large, complex problems. The program has been successful in attracting a broad spectrum of students; for example, in 2005 & 2006, half of the participants were female (24/49) and one fifth (11/49) were from underrepresented minority groups.

Teaching Computing. While statistics students must learn practical details of computing such as programming language syntax and the names of useful functions, we must strive to teach higher-level concepts of computational thinking that enable students to approach computational tasks intelligently. This includes the ability to discuss and reason about computational problems precisely and clearly. Furthermore, as computing and data technologies continue to evolve rapidly and as we enter the era of mainstream parallel and distributed computing for scientific computing, it is essential that statistics students be in a position to continue to learn new aspects of computation based on a good foundation, rather than a thin memorization of specifics and ad hoc tricks. Statistics programs must prepare the student for the future, which undoubtedly involves computing.

Since 2004, we have been developing and teaching an upper-division course in our respective departments. The two courses are similar and have been developed in close collaboration. The overarching topics are data technologies and statistical and scientific programming. Although the course has no statistics prerequisites, students work with topical and relatively large data sets, performing exploratory data analyses using advanced data technologies and "modern" computationally intensive statistical methods that they typically do not encounter in other classes. These methods (e.g., CART, kth nearest-neighbor methods, naive Bayes classifiers, hierarchical clustering, spline smoothers) are intuitive and relatively easy to describe, and give students a sense of the power of modern statistics. We have observed amazing transformations in our classes as students who initially were unsure of their abilities in computing or otherwise reluctant to work with the computer gain the confidence and skills to tackle a wide variety of data problems. It is empowering because they are involved, active participants. The students find it interesting because the data are available for compelling topical questions, and many find it refreshingly different than more traditional classes. We also have found that the course attracts many students from other majors and graduate students from other disciplines; for example, at UC Berkeley the course is now taught every semester, with an enrollment of about 75 students.

Faculty Experience. For many faculty, there is a large divide between their computational training and what today's students are expected to do. Some faculty have kept up and learned how to "compute," but many have not, and many have done so in an ad hoc manner, which conveys to students that computing is not important. This is very unfortunate, because it means that new students do not get the opportunity to learn it either. So they are in the same position as previous generations, left to learn computing by themselves, and the results typically are quite poor, resulting in students with significant misconceptions, limited abilities, and lack of confidence. How do we break this cycle and provide the opportunity for students to learn this material? One approach that we have pursued is to develop workshops specifically to teach faculty and foster Internet discussion groups for instructors.

Besides developing new computing courses, we also have worked to develop expertise among faculty and graduate students at other institutions so that they can teach this important material to the current and next generation of students. To do this, we (along with M. Hansen of UCLA and R. Peng of Johns Hopkins) are organizing workshops to help faculty acquire knowledge, skills to acquire additional knowledge, and teaching practices in these new areas. The NSF provided us with funds for a series of three workshops. The first, held in 2007, brought together computing specialists and industry consultants (people who have employees in statistical roles) to advise us on preparing material for course and curriculum development plans. Two additional workshops (one held in 2008 and another scheduled for summer 2009) focus on providing the necessary background and skills for instructors who want to teach statistical computing courses, along with examples of how to include modern data analyses projects in their courses. The materials produced for these workshops and resources from our classes are available on the Web (Nolan, Temple Lang, and Hansen 2007). We also have created electronic mailing lists, discussion boards, and a wiki for continued discussion and assistance. Overall, we aim to build a community of educators interested in incorporating computing into the statistics curriculum and sharing course materials.

Course Materials. Finding interesting and topical scientific problems with accompanying data in a form accessible to instructors who want to teach in this experiential manner can be difficult. The Internet provides a great resource for data but often falls short in supplying analysis and context. Articles that present applications are plentiful in research journals, but the analysis is typically presented as a completed work, with the pedagogically important thought process that led to the conclusions and approaches omitted. Where will educators find a wealth of materials suited to this approach to teaching statistics? Vehicles for transferring the experience with data from working statisticians to students are needed.

A project that we are experimenting with (Nolan and Temple Lang 2007) offers a novel approach to providing students with statistical experience. The idea is to enable researchers to document all of their computations and analysis process so that they can be reproduced in their entirety for both themselves and their peers (Gentleman and Temple Lang 2007) (e.g., reviewers, editors, bosses). Researchers would work in an environment that captures their writings, computations, and thought processes in an electronic notebook. In essence, this "lab notebook" would be a database of all of the activities involved in the data analysis or simulation study, and could be projected into different "views" (e.g., code, the final paper, various "dead ends") to make the information available for different audiences. An important consequence of this approach is that these rich documents will provide a flow of materials from statistics researchers involved in scientific applications to the education community. These documents will provide resources to instructors to assist teaching in new ways by opening up the thought process and experiences involved in data analysis to both instructors and students. Moreover, these documents can be displayed with interactive controls, allowing the reader to explore different analysis choices (e.g., changing the values of nuisance parameters, discarding outliers). This technological approach will support a model for passive cooperation between statisticians active in research and consulting and the community of statistics educators. Instructors will then have libraries of real case studies that include data analysis projects and current research methodologies that show how statisticians think and work.

Adjustments. Fundamental changes in the training of statisticians will not follow a prescribed, straight path. At most institutions, the training process has been running fairly smoothly for 20 years or longer. We cannot anticipate all that will happen to our programs as a result of such modifications. Even the question of where to begin is not easily answered. Changes of this magnitude will have repercussions, and it is important to make adjustments, continue on, and not turn back to the old system that supposedly "worked." How do we begin? How do we ensure that students on different pathways do not slip through the cracks? It takes a concerted effort, along with perseverance, to make significant changes to a program.

Over the past several years, the UC Berkeley Statistics Department has been making major changes in its Ph.D. program. A task force of faculty and graduate students reviewed the program, paying particular attention to the first 2 years and to whether students were being adequately prepared for research. The goals of broadening our graduate students' education and also broadening our graduate student population provided the impetus for this reform. The task force recommended that the program (1) broaden the traditional first-year course requirements of two year-long courses in two of the three areas of probability, theoretical statistics, and applied statistics to include other courses, such as the new course "Probability for Applications," as well as courses from other disciplines, and (2) require students to embark on a short-term research project, internship, or other research activity during the first summer of the program. To accomplish these recommendations, two additional key changes were needed: (3) Replace the preliminary exams, which were held in the summer between the first and second years in the Ph.D. program, with the requirement of satisfactory progress in the first 1–2 years of graduate course work, and (4) develop individual course plans for incoming students with the graduate advisor and a faculty mentor. The transition to this new program has not been without problems and has required much effort and resources. Naturally, not all of the effects of such significant changes to the program were anticipated at the start of the transition, and the program continues to evolve. Currently the general sentiment is that the program encourages increased cross-disciplinary research, and that the changes are attractive to graduate students.

2. CONCLUSION

Brown & Kass's discussion of statistical thinking is very important. The concept is what most of us recognize as the essence of statistical contributions. Yet too often, the educational focus remains on techniques and mathematical presentation of concepts because of their convenience and familiarity. Perhaps the problem is that most academic statisticians have not had the experience that Brown & Kass speak of, and the "anachronistic conception" is being passed on through the generations. At a time of great change for science and statistics, statistics education is not evolving at a sufficiently rapid rate. Educators are mostly doing the same things over and over again with minor extensions, and there are few forces to cause us to change in response to general changes in science. This is not any one individual's fault, and there are many truly vibrant and novel statistics educators in academia, but as Brown & Kass mention, this status quo is the result in the aggregate and has us concerned and frustrated. Can statisticians take on the challenge to find bold new ways to teach statistical thinking and practice? Where will the impetus come from? Senior statisticians can step up to this challenge and create a community that supports this change, including encouraging and enabling more junior statisticians who are in the midst of this sea change to take important roles in the process.

In summary, we agree wholeheartedly with most of the ideas that Brown & Kass espouse, and we are grateful that these two eminent statisticians have taken the time to write this article that challenges our field. Unfortunately, these types of articles often elicit tacit agreement but little or no action. Again, many individuals will be enthusiastic about the opportunity for change, but in the aggregate, change will be difficult. This is especially true if university programs must change, especially at a time when budgets are being squeezed. But this topic is clearly important, and vital for our field. We must find a way to effect change. Perhaps guidance should come from an organization such as the ASA. We must focus on changing the "anachronistic conception of statistics" of Ph.D. students and recent graduates, and encourage senior statisticians to seriously challenge their own perspectives and support junior faculty in designing new statistical programs that emphasize statistical thinking and reasoning. We should pool teaching resources, perhaps hold workshops to foster new ways of teaching, and develop case studies for teaching. We might even train graduate students nationally to teach important topics, such as computing, rapidly. Together, interdisciplinary science, computing, and the digital world present a change point for the field of statistics, requiring us to think about what a modern statistics curriculum would look like if we had both the freedom to change and the resources to implement such change. For too long, the field of statistics has acted more passively to such change points and responded by merely adding topics to classes, not by seeking, considering, and embracing new paradigms.

REFERENCES

Gentleman, R., and Temple Lang, D. (2007), "Statistical Analyses and Reproducible Research," *Journal of Computational and Graphical Statistics*, 16 (1), 1–23.

Hansen, M., Nolan, D., and Temple Lang, D. (2006), "Undergraduate Summer Statistics Program," available at *http://www.stat.berkeley.edu/~summer/*.

Nolan, D., and Temple Lang, D. (2007), "Dynamic, Interactive Documents for Teaching Statistical Practice," *International Statistical Review*, 75 (3), 295–321.

Nolan, D., Temple Lang, D., and Hansen, M. (2007), "Wiki: Computing in Statistics: Model Courses and Curricula," available at *http://www.stat.berkeley.edu/~statcur/*.

R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0. Available at *http://www.R-project.org*.

Wild, C., and Pfannkuch, M. (1999), "Statistical Thinking in Empirical Enquiry" (with discussion), *International Statistical Review*, 67 (3), 223–265.

Rejoinder

Emery N. Brown and Robert E. Kass

The discussants have added considerable content to our article, making many relevant points and insightful comments. Much of what they say stands on its own with no need for further remarks from us, and we are very grateful for their contributions. In broad stroke, they, along with the many people who responded when we sent them our article, strongly agree with our primary points: statistics has a problem, the discipline needs to recognize its own evolution, and important steps include encouraging statistical thinking and increasing student participation in cross-disciplinary work. A frequent next reaction is "yeah, but it will be hard," sometimes followed by "and I'm not so sure about your specific recommendations." Here we respond to this and a few other items from the discussion.

1. PATHS TO REFORM

We agree that there should be no "blanket mandates" (Johnstone) or "a [single] path" toward reform (Hedayat and Stufken). Each environment presents its own opportunities and imposes its own constraints. On the other hand, we would not back away from our four guiding suggestions to increase research opportunities, be conscious of statistical thinking, require real problem solving, and encourage cross-disciplinarity. We did not intend to be directive about either the ambitions or the specific implementations; rather, we hope that every group of training faculty will ask themselves what they can do, or do better, along these lines. We agree with Hedayat and Stufken that resources are important, but, especially in the current economic situation, a creative approach to resource-neutral curriculum reform may be called for. Furthermore, in our experience, administrators respond most supportively to investments that are perceived as likely to reap returns. We think the broad directions that we suggest can help increase research productivity and improve student satisfaction.

2. OBSTACLES

In response to the call for fewer requirements, more flexibility, and earlier research opportunities, a beleaguered colleague wrote, "I have been fighting this battle essentially forever, and I don't think I will live long enough to see victory." He added that one motivation for continuing to offer many specialized courses is a desire to satisfy potential employers, who presumably require them. We find this dubious. The primary concern of employers is to get good people, those who can interact successfully, have problem-solving skills, and can quickly learn what they need to know. Furthermore, courses that treat a topic in several lectures rather than a whole semester allow students to say that they have been taught the material. We understand that some colleagues have argued this perspective without success, but in our view departments no longer have the luxury of avoiding change.

Our article located the cause of current difficulties in an anachronistic conception of the field. When we elaborated by saying that excessive attention to mathematical foundations has led to a caricature of statistical activity as short-term consulting, we did not mean to imply that consulting labs should be closed. Along with Gibbs and Reid, we recognize these labs' potential instructional value. Rather, our point is related to an overarching attitude about the field. Statisticians can be productive and useful as consultants, collaborators, or principal scientific investigators, but the opportunities increasingly involve long-term projects with teams of multidisciplinary workers.

An additional issue, raised by Hedayat and Stufken and also by Johnstone, is whether junior faculty should be advised to pursue cross-disciplinary research. Here we would agree it is probably unwise for an untenured faculty member to make a substantial investment in a completely new area. On the other hand, many new Ph.D. recipients are desirable faculty candidates in large part because of their cross-disciplinary accomplishments, knowledge, and interests. In our view, the institutional commitment in hiring them must include opportunities and the support they need to pursue those interests, and to work in related areas if they so choose. In our experience, admittedly drawn from unusual environments at Carnegie Mellon, Harvard Medical School, and MIT, procedures for evaluation of cross-disciplinary work at the time of promotion and tenure are not especially hard to implement.

3. TEACHING STATISTICAL THINKING

We very much appreciate the ideas voiced by Nolan and Temple Lang that computation is important both pedagogically and conceptually. We have nothing to contribute aside from mentioning that discussions with Carnegie Mellon's John Lafferty and Larry Wasserman have led us to suggest an addition to our two-item summary of statistical thinking: *Computational considerations help determine the way statistical problems are formalized*. This point may seem rather advanced and specialized, aimed at the relationship of machine learning and statistics, but fundamental examples include the criteria of mean squared error and least squares. We think that elementary classes can, and should, present these, noting that other criteria may be sensible, but minimizing averaged squared differences usually leads to easy analysis and implementation, and

Emery N. Brown is Warren M. Zapol Professor of Anaesthesia, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA 02114 and Professor of Computational Neuroscience and Health Sciences and Technology, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139. Robert E. Kass is Professor, Department of Statistics, Center for the Neural Basis of Cognition, and Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA 15217.

thus widely available software. Such discussions clearly identify computation as a conceptual component of statistics.

Madigan and Gelman observe, quite rightly, that when devising approaches to teaching, statisticians often "resort to the usual mixture of introspection and anecdote." They then add that "we know of no easy way around this incoherence." There is in fact a body of methods developed by educational psychologists that, while not "easy," are nonetheless relatively straightfoward and quite useful. A long-term collaboration between Carnegie Mellon's Marsha Lovett (a cognitive psychologist) and its Department of Statistics produced some relevant findings. This research began with an analysis of problem-solving transcripts by statistical experts and statistical novices. Both groups were given elementary data analysis problems and asked to voice their thought processes as they worked. The most immediate distinction between the two groups was that when faced with a problem statement and a set of data, naive students immediately tried to find a suitable statistical technique (e.g., chi-squared, t-test), whereas the experts began by identifying the scientific question. This led to a revision of computer labs for Carnegie Mellon's most widely enrolled course *Statisical Reasoning*, in which students were required to follow an explicit series of steps: (1) understand the problem (check data format; consider study design), (2) reflect on the question (state expected findings, identify relevant variables), (3) analyze the data (classify the variables, perform exploratory analyses, conduct formal analyses, report results), (4) draw conclusions (recall expectations, consider what results mean), and (5) summarize (choose key results to report, place conclusions in a substantive context). This set of required steps was seen as "scaffolding" that could be removed gradually as the course proceeded, forcing students to internalize them (Lovett and Greenhouse 2000). This work has gone on to produce a successful self-contained online course, recently highlighted in a perspective in *Science* (see Lovett, Meyer, and Thille 2008; Smith 2009). Furthermore, students at every level of training can benefit from seeing this outline of the steps taken by experts. As Nolan and Temple Lang point out, students "need training and practice in mapping a scientific question into a statistical approach... [and using] statistical methodology in the scientific context." But our main point here is that methods exist to help instructors better align strategies with goals.

An emphasis on the interplay among questions, analyses, and conclusions is highly intuitive. As Terry Speed (quoted by Weldon, who was cited by Gibbs and Reid) noted, "if students have a good appreciation of this interplay, they will have learned some statistical thinking, not just some statistical methods." In no way, however, should appreciation of statistical practice conflict with a solid understanding of statistical theory. In fact, when principles are applied to important problems, their value becomes more readily apparent. Gibbs and Reid mention the expectation among more advanced students that evaluation will be based mainly on technique. In many courses this will remain appropriate, but, as we stated in our article, one bothersome feature of most statistics curricula—especially at the undergraduate level—is their artificial bifurcation into theory versus application. This not only undersells the field, but also sends the wrong message; as Madigan and Gelman noted, "specialization along applied versus theoretical lines is precisely the *wrong* type of specialization." One solution will come from new teaching materials that define more inclusive courses. Together with Sam Behseta and Uri Eden, we are in the process of writing a text on probability and statistics with this goal in mind. We hope that other textbook authors, and instructors, will strive to help novices experience the close interaction of theory and application found almost universally among experts.

REFERENCES

Lovett, M., and Greenhouse, J. B. (2000), "Applying Cognitive Theory to Statistics Instruction," *The American Statisician*, 54, 196–206.

Lovett, M., Meyer, O., and Thille, C. (2008), "The Open Learning Initiative: Measuring the Effectiveness of the OLI Statistics Course in Accelerating Student Learning," *Journal of Interactive Media in Education*, available at *http://jime.open.ac.uk/2008/14/jime-2008-14.html*.

Smith, M. (2009), "Opening Education," *Science*, 323, 89–93.